Rapid Responses to:

EDUCATION AND DEBATE:
Jeanne Lenzer, Charles Warlow, Jeffrey L Saver, Chelsea S Kidwell, and Sidney Starkman

Alteplase for stroke: money and optimistic claims buttress the "brain attack" campaign ● Commentary: Who pays the guideline writers? ● Commentary: Thrombolysis in stroke: it works!
BMJ 2002; 324: 723-729 [Full text]

Rapid Responses published:

▼ The efficacy analysis of the NINDS trial is flawed
Jeffrey Mann  (22 March 2002)

▼ Thrombolysis in Stroke: Not Ready for Prime Time
Robert Solomon  (22 March 2002)

▼ Guidelines and Conflict of Interest
A.C. Anilkumar  (23 March 2002)

▼ Uncertainty and publication ethics in clinical research
Howard Mann  (24 March 2002)

▼ Patients need to know true risks and benefits
Dennis P Briley  (25 March 2002)

▼ Concerns
Jerome R Hoffman, LA CA 90077 USA  (25 March 2002)

▼ Self-interest and statistical bias may obscure real patient harm.

▼ Setting the record straight
David Faxon  (24 April 2002)

▼ Re: Setting the record straight
Jeanne M. Lenzer  (24 April 2002)

▼ Alteplase and Acute Stroke
Philip B Gorelick, BM Coull, HJM Barnett, AM Buchan, G del Zoppo  (9 May 2002)

▼ Neurologist Attacks Emergency Medicine
The authors, Drs. Saver, Kidwell and Starkman, start their commentary by stating “Intravenous tissue plasminogen activator (alteplase, tPA) administered within three hours of the onset of symptoms to appropriate patients with acute ischaemic stroke is a proved treatment of substantial benefit. Data from six trials enrolling patients within three hours show this reality unequivocally.” Is that true? How can the results be unequivocal if the stroke research community has failed to agree on an objective "gold standard" method for precisely correcting for critical prognostic variables that may significantly bias the interpretation of a stroke trial's results? An imbalance in baseline stroke severity between tPA-treated and placebo patients is a confounding variable that may markedly affect the
The efficacy analysis of the NINDS trial is flawed. The NINDS trial had a significant imbalance in baseline stroke severity between the tPA-treated and placebo patients in the 91-180 minutes subgroups. In the Marler article[1], where that information was divulged, the authors stated that they made appropriate adjustments for baseline NIHSS stroke severity scores as a confounding masking covariate. However, how scientifically accurate was that adjustment? In a recent telephone communication with Dr. Marler (private communication - March 2002), he admitted that the adjustment was not based on solid EBM evidence from previous scientific studies. Instead, it was based on a logistic regression equation using the raw data from the NINDS trial. The question then remains - is that statistical technique, which is only applied a posteriori after a research study has been concluded, the optimum way to correct for imbalances in baseline stroke severity? Roberts and Torgenson in their article[2] on methods of appropriately reducing the negative effects of chance bias, state "choice of baseline characteristics by which an analysis is adjusted should be determined by prior knowledge of an influence on outcome rather than evidence of imbalance between treatment groups in the trial." In other words, they are saying that the "gold standard" corrective-adjustment should be determined apriori - before any stroke trial is performed, and not after the fact. Is there a reasonably objective "gold standard" corrective-adjustment available that is based on solid EBM evidence gained from prior knowledge? Indeed, there is! Adams[3] demonstrated the marked effect that variations in baseline stroke severity have on the rate of excellent stroke outcome, and his TOAST graph demonstrates how steep and exponential is the curve relating baseline stroke severity to the rate of excellent stroke outcome. In a critical essay[4], I used the data from the NINDS trial and applied it to the TOAST graph's curve, and demonstrated that the efficacy effects of tPA in the 91-180 minutes group may entirely be explained by chance bias (imbalance between groups in baseline variables that may influence outcome). A formal version of that essay is due to be published in a medical journal in May, and a mathematical re-analysis of the NINDS trial's results (based on that particular perspective) may be presented at the Stroke Thrombolysis symposium in Lyon at the end of May.

The authors appear to be ardent believers in the accuracy of the NINDS trialist's interpretation of the NINDS trial, and they even quote the ATLANTIS trial's < 3 hour results[5] as confirming the efficacy of tPA. However, the placebo patients (median baseline NIHSS score of 12) in the ATLANTIS trial were much sicker than the tPA-treated patients (median baseline NIHSS score of 9) and the TOAST graph predicts a roughly 5-10% chance bias effect for every change in baseline NIHSS score of 1 point. Is that a fair comparison? Finally, the authors also state "these data have been reinforced by a pooled analysis of individual patient data from the six alteplase trials, involving 2776 patients from over 300 hospitals in 18 countries, and fully adjusting for any imbalances in baseline entry characteristics among patients allocated to active treatment and placebo (TG Brott et al, 27th International Stroke Conference, San Antonio, Texas, 2002)". Hopefully, the authors can be persuaded to provide precise details of the specific methodology that was used to fully adjust for any imbalances in baseline entry criteria, including a scientific comparison between those statistical adjustment techniques and the logistic regression equation used to develop the TOAST graph. In fact, if the authors are so convinced about the efficacy of tPA in acute ischemic stroke - perhaps they can respond to the ultimate scientific challenge, and make public the baseline NIHSS stroke severity scores and the rates of excellent stroke outcome at 3 months of those 2776 patients, so that an independent analyst can plot the rate of excellent stroke outcome against the baseline NIHSS stroke score for each patient-score level for all those placebo and treated patients. By plotting those two curves, imbalances in baseline stroke severity between the treated and placebo patients will no longer have to be considered, and the raw scientific data will be able to speak for itself - independent of any person's subjective biases.

Accuracy in the world of medical science is essential, and it is mandatory that the stroke interventionist community obtain a very accurate reflection of tPA's true efficacy in acute ischemic stroke before it continues to offer such risky therapy to the public. The risks of tPA therapy are substantial and include hemorrhagic transformation of an infarct secondary to tPA administration, inadvertently treating stroke mimic/TIA patients with tPA, misreading...
a CT scan that has subtle signs of a primary ICH, and protocol violations when administering tPA therapy. The authors own recent article6 demonstrates that the risk of hemorrhagic transformation of an infarct is directly related to the baseline NIHSS score, and that the rate of symptomatic hemorrhagic transformation is extremely high (>30%) in stroke patients with a baseline NIHSS score >15. While the risks associated with tPA therapy are substantial and clear, the true benefit of tPA therapy (its precise degree of efficacy) is still an unanswered question. Public disclosure of the raw scientific data may help resolve that unanswered question, and it may also give discerning clinicians the opportunity to more accurately calculate the risk:benefit ratio of tPA therapy in acute ischemic stroke.

References:


2) Roberts C. Torgerson DJ. Understanding controlled trials: baseline imbalance in randomised controlled trials. BMJ. 319(7203):185-185, 1999 July 17


4) Mann J. Truths regarding the NINDS tPA for acute ischemic stroke trial: Setting the record straight! Available at http://www.homestead.com/emguidemaps/files/tpaforstroke.html


6) Kidwell, Chelsea S. MD; Saver, Jeffrey L. MD; Carneado, Joaquin MD; Sayre, James PhD; Starkman, Sidney MD; et al. Predictors of Hemorrhagic Transformation in Patients Receiving Intra-Arterial Thrombolysis. Stroke. 33(3):717-724, March 2002

Robert Solomon, Assistant Professor of Medicine Ohio Valley Medical Center, Wheeling, WV 26002 USA

Send response to journal: Thrombolysis in Stroke: Not Ready for Prime Time

Dear Editor,

A number of interesting but highly contentious points have been made in the commentary by Saver, Kidwell, and Starkman (“Commentary: Thrombolysis in stroke works”) on the article by Lenzer.

Saver et al assert that Lenzer’s recommendation for avoidance of “all potential bias” is “unworkable and undesirable...extreme financial correctness” that would leave development of clinical guidelines to “ill equipped” non-experts.

Lenzer's recommendation is neither unworkable nor undesirable. "Financial correctness" is exactly what is needed to assure objectivity. One need not have financial ties to proprietary concerns in order to be an expert. In fact, one need not even have participated in the research in order to be an expert. One need only have a thorough understanding of the application of the therapies under consideration for inclusion in the guidelines and a comprehensive knowledge of the published literature that might serve to inform decisions made about what to include in the guidelines. To the extent that having participated in related research might prove valuable, the financial conflicts that arise therefrom could be mitigated by funding more research with public dollars, and that is clearly a direction in

http://bmj.bmjjournals.com/cgi/eletters/324/7339/723
which we should be moving.

Given their acknowledged financial ties, however, it is understandable that they would take this position. In this connection it is interesting to note the refusal by Susan Hellmann, MD, Chief Medical Officer for Genentech, to endorse a report of the Association of American Medical Colleges Task Force on Financial Conflicts of Interest in Clinical Research, “Protecting Subjects, Preserving Trust, Promoting Progress–Policy and Guidelines for the Oversight of Individual Financial Interests in Human Subjects Research” (http://www.aamc.org/members/coitf/). The report was approved by the AAMC Executive Council December 14, 2001 and was endorsed by 26 of the 28 members of the Task Force. (One member, Hedrick Smith, withdrew his name because he was unable to participate in the drafting of the report due to unexpected professional commitments arising from the September 11th attacks.) Hellmann declined her endorsement due primarily to her concern that the recommendations “present an impediment to research innovation.”

Saver et al claim, as if it were a statement of fact, that "Thrombolytic agents are highly efficacious in stroke." This assertion is supported by not a single study in the published literature. The only randomized, controlled trial (RCT) that appeared to show benefit, the NINDS study, yields a number needed to treat of 9 and a number needed to harm of 17 - numbers that hardly support a statement that the therapy is "highly efficacious." All of the other numbers that seem positive are the product of statistical manipulations and data snooping in an effort to demonstrate benefit by post-hoc analysis of unplanned subgroups, methods that are well understood to be useful only to generate hypotheses to test in further prospective RCTs.

Saver et al opine that "...the real scandal in acute stroke care is not that thrombolytic therapy is being used, but that it is not being used often or wisely enough." There is absolutely nothing scandalous about this fact of current practice, arising as it does from the minuscule percentage of patients who are potential candidates for this therapy and from physicians' prudent wariness regarding a modality that remains unproven.

The authors note, "Only 1-2% of acute stroke patients in the US are receiving thrombolytic therapy." They suggest, "Public education campaigns and reorganisation of medical services can substantially increase the proportion of patients treated." It is possible that public education campaigns could increase the number of patients who seek medical attention immediately upon the onset of a stroke syndrome. Treatment of such patients within 90 minutes of onset of symptoms might yield a larger number of favorable outcomes. It must be noted, however, that unless this takes place within a randomized, placebo-controlled trial, we will not know whether patients with favorable outcomes obtained benefit from the treatment or were, in fact, experiencing transient ischemic attacks.

According to Saver et al, "Concerns about the everyday effectiveness, as opposed to clinical trial efficacy, of thrombolytic therapy have been raised and constitute a call to action, not resignation." The Cleveland Area Experience should, indeed, constitute "a call to action." The action required, however, is the re-examination of the therapy in question and a move toward an attempt to replicate the results of the only RCT (NINDS) that appeared to show benefit. Now that it is apparent from results reported in a non-selective manner that the outcomes of NINDS may not be readily replicated in real-world clinical practice, it is appropriate to ask whether they can even be replicated in another formal RCT.

The authors declare that "Physicians caring for acute stroke patients can and should master the key elements of thrombolytic care or allow patients to be diverted to specialised acute stroke centers where thrombolytic therapy can be expertly administered." Physicians caring for acute stroke patients can and should master the key elements of acute stroke care that are of proven benefit. Fibrinolytic therapy is not yet one of those elements. Potentially unstable patients should not be diverted to more distant facilities in order to receive a treatment with an apparently unfavorable risk:benefit ratio, especially given that the vast majority of them will be found not to be candidates for that treatment,
according to the NINDS protocol. Furthermore, in many communities only a small minority of stroke patients present via ambulance, and patients in crisis do not triage themselves to hospitals that offer specific therapies.

Saver et al aver, "Ideally, as many patients as possible would be treated within 90 or 120 minutes of onset, when benefit is maximal." It may be true that patients treated within 90 or 120 minutes will derive enough benefit from the treatment to alter the risk:benefit ratio sufficiently to warrant a positive recommendation. But that is a hypothesis that has not yet been prospectively validated. If it ever should be prospectively validated, it is doubtful that the number of patients presenting early enough to be treated within 90 or 120 minutes will ever be significant.

Robert C. Solomon, MD, FACEP
Faculty, Emergency Medicine Residency
Ohio Valley Medical Center, Wheeling, WV, USA
Clinical Assistant Professor of Medicine
West Virginia School of Osteopathic Medicine
rcsmdnet@nauticom.net

Guidelines and Conflict of Interest

23 March 2002

A.C. Anilkumar, Pediatrician
Hardin Memorial Hospital,
Elizabethtown, Kentucky, KY
42701

Dear Sir, I am wondering whether the appearance of vasopressin as a pharmacological agent to use in Ventricular fibrillation and Ventricular tachycardia (Comprehensive ECC Algorithm (c)2001 American Heart Association) also has something to do with Industry Sponsorship. Unless more and more critical evaluations into these dealings were done, one might be forced to think that the current level of care is completely outdated and new drugs are the way to the future. Thank you for the review on Alteplase.

Dr. A.C. Anilkumar, Pediatrician

Uncertainty and publication ethics in clinical research

24 March 2002

Howard Mann, Program Associate, Division of Medical Ethics University of Utah School of Medicine Salt Lake City 84132

Editor--The report by Lenzer, and the associated commentary by Saver and colleagues, raise many serious issues, among which is the residual state of uncertainty concerning the efficacy of tPA in acute ischemic stroke.

Confronted by opposing interpretations of the aggregate data published to date -- and the now known baseline imbalance in stroke severity in the NINDS trial -- the practising clinician is presented with a conundrum: what action do the data support?

In this regard, the reported unwillingness of the NINDS trial's investigators and sponsor to provide patient-level data for additional analysis is particularly disturbing.

Emanuel and colleagues [1] have described seven requirements for the ethical conduct of clinical research, among which is social and scientific value. Social value presupposes the public dissemination of research results. This requirement is elaborated upon in A Standard for the Scientific and Ethical Review of Trials (ASSERT) I have formulated [2].

Implicit in this requirement is the necessity for the public dissemination of the complete data set acquired during a clinical trial, allowing interested investigators to apply recognized analytic techniques in an attempt to resolve (or diminish) residual uncertainty concerning the clinical implications of the trial's results.

With respect to the NINDS trial, this ethical requirement has clearly not yet been met.
Howard Mann, M.D. Program Associate Division of Medical Ethics University of Utah School of Medicine


Patients need to know true risks and benefits

25 March 2002

Dennis P Briley, Consultant Neurologist Stoke Mandeville Hospital, Mandeville Road, Aylesbury, Bucks HP21 8AL

I read the commentary on thrombolysis by Lenzer with interest (1). There is one additional concern about the NINDS trial that Lenzer did not note. The trial provided for early treatment of hypertension (2). This is an intervention that may worsen the outcome of acute stroke. Hypertension was managed in the same way in both the placebo group as well as the thrombolysis group. It is a possibility that the good outcome in the thrombolysis arm of this trial reflected a bad outcome from an active, and potentially harmful, treatment of the placebo group. I agree with Professor Warlow that the results are not convincing, but further trials are warranted (3). Our patients need to know what the true risks and benefits of this treatment are.


Concerns

25 March 2002

Jerome R Hoffman, Professor UCLA School of Medicine, LA CA 90077 USA

I am far less sanguine, however, about the BMJ’s decisions to solicit the two commentaries that accompany Ms Lenzer's article. Although it is hard to quarrel with Dr. Warlow’s thoughtful essay, it remains true that all four of the authors have financial relationships with the makers of tPA. Couldn’t the BMJ have found even a single commentator without such a tie? I am far less sanguine, however, about the BMJ’s decisions to solicit the two commentaries that accompany Ms Lenzer's article. Although it is hard to quarrel with Dr. Warlow’s thoughtful essay, it remains true that all four of the authors have financial relationships with the makers of tPA. Couldn’t the BMJ have found even a single commentator without such a tie? Is the BMJ tacitly supporting Saver et al’s bizarre claim that it is impossible to be an expert on a subject about which one does not have a conflict of interest – even when the subject is conflicts of interest themselves?!

Equally distressing is the BMJ’s decision to publish an unopposed attack on the scientific bases of the opposition to tPA. Ms Lenzer is not a scientist, and therefore it was most appropriate that she merely referenced some of the concerns of critics, without having tried to argue any of their points in detail. Her contention clearly seems to be that although there is substantial disagreement among people knowledgeable about the topic, financial conflicts of interest may influence the manner in which controversial issues are addressed. It would have been perfectly reasonable for the BMJ to have allowed the article to stand on its own; indeed it is hard to imagine that the BMJ would have solicited not one but two contrarian commentaries for an article praising the use of tPA in acute ischemic stroke. (The BMJ could hardly have predicted Warlow’s lukewarm take on tPA, given that he is an investigator in IST-3, which "is designed to generate the data needed to persuade everyone involved in stroke medicine [epidemiologists, public health physicians, politicians,
stroke physicians] that rt-PA thrombolysis should be more widely available".\) Especially in light of the enormously positive spin placed on tPA by both the lay and medical media, and the relatively limited opportunity critics have to express their opposition (in either the medical or lay media), it is disturbing that the BMJ solicited comment only from supporters of tPA.

It is ironic that Saver et al’s piece was published immediately above an article from the BMJ’s series on evidence based medicine, given that they repeatedly resort to the type of distorted and selective marshalling of "evidence" that violates fundamental principles of clinical epidemiology. Allowing these authors to dismiss the concerns referenced by Ms. Lenzer – with the implication that they are the experts, and she is merely an "intemperate" journalist – and without dissent from opposing experts, not only encourages readers to pay less attention to the issue of tPA for stroke. It also discourages concern about the paper’s primary subject, the influence of money on medical policy recommendations. "So what if there are conflicts of interest", one might ask, "if (as Saver et al get to proclaim unilaterally) the evidence is so clearly favorable"?

If the BMJ felt absolutely compelled to solicit an attack on criticism of tPA, why did it not also solicit a balancing paper – to allow readers to understand the debate, and reach their own conclusions? I am certain that among the many of us who believe the evidence is far from conclusive, and who worry that widespread community use of tPA for stroke might do more harm than good, any number would have been (and remain) willing to write that very paper.

JRH has provided expert consultation in lawsuits against physicians involving the issue of use of tPA in acute stroke. He has taken no personal compensation for this work, having donated all fees to the UCLA EM Residency Program.

**Self-interest and statistical bias may obscure real patient harm.**

---

We appreciate and agree with Dr Warlow’s commentary.

No matter how earnestly purported by Saver and colleagues, intravenous tPA is not a proven treatment for ischaemic stroke, and payment by tPA manufacturers to guideline and research consultants degrades the premise of unbiased expertise. Drs Saver, Kidwell, and Starkman disclose 81 ties to for-profit companies with interests vested in expensive therapies for stroke. Sponsorship of this magnitude does not “channel the self interest of profit making companies to improving stroke care,” it purchases spokesmen for manufacturers seeking to add credibility to their wares. Rather than improve the venerability of its payees, such sponsorship erodes our trust. We grant that it may be possible for someone with sponsorship to perform skeptical and impartial research, but submit that it is impossible to determine whether such research is truly unbiased. (We are happy to see exceptional cases, where sponsors have encouraged research that could cast doubt on their products and lead to financial loss.[1]) Being so much simpler to believe the scientist without such ties, we wonder if the insistence of certain doctors to be allowed to continue their financial relationships is based on scientific curiosity, or further financial gain.

We also are troubled by the claim that tPA has been proven in pooled data from 6 trials. Of these, only one, the NINDS trial, was randomized, controlled, and had as a primary a priori endpoint the efficacy of tPA given within 3 hours from stroke onset. (This trial is counted twice by Saver and colleagues.) The remainder were post-hoc analyses of earlier trials that both routinely ignored negative results regarding their own primary hypotheses and (using the same flawed pooling methods) demonstrated increased mortality in their tPA treatment groups (table). Such data manipulation goes contrary to meta-analysis methodology and confounds
rather than promotes the truth.[2]

<table>
<thead>
<tr>
<th>Trial (&lt;3h subgroups)</th>
<th>Three month mortality</th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>tPA</td>
<td>placebo</td>
<td>P</td>
<td></td>
</tr>
<tr>
<td></td>
<td>%</td>
<td>n</td>
<td>%</td>
<td>n</td>
</tr>
<tr>
<td>ATLANTIS A and B</td>
<td>17</td>
<td>4/23</td>
<td>5</td>
<td>2/38</td>
</tr>
<tr>
<td>ECASS I and II</td>
<td>18</td>
<td>24/130</td>
<td>12</td>
<td>14/115</td>
</tr>
<tr>
<td>Subtotal</td>
<td>18</td>
<td>28/153</td>
<td>10</td>
<td>16/153</td>
</tr>
</tbody>
</table>

It is important to compare the reported benefit and harm in the NINDS trial, the single rigorous randomized investigation of 3 hour tPA to date. For every 9 patients treated, one benefited. For every 16 patients treated, one was harmed by symptomatic cranial hemorrhage. For every 34 patients treated, one died as a result of such cranial hemorrhage. Thus, for about every 4 patients who benefited, one died. These numbers do not support the statement that tPA therapy is “highly efficacious.”

Saver and colleagues write that the NINDS “trial data underwent two independent audits, by an autonomous contractor funded by and reporting to the National Institutes of Health (not Genentech) and by the US Food and Drug Administration.” However, no information is provided about the results of these audits. More importantly, despite requests by multiple parties unaffiliated with Genentech, the NINDS investigators have to date refused to release their data for independent verification.

Given the animosity raised by this debate, we wonder if the results of the NINDS trial will ever be verified by independent reanalysis of its data or by confirmation in a future unbiased randomised trial. We wonder if the future of medical research will rely on investigators with such unashamed ties to for-profit corporations. Finally, until claims of positive results from NINDS are verified independently, we strongly disagree with the premise that the lay public should be targeted in campaigns designed to increase the number of stroke patients treated with tPA, given its potential for catastrophic harm. With a potential treatment window of only 3 hours, pressured by patients unversed in the narrow eligibility criteria for tPA treatment who demand treatment all the same, we believe that clinicians unfamiliar with the specific risks and benefits of tPA will not be able to properly educate their patients, obtain unrestricted informed consent, or accurately screen out the majority of stroke patients who are ineligible for therapy. In this scenario, both patients and physicians are misled by promotional pressures to treat as many stroke victims as possible, without equal weight being given to the specific truths that such treatment has a measurable and significant potential for death, and that in actuality only a small minority of stroke victims are eligible for such treatment, given the narrow constraining protocol dictated by NINDS.

James Li, MD, Assistant Professor of Medicine
Larry A. Nathanson, MD, Instructor of Medicine
Division of Emergency Medicine, Harvard Medical School, Cambridge, Massachusetts.

Trevor J. Mills, MD, Assistant Professor
Division of Emergency Medicine, Southwestern Medical Center, University of Texas, Dallas, Texas.

Karin E. Netland, MD, MPH, Clinical Faculty, Emergency Medicine Charity Hospital, LSU School of Medicine, New Orleans, Louisiana.

Richard Paula, MD, Clinical Faculty, Emergency Medicine Tampa General Hospital, University of South Florida, Tampa, Florida.
Douglas Ragland, MD, Clinical Faculty, Emergency Medicine
John C. Lincoln Hospital, Arizona School of Health Sciences, Phoenix, Arizona.

Disclosures: None.

REFERENCES


Each author declares that he or she has not testified as a paid expert witness in a trial involving tPA for stroke use.

Setting the record straight

David Faxon, MD
President, American Heart Association

Send response to journal:
Re: Setting the record straight

24 April 2002

This is to set the record straight on claims made in the article "Alteplase for stroke: money and optimistic claims buttress the 'brain attack' campaign."

In accord with its established policies, the American Heart Association implemented a thorough, multi-leveled process to prevent any individual or special interest from unduly influencing its "Guidelines 2000 for Cardiopulmonary Resuscitation and Emergency Cardiovascular Care." In addition, a formal evidence-based system was used to evaluate scientific data, formulate recommendations and classify final recommendations.

Guidelines development took more than 18 months and involved more than 300 of the world’s leading resuscitation experts. The final published recommendations resulted from countless hours spent by panel committees, international scientific councils, an international editorial board and members of the American Heart Association’s Emergency Cardiovascular Care committee and writing groups.

All those who developed the guidelines adhered to the American Heart Association’s stringent conflict of interest policies and procedures. Participants submitted disclosure forms reporting all relevant relationships with external organizations, groups and companies. Panel and subcommittee chairs reminded participants to report these relationships and if necessary, asked them to refrain from discussion or voting. The relationships reported by panelists are published in the "Proceedings of the Guidelines 2000 Conference," in the Annals of Emergency Medicine.

In its more than 75 years, the American Heart Association has earned its reputation as a trusted authority by remaining an independent, objective leader focused on reducing disability and death from cardiovascular disease.

Sincerely,

David Faxon, MD
President, American Heart Association
Cardiology section chief, University of Chicago

Re: Setting the record straight

Jeanne M. Lenzer
Ellenville, NY 12428 USA

Send response to journal:

24 April 2002

To the editor:

In "Setting the Record Straight" Dr Faxon fails to set anything straight. Protestation aside, he hardly attempts to refute a single point in my article. He does state that "Participants submitted disclosure forms reporting all relevant relationships with external organizations, groups and companies." He fails to mention, however, that the American Heart Association
Re: Re: Setting the Record Straight

(AHA) keeps those "disclosures" secret - refusing to allow the public to know what has been "disclosed." (This is particularly troubling given the information reported in my article.)

I suspect the public would be quite interested to see these disclosures - and challenge the AHA to change its policy so that "disclosure" means public disclosure.

Ultimately, however, as long as public funding is curtailed, many non-profit organizations will be driven into the arms of corporations. Unless we are willing to support a degree of public funding necessary to attain truly disinterested and objective science, potentially critical conflicts of interest of this sort are inevitable.

Jeanne Lenzer,
Medical investigative journalist

Alteplase and Acute Stroke

Scientific debate is the crucible in which to test evolving hypotheses and to evaluate claims for benefit of new treatment strategies. The most fruitful discussions are conducted between informed scientists who may agree or disagree with the new observations and the conclusions. The acceptable milieu is the scientific meeting or the peer-reviewed pages of the leading scientific journals. Chairpersons and editors are obliged to cite reliable written reference to support their claims. They must make every effort to avoid the folly and the pitfalls of hearsay evidence and argumentum ad hominem.

By contrast we were surprised, nay shocked, that the BMJ allowed the observations of a "medical investigative journalist" to lead a discussion of the value of thrombolysis with rt-PA after an acute stroke. Serious stroke neurologists rather would have preferred and expected an ongoing process of high-level scientific enquiry and debate from a journal such as BMJ. Lenzer’s review of the science is coupled with accusations of financial self-interest directed not just at the American Heart Association (AHA) but implies the same of the National Institute of Neurological Disease and Stroke and of all the serious-minded and diligent stroke physicians who took part in an extremely complex and carefully conducted trial. To add weight to the credibility of the insinuations she refers to a list of other studies that may have been biased because they received financial aid from industry. She does not mention the fact that the data from the rt-PA Trial that she alleges to be imperfect were held and analyzed by the independent investigators and the NINDS scientists.1

Attempting to discredit the endorsement by the FDA and others of the NINDS rt-PA trial, Lenzer singles out for emphasis the failure of the Canadian and American Associations of Emergency Medicine to accept the conclusions of their colleagues in stroke neurology. An unknown number of members of these two associations have not been prepared to regard acute stroke as a medical emergency. The named opponent in the American Association, Dr. Jerome Hoffman, dissented from his colleagues on an AHA consensus panel. Lenzer presumes that Hoffman’s survey of the evidence is accurate and that the other members of the panel including many with special expertise in stroke were under financial obligation to Genentech and therefore not to be trusted.

Certainly there are obstacles to be faced when stroke is accepted as an emergency. The American College of Emergency Physicians recognizes this point in their alert endorsement of rt-PA use in acute ischemic stroke (www.acep.org.html). Emergency room physicians must become experts in stroke diagnosis or must arrange that experts are available 24 hours a day in their ER’s. They must have access to immediate and expert CT examination and interpretation. Decisions must be made quickly for the individual patient that thrombolytic treatment is safe, or conversely not indicated for reasons of safety or by the recognition of alternative diagnoses: hemorrhage; tumor; seizure; transient attack, not a stroke. Without the requisite staffing and appropriate organization and skill, rt-PA treatment can be dangerous and is contraindicated. A gap in prerequisite organization is not an argument against the value of rt-PA in reducing stroke disability.
There has been misuse of rt-PA with disastrous results, as in the Cleveland report, when the evolving strict protocol was ignored.2 Contrary to this poor report the increasing experience in a number of well-equipped cities replicating the good results of the NINDS trial are referred in the Saver report and in other publications but have not been considered by Lenzer.3,4,5,6,7 In the reports where success has attended the administration of rt-PA, a major community campaign has been vigorously pursued to educate physicians and healthcare professionals, including ambulance service personnel, in the stringent time requirements, the absolute demand of pre-treatment investigation and the contraindications to the use of rt-PA.

Warlow adds to the innuendo about the flawed nature of the positive rt-PA trials. Guilt by association is implied by his writing about data distortion reportedly carried out in other studies to satisfy industry. His personal opinion is expressed that the results from rt-PA administration were not “convincing enough to change my practice”. The fact that only about 600 patients were included was one of the reasons advanced for the scepticism. Contrariwise only 509 patients symptomatic with severe carotid stenosis in the European Carotid Surgery Trial (ECST) established the efficacy of carotid endarterectomy in stroke prevention. The ECST ethics committee recommended that the trial be stopped because of the very evident benefit. For “severe” patients the North American Symptomatic Carotid Endarterectomy Trial (NASCET) was stopped by its ethics committee at 659 individuals. Large numbers are not what count most in deciding the size or credibility of a randomized trial. It is the frequency and severity of outcome events, coupled with the meticulousness of the data gathering, a minimal loss to follow-up, adherence to the protocol, and scrupulous adjudication of outcome events. Definitive trials may be done with even less than 500 patients. Another expressed concern was inequality in some of the baseline characteristics. Statisticians are capable of adjusting for unexpected imbalance in baseline characteristics. This adjustment was done and clearly reported by the rt-PA investigators.8

Meta-analyses from the randomized trials involving 5216 patients were described as “not particularly convincing of benefit.” This comes as no surprise as these trials included patients whether treated early or later.9 A declining benefit even between 90 minutes and 180 minutes has been noted in an analysis of pooled data from 2776 patients.4 By scrupulous use of the most up-to-date evidence from sophisticated brain imaging a trial of patients between 3 and 6 hours may be justified. It is difficult to dismiss the positive evidence from the 1545 patients with a 3-hour cut-off already submitted to randomized trials and invite patients to submit to placebo in a new study that includes all time-intervals.

A recent report from the Association of British Neurologists concludes that the UK has only 358 consultant neurologists or one per 177,000 population.10 No other European country has fewer than 1 per 38,462 population. The British report concludes that “it would be inappropriate for neurologists to participate in acute general medical intake in emergency rooms.” By contrast in Canada there are almost half as many neurologists with special interest in stroke (150) as the total number of all neurologists in all the various special subdivisions of neurology in the UK. The USA figures are at least of the same proportion as the Canadian. Because the care of acute stroke patients is labour-intensive and demands the attendance of stroke experts, it may be some time before this new treatment strategy will be accepted, let alone implemented in the UK. We wonder if a measure of the scepticism about the value of rt-PA in acute stroke is not prejudiced by the probable inability of the British health-care system as presently constituted to implement the treatment safely.

Saver et al have summarized admirably the evidence favoring the use of thrombolysis in stroke prevention. The speculation of the contrarians writing in the BMJ may deny this strategy to many patients who would benefit. Making a virtue of necessity may delay the marshalling of the professional expertise to bring this modern advance to segments of the medical world. This will be regrettable.
P.B. Gorelick, M.D., Department of Neurologic Sciences, RUSH Medical College, Chicago, Illinois, USA.

B.M. Coull, M.D., Department of Neurology, Arizona Health Sciences Center, Tucson, Arizona, USA.

H.J.M. Barnett, M.D., Professor Emeritus, University of Western Ontario; Scientist, Robarts Research Institute, London, Ontario, Canada

A.M. Buchan, B.M., F.R.C.P., Department of Clinical Neurosciences, University of Calgary and Calgary Regional Health Authority, Calgary, Alberta, Canada

G. del Zoppo, M.D., Department of Molecular & Experimental Medicine, Scripps Clinic & Research Foundation, La Jolla, California, USA.

Please address correspondence to: Philip B. Gorelick, M.D., Department of Neurologic Sciences, RUSH Medical College, 1645 West Jackson Boulevard, Suite 400, Chicago, Illinois 60612 USA
Email: Philip_B_Gorelick@rush.edu

Competing interests: PB Gorelick, MD has received fee payment or reimbursement from Boehringer Ingelheim for speaking engagements, consultation and attending a symposium.

References


Neurologist Attacks Emergency Medicine

Robert C. Solomon, Clinical Assistant Professor of Medicine, Ohio Valley Medical Center, Wheeling, West Virginia, USA 26003

Send response to journal:
Re: Neurologist Attacks Emergency Medicine

Dear Editor:

I could not help being amused by Gorelick's nose-in-the-air dismissal of Lenzer as a mere "investigative journalist" - while at the same time he accuses her of employing argumentum ad hominem. As Lenzer's article did no such thing, I wonder if Gorelick learned Latin (or debate) in the same place where he learned about the capabilities of emergency physicians to diagnose and manage acute stroke as a medical emergency. Oh, wait.... That really WAS ad hominem. Well, whatever fits. And, given the manner in which Gorelick demeans the commitment of emergency physicians to excellent care of stroke patients, it fits.

By the way, his reference to an "alert endorsement" of rt-PA for stroke by the American College of Emergency Physicians is at the same time a distortion of the position taken by the College and an excellent illustration of my concerns that the ACEP policy statement is too easily interpreted as positive - when it was clearly intended to be decidedly neutral.

Robert C. Solomon, MD, FACEP Faculty, Emergency Medicine Residency Ohio Valley Medical Center, Wheeling, WV Clinical Assistant Professor of Medicine West Virginia School of Osteopathic Medicine

rcsmd@nauticom.net

Lies, Damned Lies, and Statistics

Gary J Minter, epidemiologist North Carolina USA Department of Health 27609

Send response to journal:
Re: Lies, Damned Lies, and Statistics

The American novelist and social commentator Mark Twain once said, "There are three types of lies: lies, damned lies, and statistics."

Statistics can be very useful in discovering patterns, or in misleading people who don't understand their use, or who are not trained in critical thinking and analysis.

I have worked for the Epidemiology Division of the North Carolina USA Department of Health and Human Services for ten years, focusing on HIV/AIDS, rabies, certain occupational and environmental issues, and cancer.

During this time I've read thousands of medical and scientific journal articles and abstracts, primarily concerning HIV, but also many on autoimmune disorders and cancer. My most valuable learning occurs outside of work hours, when I go to the NCSU College of Veterinary Medicine Library and read all articles in various journals, including AIDS Research and Human Retroviruses.

I've seen examples of research which is brilliantly done, and seemingly unbiased. I've also read of research which seems to be biased for or against a particular view. Almost all clinical trials are done in accordance with strict procedural standards, but the interpretation and selection of which trials to publish, which results to emphasize, and how to interpret the results gives much "wriggle room."

At a recent conference in Bethesda, Maryland I posed this concern to the editors of both the New England Journal and JAMA. They admitted there are no easy answers, and that there is much judgment involved in peer review. Hopefully, those reviewing an article or a study are familiar enough with the subject matter that problems or inconsistencies can be weeded out. But there are no guarantees.

The same holds true of journalists. There have been cases of fraudulent journalism--not just plagiarism, but invented sources and stories invented by the author's imagination!

Scientists, clinicians, journalists are all human, and we all have personal perspectives, emotions, and the potential for bias, whether deliberate or subconscious.
I tend to be a bit of a purist, and am sometimes overly skeptical and cynical of human motivation. However, in recent decades there has been a definite shift in both research and clinical practice toward "product development" and marketing.

The medical industry--and it is now more of an industry than a "calling"--needs to be extremely vigilant about not only the precision and adherence to research guidelines of a particular clinical trial or study, but also about the motives and the overall context of a particular study.

Generally, the larger the sample size, the better, but valuable insights are often made from a seemingly chance observation or isolated case, as when it was observed that bread mold inhibited growth of bacterial colonies, and penicillin was "discovered." Of course, Mother Nature "developed" penicillin over a billion years or so, but Fleming had the wisdom to make use of it.

I will not comment on the specific issue at hand, because it is not in my field of expertise. Yet I am as qualified as any neurologist to discuss issues of medical and scientific ethics, application of statistics, the role of finance in drug development and marketing, and also the role of advertising in medical journals and the mass media (yes, I once dabbled in politics and published my own small, non-profit issues journal).

In the end, there are many factors which must be considered in researching the safety and efficacy of any treatment. Generally, the more cases which can be considered and fairly compared, the better, because it is more difficult for random factors or bias to influence overall patterns within a huge number of samples.

Full disclosure of conflicts is wise, and sadly has become necessary in these modern times, precisely there are deep and justifiable uncertainties about the motivation, selection, and interpretation of scientific studies due to the immense financial pressures involved in pharmaceutical marketing.

There are other conflicts beyond financial ones, and often I have caught myself being tempted to act in a manner so as to give in to my own predispositions or self-interest. We are all human.

Constant vigilance, questioning, skepticism, and "reality checks" are needed to wash out the inherent biases in all of us flawed mortals.

That is the difference between faith/religion/politics/business, and science: the former are motivated and characterized by emotion, desire, self-interest, or other subjective biases, even though they can make use of technological or scientific methods to advance their goals.

Science and academia should ideally be uninfluenced by motives or goals, and be instead a neutral quest for objective truth, whatever it may turn out to be.

Sincerely,

Gary James Minter
epidemiologist, North Carolina Department of Health and Human Services Raleigh, NC USA 27609
founder and first volunteer, TEST Truth and Ethics in Science and Technology Research Triangle Park, NC, USA
garyjminter@alumni.duke.edu www.minter-g.com

Re: Re: Setting the record straight

14 May 2002
Leaving aside the question of how qualified a non-MD journalist might or might not be to make truly medical judgements in the research context, the link between corporate money and the provision of health care services cannot be overinvestigated, evidenced by the New York Times' recent series on hospital purchasing coalitions and the resulting Congressional hearings.

"Co-optation" of researchers and medical professionals can take many forms, direct and indirect, subtle and overt. The key is not to deny the predictability of corporate behavior--always with an eye on the goal of return on investment--but, rather, to open one's eyes to that behavior and be aware of it, always looking through the fog and flak of whatever spin the co-opters and co-optees may put on an issue to answer the question: what would be best for the patient if the patient was your brother, your sister, your mother your father?

James Unland, Editor Journal of Health Care Finance

**An open letter to the stroke interventionist community**

20 May 2002

The debate about the risk:benefit ratio of tPA therapy in acute ischemic stroke continues.

Read the following dialogue articles in the May 2002 issue of the Western Journal of Medicine.

1) Truth about the NINDS study: setting the record straight - Jeffrey Mann
   Available at http://www.ewjm.com/cgi/content/full/176/3/192

2) Why were the benefits of plasminogen activator (tPA) exaggerated? - Griffin Trotter
   Available at http://www.ewjm.com/cgi/content/full/176/3/194

3) Thrombolysis for acute ischemic stroke: still a treatment for the few by the few - Joanna M Wardlaw, Richard I Lindley, Steff Lewis
   Available at http://www.ewjm.com/cgi/content/full/176/3/198

The NINDS trial is the only RCT that has shown any benefit from tPA therapy and many clinicians have questioned the validity of the NINDS trial for a number of reasons -- that all the other RCTs have failed to demonstrate the efficacy of tPA, that the NINDS trial's results could possibly be due to chance [1], that the NINDS trial was far too small in size, that the NINDS trial's results are not readily transferable to clinical practice because 50% of the patients were treated in < 90 minutes (a level of clinical performance that is unlikely to be achievable in community practice), and that certain post-marketing studies have shown higher rates of secondary ICH in community practice, which may/may not be due to protocol violations. However, despite these various criticisms, no one has suggested that the NINDS trial is internally invalid because of methodological or interpretative flaws.

My personal analysis of the NINDS trial suggests that the NINDS trial, as originally reported in the NEJM, is internally invalid. When I originally wrote my dialogue piece about 9 months ago, I wondered to what degree the imbalance in baseline stroke severity between the tPA-treated patients and the placebo patients in the NINDS trial's 91-180 minutes groups, could account for the favorable results of the NINDS trial -- and I therefore wondered to what degree the NINDS trial could be perceived to be internally invalid. The answer is still not entirely clear to me, and I wonder how clear it will be to the stroke interventionist community after they read my dialogue piece in the wjm.

In the original NEJM paper, the NINDS trialists gave no indication that there was any imbalance in baseline stroke severity between the tPA- treated and placebo patients. They also did not comment on the unexpected fact that the odds ratio (OR) for an excellent stroke outcome was 1.7 (mRS<=1) for the 0-90 minutes groups and 2.4 (mRS<=1) for the 91-180 minutes groups, which suggested that tPA was more effective if given later.
rather than earlier. The NINDS trialists were obviously disconcerted by this finding, and they continued to analyse their data in private (interestingly, although they generated many hypotheses to account for the unexpected finding, it was obviously due to the fact that they did not correct for the imbalance in baseline stroke severity between the treated and placebo patients). Five years later, the NINDS trialists supplied new figures for the ORs based on a re-analysis of their raw data using an unspecified statistical adjustment for baseline NIHSS scores and other baseline variables - in the article by Marler [2]. However, they only supplied a revised OR for the global statistic stroke outcome measure, and they apparently never published revised OR figures for the most commonly used stroke outcome measure - the modified Rankin score (mRS<=1). Without knowing the revised OR for a mRS<=1, how can a stroke interventionist compare the revised results to the original OR for a mRS<=1, or to the excellent stroke outcome results obtained from other stroke trials, which are usually reported as mRS<=1? Without understanding the statistical adjustment that the NINDS trialists used to create their revised OR figures, how can a stroke interventionist determine whether it adequately corrected for the confounding variable of imbalances in baseline stroke severity? According to the TOAST study [3], the graphical curve delineating the natural relationship between baseline stroke severity and the rate of excellent stroke outcome in untreated stroke patients is very steep, and each one point change in baseline NIHSS stroke severity score could cause a 5-10% difference in the absolute rate of an excellent stroke outcome due to the natural course of the disease. Therefore, any statistical adjustment that the NINDS statisticians used to correct for imbalances in baseline stroke severity has to be fully compatible with that fact. It would be extremely useful if the NINDS statisticians publicise the methodology of their statistical adjustment, so that other independent statisticians can judge whether it is compatible with the TOAST graph.

In that same Marler article [2], the NINDS trialists plotted the re-estimated efficacy of tPA against time-to-treatment as a hypothetical model graph-curve (figure 2) and it can be seen that tPA has borderline-positive efficacy after 150 minutes, and the lower boundary of the 95% confidence interval line intersects the "no benefit for tPA" X axis at 150 minutes (see reference number [4] for a hand-drawn copy of figure 2 - the copy is freely viewable online). That model graph-curve suggests that tPA has limited (and equivocal) efficacy after 150 minutes, which would be compatible with Jerome Hoffmann's observation [5] that "the study showed an overall 11-13% absolute benefit with tPA treatment; however, a recent report [2] by the NINDS authors 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) group". Also, the fact that the lower boundary of the 95% confidence interval crosses the "no benefit for tPA" X axis at 150 minutes strongly suggests that the NINDS trial's sample size was too small. Gordon Guyatt states in JAMA's Users' Guides to the Medical Literature [6] "In a positive trial establishing that the effect of treatment is greater than zero, look at the lower boundary of the confidence interval to determine whether the sample size has been adequate. If this lower boundary - the smallest plausible treatment effect compatible with the data - is greater than the smallest difference that you consider important, the sample size is adequate and the trial is definitive. If the lower boundary is less than this smallest important difference, the trial is non-definitive and further trials are required".

Does the stroke interventionist community fully understand how that hypothetical model graph was devised, and whether it is statistically valid and accurate? Should the NINDS trialists have plotted the model graph using the OR for the most frequently used favorable stroke outcome measure - the mRS<=1 (instead of the global statistic)? What would the revised efficacy figures be if the NINDS trialists had used relative risk measurements rather than odds ratio measurements to quantify the benefit of tPA therapy - after adjusting for imbalances in baseline variables?

Barbara Tilley, the primary statistician for the NINDS trial, co-wrote in an article [7] that "odds ratios have been used extensively in data collected both prospectively (for example, a cohort study) or retrospectively (for example, a case control study) when logistic
regression is used to adjust for covariates. However, the odds ratio is less commonly used to communicate the primary results of a clinical trial because of its lack of clinical interpretation. It measures neither a relative size nor an absolute size difference for the treatment effect on the outcome. As an alternative, the relative risk (RR), unlike the odds ratio, measures the relative size of the treatment difference, that is, the ratio of the response probability in a treatment group versus the response probability in a reference group. In that same article, Barbara Tilley re-analysed the NINDS trial's data using a log link model rather than a logit link model, and found that the log link model estimated the benefit of tPA to be 1.32 (relative risk estimate) for part II of the NINDS trial while the logit link model had previously estimated the benefit of tPA to be 1.73 (odds ratio estimate). In fact, Barbara Tilley wrote further "Considering the NINDS t-PA stroke trial, if the analysis plan had specified a goodness-of-fit test to choose the link function, we would have conducted the global test for the treatment effect based on the log link rather than the logit link. We could then interpret the results using an estimate of relative risk instead of the odds ratio, -----". It is refreshing to note that the NINDS statisticians agree that it is better to use relative risk measurements, rather than odds ratio measurements, to determine the efficacy of tPA, and it is interesting to see how much lower the RR figure is than the OR figure. Although I am very sceptical of the accuracy of the RR figure of 1.32 (because the NINDS statisticians gave no indication that major statistical adjustments were needed to correct for imbalances in baseline stroke severity between the treated and placebo patients) note that the RR figure of 1.32 was for the entire time span of 0-180 minutes. Can you imagine, after examining the graph in reference number 4, how low the RR figure must be for patients treated between 91-180 minutes, and particularly between 150-180 minutes? The relative benefit of tPA for patients treated in the 91-180 time period may only be a fraction of the benefit experienced by patients treated in the 0-90 time period (the estimation of relative benefit is based on the simple awareness that most of the patients treated in the 0-90 minute time period were treated between 60-90 minutes and the estimated odds ratio for a favorable stroke outcome is 2.8-3.8 for that time period, while the estimated odds ratio for the 91-180 time period is between 1.3-2.8, and most of those patients were actually treated between 150-180 minutes and the estimated odds ratio was 1.3-1.8 for the 150-180 time period -- note that all of the estimated odds ratio figures are derived from the Marler graph (reference number [4]).

If one cannot precisely determine the adjusted RRR for patients treated between 91-180 minutes, then one cannot precisely calculate the absolute risk reduction (risk difference) for those patients. Barbara Tilley wrote [7] a "risk difference (RD) is the difference in response probabilities between two groups, which is sometimes considered clinically more interpretable than the relative risk, because it indicates an absolute but not relative treatment size difference in response probability. It is an intuitively appealing measure of treatment efficacy in clinical trials. The risk difference is useful in that it provides an estimated amount by which a particular response might be increased or reduced if a specified treatment is removed, and is a particularly important concept when treatment benefits are offset by side-effects and/or by a high cost." In other words, Barbara Tilley is suggesting that one needs to know the absolute benefit of tPA therapy (RD) in order to calculate a risk:benefit ratio for tPA therapy. I agree - but how can a stroke interventionist determine the risk:benefit ratio of tPA therapy for stroke patients treated between 91-180 minutes, assuming that the absolute risk of a major side-effect (symptomatic ICH) is approximately 6%, if he does not precisely know the absolute benefit of tPA therapy for those same patients (after having to make precise statistical adjustments to correct for imbalances in baseline stroke severity between treated and placebo patients)? Also, any absolute benefit figure that is used for the 91-180 minute time span would only be an "average" figure and one really needs to know the absolute benefit for different time spans throughout that 90 minute time period (eg. 90-120 minutes, 120-150 minutes, 150-180 minutes) because the efficacy of tPA continuously wanes throughout the 91-180 minute time period.(To better appreciate the varying relationship between RRR and the NNT with respect to variations in time-to-treatment and/or variations in baseline stroke severity, see reference number [8] which is freely available online)
Should stroke interventionists be obliged to offer the stroke patient more precise information about the risks and benefit of tPA therapy - with particular attention paid to natural variations in baseline stroke severity and "real life" variations in time-to-treatment?

For instance, what are the answers to the following three questions with reference to 4 hypothetical stroke patients?


Question 1: What is the likelihood of an excellent stroke outcome (mRS<=1) without any tPA therapy for those 4 patients?

Question 2: What is the likely benefit of tPA therapy for those 4 patients if they are treated in the following time frames = < 90 minutes, 90-120 minutes, 120-150 minutes and 150-180 minutes?

Question 3: What is the likely secondary ICH rate for those 4 patients if they are treated in the following time frames = < 90 minutes, 90-120 minutes, 120-150 minutes and 150-180 minutes?

How do you answer question number 1?

I believe that the answer to question number 1 has to take the natural course of the disease into precise account, and it is well known that patients with mild strokes have a much better rate of excellent stroke outcome than patients with severe strokes (due to the natural course of the disease).

My "best guess" answer to question number 1 is:

Expected rate of excellent stroke outcome for patient number 1 (baseline NIHSS score of 8) = 56%. Expected rate of excellent stroke outcome for patient number 2 (baseline NIHSS score of 12) = 42%. Expected rate of excellent stroke outcome for patient number 3 (baseline NIHSS score of 16) = 23%. Expected rate of excellent stroke outcome for patient number 4 (baseline NIHSS score of 20) = 13%.

The "expected" rate of excellent stroke outcome numbers come from the TOAST graph curve [3]. Do you use different "expected" figures? From where do you obtain your "expected" figures?

I believe that a stroke interventionist should be able to supply a stroke patient with a numerical figure for the expected rate of excellent stroke outcome - derived from the evidence-based medical literature - so that the individual stroke patient can estimate his likelihood of an excellent stroke recovery (based on the natural course of the disease) before he tries to establish whether tPA therapy (or any other stroke therapy) can offer him a better rate of recovery. If the stroke interventionist community does not agree with the estimated figures from the TOAST graph, then why does it not perform a prospective study on 10,000 acute ischemic stroke patients (untreated) and measure their baseline NIHSS stroke severity scores and their rate of excellent stroke outcome at 3 months? It would then be possible to draw a TOAST-like graph showing the precise relationship between the baseline NIHSS score and the rate of excellent stroke outcome without having to use a logistic regression equation to draw a "best-fit" graph (because the graph would be plotted from actual baseline NIHSS scores for each level of baseline stroke severity from a NIHSS score of 1-25). By plotting that graph, the stroke interventionist community would have established a "gold standard" curve that would allow it to determine the true benefit of tPA therapy (or any other stroke therapy) - by comparing the individual treated patient's results to the "gold standard" graph.

How do you answer question number 2?
I believe, that in the absence of reliable results from a RCT that is perfectly randomized for baseline stroke severity, that one can only answer question 2 accurately if one plots the results of tPA therapy for different treatment times (<90 minutes, 90-120 minutes, 120-150 minutes, 150-180 minutes) for each level of baseline stroke severity (from a NIHSS score of 1-25) on top of that "gold standard" graph, so that an automatic correction is made for imbalances in baseline stroke severity. If the tPA-treated patient's rate of excellent stroke outcome figure is higher than the untreated patient's figure (for each baseline NIHSS score), then tPA obviously works, and the degree of efficacy will be obvious by seeing how much higher the treated patient's figure is above the untreated patient's "gold standard" figure - for varying baseline stroke severity scores and varying times-to-treatment.

How do you answer question number 3?

I believe that one can really only answer question 3 accurately if the stroke interventionist community plots the secondary ICH rate results similarly - by plotting the secondary ICH rate against the baseline NIHSS score and by plotting 4 sets of curves for the different treatment times - because it is well-known that the secondary ICH rate depends on both the baseline stroke severity and the degree of delay in administering tPA therapy. Kidwell [9] showed that the secondary ICH rate increased markedly with baseline NIHSS scores > 10, but their published results are of limited value because the study sample size was too small, because they used IA tPA, and because they used a subgroup analysis with subgroups that are too broad (see reference number [8] for a handpainted copy of their diagram showing the hemorrhagic transformation results - it is freely viewable online). I think that the ICH rate needs to be plotted for each NIHSS score between 1-25 (and not from limited subgroup data, which is too crudely inaccurate) and for varying treatment times. The result-analysis may show that patients with severe strokes (NIHSS score > 15), who are treated later ( > 150 minutes after stroke onset), have much higher secondary ICH rates and that the risk:benefit ratio of tPA therapy for those patients is significantly greater than 1.0.

The stroke interventionist community already has enough pooled data from a number of stroke trials [10] to be able to plot those curves.

By using the pooled data to plot those curves, the variable of imbalances in baseline stroke severity will no longer have to be considered a confounding variable, and the true risk:benefit ratio of tPA therapy will immediately become apparent -- by examining the excellent stroke outcome rate and the secondary ICH rate for each baseline NIHSS score for a variety of treatment times (< 90 minutes, 90-120 minutes, 120-150 minutes, 150-180 minutes).

"Statistics are like a bikini: what they reveal is suggestive, but what they conceal is vital" - Aaron Levenstein.

I think that the stroke research community needs to immediately make the pooled raw data from all the tPA-for-stroke trials publically available, so that the vital truth about tPA can be revealed.

The vital truth regarding tPA's efficacy in acute ischemic stroke cannot be discovered by examining the result-analyses of the major tPA-for-stroke RCTs (NINDS, ECASS, ECASS II, ATLANTIS) as they were originally reported in the medical literature, because they did not address the confounding effect of the most critical prognostic variable - imbalances in baseline stroke severity between the tPA-treated and placebo patients. Even more recent post hoc re-analyses of those trials, such as the re-analysis of the ATLANTIS trial's < 3 hours patient group [11] and the meta-analysis of the NINDS, ECASS, ECASS II and ATLANTIS trials [12] make no attempt to correct for imbalances in stroke severity between the treated and placebo patients, thereby rendering their conclusions moot.

I think that if the stroke interventionist community decides that the NINDS trial was too
small in size and too poorly randomized, and that arbitrary, and inordinately complex, post hoc statistical adjustments cannot accurately correct for the marked imbalance in baseline stroke severity in the 91-180 minutes groups -- then the need to perform a much larger tPA-for-stroke RCT, that is precisely balanced for critical prognostic variables like baseline stroke severity, becomes imperative. The stroke research community may decide to actively debate the issues that I have discussed, but if genuine "uncertainty" still persists, then the need to perform a new RCT will likely remain -- because it makes no sense to base an entire "brain attack" industry on the results of a single clinical trial, if its legitimacy remains forever controversial and doubtful. In the interim period, the stroke interventionist community could debate what to do with the presently accumulated raw data, and it could also debate the advisability, and correct methodology, of analysing the pooled raw data from multiple RCTs and observational studies (as I have just described).

Finally, I think that stroke researchers, who deem themselves tPA-experts, should actively engage tPA-contrarians and tPA-sceptics in "open" scientific debate regarding the utility of tPA in acute ischemic stroke. The tPA-experts should specifically avoid adopting an "all-knowing" attitude as exemplified by the statement made by James Grotta in his editorial in the February 2002 issue of the Stroke Interventionalists' publication [13], when he wrote "The data supporting the efficacy of thrombolysis within the first 3 hours after stroke onset in patients who qualify, as nicely reviewed by Dr Pancioli, is now unanimously endorsed by those who know best - ie. those Neurologists and Emergency Physicians who have made the commitment to take care of acute stroke patients for a living and are experienced in using thrombolytics." The implication of that statement seems to be that other neurologists and emergency physicians, who do not actively participate in stroke research, are incapable of accurately analysing the evidence-based medical literature, and they should remain passively dependent on the subjective judgments of "those who know best". The stroke experts are obviously at the cutting edge of stroke research, and they are the people most keenly aware of potentially productive research endeavours in the area of biological and mechanical thrombolysis. However, I don't agree that stroke experts are automatically more adept at judiciously assessing the "weight" of the published tPA-for-stroke medical literature, and that they will automatically come to a more rational and better balanced conclusion. It is my personal belief that any conscientious physician, who has acquired a modicum of evidence-based medicine skills, can painstakingly review the tPA-for-stroke medical literature and come to independent judgments, that may be closer to the vital truth than the biased judgments of some of the self-appointed tPA-experts. History, of course, will be the final arbiter, and the future will determine who is closer to realizing the vital truth about tPA therapy in acute ischemic stroke -- the tPA-experts, who have a fixed set of a priori biases, and who are therefore absolutely convinced that they know best, or the tPA-sceptics who are likely to remain more open-minded about the issues.

Jeff Mann.

References:


We thank BMJ readers who have posted responses to our commentary and are grateful for the opportunity to address them.

Dr. Jeffrey Mann wildly overstates his case. In theory, it is best to adjust for baseline variables based on pretrial knowledge of influence on outcome. In practice, sufficiently detailed knowledge regarding the influence of baseline variables on clinical outcome in populations precisely resembling those being enrolled in a clinical trial is almost never available. As a result, virtually all clinical trials adjust for influences actually observed in the enrolled control population, as a cursory glance at trials published in leading medical journals, including the BMJ itself, will attest. Moreover, it is scientifically unsound to adjust for baseline imbalances using estimates of influence upon clinical outcome that are known to be inappropriate for the enrolled clinical trial population. Yet this well-recognized error is just what Mann commits in his posting and in his Western Journal of Medicine essay. Because spontaneous improvement and worsening frequently occur early after stroke onset, NIHSS stroke scores 12-24 hours after stroke onset are expected to correlate much more substantially with final clinical outcome than NIHSS stroke scores 1-3 hours after stroke onset. [1] The TOAST trial population was enrolled up to 24 hours after stroke onset, with virtually no patients enrolled within 3 hours of onset. It is self-evidently an inappropriate population to use in estimating influence upon outcome of baseline NIHSS scores in both NINDS TPA Trial 1 and NINDS TPA Trial 2, each enrolling only under 3 hour patients. The two NINDS TPA trials were the first studies ever to characterize well the course of a large population of under 3 hour stroke patients. In such a case, it is solid evidence based medicine, indeed obligatory EBM, to adjust for influences of baseline
variables using effects observed in the control group actually enrolled in the clinical trial.

Dr. Mann appears generally insensitive to important population distinctions between stroke studies. In a discussion regarding the hemorrhagic risks of intravenous TPA given within 3 hours of onset, he invokes our report on predictors of hemorrhage in patients given intra-arterial thrombolysis within 6 hours of onset. In fact, most patients in our recent analysis were treated with intra-arterial thrombolysis because they were NOT eligible for intravenous tPA. It is inappropriate to extend findings from a different procedure with a different time window and different baseline characteristics to the NINDS trials.

Dr. Solomon’s posting contains much opinion, but little actual analysis. He wishes we were wrong, but is unable to demonstrate that we are. Two glaring errors in his post are worth pointing out. First, his statements regarding the “number needed to harm” with TPA are completely misleading (as are those of Li and colleagues). TPA causes more patients to bleed – the number needed to treat to cause symptomatic intracerebral hemorrhage does approximate 17. However, TPA also prevents an approximately equal number of patients from experiencing symptomatic worsening from stroke extension, cerebral herniation, and other complications of large infarcts. Baldly stated, if you receive TPA, the risk is increased that you may bleed and die. If you don’t receive TPA, the risk is increased that you may herniate and die. The salient number needed to harm is the net sum of these two factors, and across all under 3 hour trials, there is no net harm.

A second misrepresentation advanced by Dr. Solomon is his suggestion that in many communities only a small minority of stroke patients present via ambulance. We are unaware of any published data that would support such a claim, and he fails to reference any. In fact, published studies suggest just the opposite, that 35-70% of all acute stroke patients are transported to the emergency department by emergency medical services.[2] This number is likely to be even higher for patients who are candidates for thrombolytic therapy.

Our UCLA colleague, Dr. Hoffman, makes the rather startling suggestion that it was unfair of us to refute the incorrect scientific statements in Ms. Lenzer’s article because Ms. Lenzer is not a scientist. We could not disagree more. Writings of both journalists and scientists should be held to one standard—the truth. We are sure Ms. Lenzer would agree, pace Hoffman. Dr. Hoffman also wishes the BMJ had solicited commentaries more accurately reflecting the balance of opinion on TPA in acute stroke. So do we. It is important, however, to realize what such balanced commentaries would look like. The typical opinion page dichotomy of one pro opinion and one con opinion would give an entirely misleading view of the state of informed opinion. Among American stroke experts, there is overwhelming consensus that TPA is efficacious. A balanced set of expert commentaries would include an order of magnitude greater number of positive opinions than negative opinions. We join with Dr. Hoffman in urging the BMJ to solicit such a representative set of publications. Lastly, Dr. Hoffman erects a straw man. We never suggested that “it is impossible to be an expert on a subject about which one does not have a conflict of interest.” Rather, we merely suggested that many experts will have minor competing interests. In this regard, it is ironic that Dr. Hoffman has modified his posting since it originally appeared, adding a notice of his own financial competing interest with regard to the use of TPA in stroke that was not included originally. If he were to be true to the absolutist position on financial conflicts that he has advanced, he will now absent himself from advising influential and independent organizations on the TPA in stroke issue.

Contrary to Li and colleagues, we did not count the NINDS TPA trials twice. We counted NINDS TPA Trial 1 once and NINDS TPA Trial 2 once. Clearly they would prefer that there had been only one NINDS TPA trial, but wishing does not make it so. We are mystified by the continued misreading of the NINDS-TPA Trials NEJM report as representing a single trial.[3] The first way you can tell that there were two trials in the study is by their names. The first trial was named Trial 1. The second trial was named Trial 2. Trial 1 had as its primary prespecified endpoint early improvement at 24 hours by 4 points or more on the NIHSS. It narrowly missed reaching statistical significance on this endpoint, but several
final 3 month clinical endpoints were positive. Trial 2 was launched after the completion of enrollment in Trial 1, enrolled a completely new set of patients, and was analyzed separately with regard to its prespecified primary endpoint, a global measure of final functional outcome 3 months after stroke. Trial 2 was positive on this prespecified primary endpoint. However devoutly the TPA contrarians wish they only had a single NINDS-TPA trial with which to contend, the fact is there were two trials, as the FDA recognized when ascertaining that the evidence for the benefit of TPA in acute stroke was quite adequate to approve the indication. Li and colleagues also claim that our pooled analysis of under 3 hour data from intravenous TPA trials is incorrect. Once again, this is mere assertion—they advance no actual argument—and it is wrong. In contrast, we can easily point out how the meta-analysis in the table they provide is itself “contrary to meta-analysis methodology and confounds rather than promotes the truth.” Their table includes only the 4 trials with higher mortality in the treatment group than in the placebo group and leaves out entirely the 3 trials with lower mortality in the treatment group than in the placebo group. This selective inclusion in a meta-analysis of only trials with data favorable to one’s argument is a fundamental violation of meta-analytic methodology. Across all seven trials with available data (NINDS 1 and 2, ECASS 1 and 2, ATLANTIS A and B, Haley 1993), death occurred in 83/479 (17.3%) of TPA treated patients and 83/478 (17.4%) of placebo treated patients (p=0.9). Thus, the number needed to treat to produce benefit from TPA is as low as 2, the number needed to treat to cause net harm approaches infinity. These numbers amply support the statement that TPA is highly efficacious.

Once again, we will close trying to find common ground. We concur, as before, with calls to bar experts with major financial competing interests from service on guideline committees and to require experts with minor financial competing interests to disclose them publicly. We also concur with calls to make public detailed raw data from the pivotal NINDS-TPA trials and from the pooled analysis of all 6 major intravenous TPA trials soon to be published. We are confident that the effect of TPA therapy, used rightly, is robust and will stand up to detailed scrutiny. Lastly, we reiterate our concurrence with policy statements of the Brain Attack Coalition and the American College of Emergency Physicians that urge emergency physicians and other acute care providers to become expert in acute stroke care, including the use of TPA for acute stroke, or to place their hospitals on standby and divert patients to designated stroke centers where therapy can be expertly delivered.[4,5]


**Ongoing scientific debate allows the scientific truth to be more clearly established**

Jeffrey Mann, Physician

In his commentary [1] on Jeanne Lenzer's article, Philip Gorelick stated "Scientific debate is..."
the crucible in which to test evolving hypotheses and to evaluate claims for benefit of new treatment strategies. The most fruitful discussions are conducted between informed scientists who may agree or disagree with the new observations and the conclusions. The acceptable milieu is the scientific meeting or the peer-reviewed pages of the leading scientific journals." I agree with that statement, but I also feel that online discussions between informed scientists can supplement those traditional venues as a serious forum for scientific debate. Critics, who are not personally involved in a particular area of research, often do not attend scientific meetings frequented by super-specialists, and that phenomenon robs those super-specialists of the opportunity to become aware of contrasting points-of-view. The rapid response section of the bmj is a wonderful idea, because it allows for an ongoing debate in the "open" forum of public space, thus giving interested readers the opportunity to view the facts from two (or more) conflicting points-of-view.

I applaud Dr. Jeffrey Saver for taking the trouble to reply to my criticism of his views [2] in his recent rapid response letter to the bmj [3]. I think that it is eminently fair that Dr. Saver be given the opportunity to reply to his critics, because it gives interested readers an opportunity to hear both sides of an argument. Being able to view "reality" through the prism of contrasting viewpoints gives the reader an opportunity to more clearly discern the "scientific truth". However, the subject of tPA-for-stroke trial methodology is very complicated and many readers may not be able to clearly appreciate the nuances of the debate without further explication. It is obvious that each countering reply presents the best face on the author's argument -- and subtle, but important, facts may not be obvious to the uniformed reader without further explication.

Dr. Saver states [3] "Dr. Jeffrey Mann wildly overstates his case. In theory, it is best to adjust for baseline variables based on pretrial knowledge of influence on outcome. In practice, sufficiently detailed knowledge regarding the influence of baseline variables on clinical outcome in populations precisely resembling those being enrolled in a clinical trial is almost never available. As a result, virtually all clinical trials adjust for influences actually observed in the enrolled control population, as a cursory glance at trials published in leading medical journals, including the BMJ itself, will attest. Moreover, it is scientifically unsound to adjust for baseline imbalances using estimates of influence upon clinical outcome that are known to be inappropriate for the enrolled clinical trial population. Yet this well-recognized error is just what Mann commits in his posting and in his Western Journal of Medicine essay." I agree with Dr. Saver that sufficiently detailed, and appropriate, knowledge regarding the influence of baseline variables on clinical outcome is not always available and that trialists therefore have to rely on a less optimum approach -- using a post hoc statistical adjustment to correct for the most critically important baseline prognostic variables. Dr. Saver states further "it is solid evidence based medicine, indeed obligatory EBM, to adjust for influences of baseline variables using effects observed in the control group actually enrolled in the clinical trial." If that EBM practice is indeed obligatory - and I fully agree that it is - then why did the NINDS trialists not make the obligatory adjustment in their original analysis of the NINDS trial, as initially reported in the NEJM in 1995? Also, why did all the other tPA-for-stroke RCTs (ECASS, ECASS II, ATLANTIS) fail to adjust for the influence of baseline stroke severity variations when interpreting their trial's results? Even more recent re-analyses in the medical literature, such as Gregory Albers' review of the ATLANTIS trial's < 3 hours patient group [4] and Marc Fisher's meta-analysis of the NINDS, ECASS, ECASS II and ATLANTIS trials [5] make no attempt to correct for imbalances in baseline stroke severity between treated and placebo patients when analysing the raw data. Has the stroke research community been scientifically delinquent -- if it has willfully ignored the confounding variable of "variations in baseline stroke severity" when analysing the raw data of tPA-for-stroke trials? What is wilder -- a total failure of clinical trialists to make any statistical adjustment for a critical prognostic variable (baseline stroke severity variations in tPA-for-stroke trials), or the imperfect use of a statistical adjustment tool to better understand the implications of that failure?

Is it essential to correct for imbalances in baseline stroke severity in tPA-for-stroke trials?
My dialogue essay published in the wjm [6] suggests that it is critically important because of the steep slope angle of the graph curve relating baseline stroke severity to the rate of excellent stroke outcome in untreated patients -- each one point change in baseline NIHSS stroke severity score could cause a 5-10% difference in the absolute rate of an excellent stroke outcome due to the natural course of the disease. If the slope-angle of the graph curve wasn't that steep, then the interpretative-error that would occur if that prognostic variable was ignored would not be that significant. However, baseline stroke severity has a major impact on the final prognosis, and any stroke severity imbalances between treated and placebo patients must be corrected for in ALL stroke trials.

Dr. Saver argues that my use of the TOAST graph is invalid. His reasoning is: "Because spontaneous improvement and worsening frequently occur early after stroke onset, NIHSS stroke scores 12-24 hours after stroke onset are expected to correlate much more substantially with final clinical outcome than NIHSS stroke scores 1-3 hours after stroke onset.[1] The TOAST trial population was enrolled up to 24 hours after stroke onset, with virtually no patients enrolled within 3 hours of onset. It is self-evidently an inappropriate population to use in estimating influence upon outcome of baseline NIHSS scores in both NINDS TPA Trial 1 and NINDS TPA Trial 2, each enrolling only under 3 hour patients." I agree with Dr. Saver -- the fact that there are variations in baseline stroke severity measurements, depending on how soon after stroke onset the measurements are made, means that the TOAST graph cannot be utilized as an absolute "gold standard" measuring stick when it comes to making accurate predictions about final stroke outcome. However, Dr. Saver fails to mention that I fully acknowledged that fact in my open letter to the stroke interventionist community [7]. I specifically acknowledged that any information derived from the TOAST graph only allows for a "best guess" estimate, and that it cannot necessarily serve as a "gold standard" measurement. In that rapid response letter, I stated "If the stroke interventionist community does not agree with the estimated figures from the TOAST graph, then why does it not perform a prospective study on 10,000 acute ischemic stroke patients (untreated) and measure their baseline NIHSS stroke severity scores and their rate of excellent stroke outcome at 3 months? It would then be possible to draw a TOAST-like graph showing the precise relationship between the baseline NIHSS score and the rate of excellent stroke outcome without having to use a logistic regression equation to draw a "best-fit" graph (because the graph would be plotted from actual baseline NIHSS scores for each level of baseline stroke severity from a NIHSS score of 1-25). By plotting that graph, the stroke interventionist community would have established a "gold standard" curve that would allow it to determine the true benefit of tPA therapy (or any other stroke therapy) -- by comparing the individual treated patient's results to the "gold standard" graph." In other words, I fully acknowledged the fact that the TOAST graph can only be used as a "relative" comparison, and I have already suggested that the stroke research community should debate the issue further and "standardize" the required measurements by appropriate means. The stroke research community can also debate the time-point issue further and decide at what time-point after an acute stroke, the baseline stroke severity measurement should be made. However, readers should not miss the essential point that I was trying to make -- that using information gleaned from the TOAST graph (which is only "relatively" accurate) is probably a "second- best" solution, that should only be necessary if stroke RCTs cannot perfectly randomize the treated and placebo patients for baseline stroke severity, and if the stroke research community has not established a "gold standard" post hoc statistical method of accurately adjusting for imbalances in baseline stroke severity, that is acceptable to the entire international stroke research community.

Is there a "gold standard" statistical adjustment that should be applied post hoc to accurately correct for imbalances in baseline stroke severity between treated and placebo patients in tPA-for-stroke trials? Has the entire stroke research community debated all the alternative methods of making that statistical adjustment, and have they jointly decided on the optimum methodology? Are stroke interventionists even aware of the statistical methodology that was used by the NINDS trialists to correct for baseline stroke severity imbalances in the NINDS trial, and do they agree on its appropriateness and accuracy? Dr.
Saver did not mention the fact that the only tPA-for-stroke trial that has made any post hoc statistical correction for stroke severity imbalances was the NINDS trial -- and that very important fact was only reported in the medical literature [8] for the first time 5 years after the NINDS trial's results were initially reported in the NEJM. What are the implications of that post hoc statistical correction on the "accurate" interpretation of the NINDS trial's raw data? Dr. Saver has skirted that issue completely. In my open letter [7] I analysed the implications of the NINDS trialist's post hoc statistical adjustment - accepting, for argument sake, their published figures - and I estimated that it implies that tPA has marginal, and equivocal, efficacy approximately > 150-180 minutes after stroke onset. That estimation was based on a "guessedestimated" relative risk reduction (RRR) for the 150-180 minute time period that must be much lower than 1.32 (the RRR figure for the entire 0-180 time period). Does Dr. Saver know what's the RRR of tPA for stroke patients treated between 150-180 minutes after stroke onset in the NINDS trial -- after making the appropriate statistical correction for imbalances in baseline stroke severity?

Dr. Saver also stated "Dr. Mann appears generally insensitive to important population distinctions between stroke studies. In a discussion regarding the hemorrhagic risks of intravenous TPA given within 3 hours of onset, he invokes our report on predictors of hemorrhage in patients given intra- arterial thrombolysis within 6 hours of onset. In fact, most patients in our recent analysis were treated with intra-arterial thrombolysis because they were NOT eligible for intravenous tPA. It is inappropriate to extend findings from a different procedure with a different time window and different baseline characteristics to the NINDS trials." If interested readers read my actual words from the open letter [7] they will note that I stated "Kidwell showed that the secondary ICH rate increased markedly with baseline NIHSS scores > 10, but their published results are of limited value because the study sample size was too small, because they used IA tPA, and because they used a subgroup analysis with subgroups that are too broad." In other words, I readily acknowledged that one could not accurately extrapolate their results because they used intra-arterial tPA and not intravenous tPA (among other reasons). However, does Dr. Saver argue that the results for intravenous tPA show a markedly different pattern of secondary ICH compared to intra- arterial tPA? Consider a recent article by Tanne [9] and note in figure 1 (risk of all ICH by baseline stroke severity) that the rate of secondary ICH goes up dramatically with baseline NIHSS stroke severity levels > 10. The only point that I was trying to make is that one has to carefully consider variations in baseline stroke severity when trying to estimate the likelihood of a secondary ICH -- just as one needs to take variations in baseline stroke severity, and variations in time-to-treatment, into precise account when judging the likely efficacy of tPA. How else can one accurately predict the absolute risk:benefit ratio of tPA therapy when treating different acute ischemic stroke patients with strokes of varying stroke severity - at varying times-to-treatment?

Finally, Dr. Saver stated "We also concur with calls to make public detailed raw data from the pivotal NINDS-TPA trials and from the pooled analysis of all 6 major intravenous TPA trials soon to be published." On this point, there is absolutely no contention, and I fully support the call to make public the raw (patient-level) data from the NINDS and other intravenous tPA trials.

Jeff Mann.

References:


5. Fisher, Marc MD; Ringleb, P. A. MD; Schellinger, P. D. MD; Schranz, C. MD; Hacke, W. MD. Thrombolytic Therapy Within 3 to 6 Hours After Onset of Ischemic Stroke: Useful or Harmful? Stroke. 33(5):1437-1441, May 2002.


The NINDS Stroke Study Group Response
2 July 2002

The NINDS investigators would like to respond to the recent articles in the British Medical Journal (1) and Western Journal of Medicine (2-4). Please excuse the necessary length of our combined reply.

First Dr. Mann’s comments (2). We also encourage physicians to think independently and would like to provide further information in response to Dr. Mann’s four main points. This is done in the spirit of discussion and not debate. We hope the readers of the BMJ and WJM will take time to consider our response. We thought our previous peer-reviewed reports had addressed Dr. Mann’s questions.

Dr. Mann writes that in order to obtain valid results, critical prognostic variables have to be prespecified, and corrected for, in the design of any randomized controlled trial. This is true, but Dr. Mann’s choice of prognostic variables is different than ours. What we think is important is predicting who will respond to TPA. Dr. Mann is most concerned with who will do well despite therapy. There is good evidence from the NINDS study that over a broad range of the baseline NIH stroke scale, patients treated with TPA do better than those who are not treated with TPA. Contrary to Dr. Mann’s assertion, NIH Stroke Scale was a prespecified variable that was known to predict outcome and it was corrected for in the usual way in the original publication (5). The data in the two trials composing the study confirm that NIHSS does not reliably predict response to therapy (6). The only group in which the benefit is not apparent without correcting for other variables is the group with very mild strokes having 0-5 on their baseline NIHSS (9% of the patients in the study, 7% <90 minutes, 11% >90 minutes). In this case, there are few patients to evaluate and the outcome variables may not have been sensitive to different degrees of minimal disability. As Dr. Mann points out, almost all of these patients had minimal or no disability at three months whether or not they were treated with TPA. Even though it makes the argument more complicated, we would also like to point out that there are other baseline variables that predict outcome. Age of the patient is just one example.

Dr. Mann writes that randomization into the TPA and placebo groups was flawed in the
National Institute of Neurological Disorders and Stroke (NINDS) trial. There is no evidence that the randomization process was flawed. The result of the randomization, however, was not an equal assignment of baseline stroke scale in every small subgroup. This happens in all trials. Fortunately, as you will see below, the effect of the drug is so large that it overpowers these imbalances. In the 91-180 minute group, the average baseline NIHSS score was lower for the patients assigned randomly (in a process that was not flawed) to the placebo group. We feel that our published analyses did account for this imbalance in our assertion that the drug reduced disability at three months. We hope that the data provided here will make the validity of the statistics more clear.

We understand how Dr. Mann could at first glance suspect that this imbalance in randomization could alone account for the apparent effectiveness of TPA shown in the NINDS trial. We had recommended as a group that there be a goal of one hour door to needle time because it was so apparent and logical to us that treating early was critical (7). Several years after the results of the trial were published, though, it was clear that the door to needle time varied widely and that physicians were delaying treatment of patients even within the three hour window. Therefore we looked more thoroughly to see if time to treatment predicted a better response to treatment. In our post-hoc analysis (8), it did appear that door to needle time was important even within three hours from stroke onset. What we failed to make clear was that the baseline difference in stroke scale in the 91-180 minute group did NOT account for the effectiveness of TPA in the entire trial. In other words, there were not so many more patients in the TPA treated group with baseline NIHSS 0-5 that were treated 91-180 minutes that their predictably good outcome could have explained the entire effect of the drug. To make this as clear as possible, we provide the following data tables with the qualification that they do not account for other significant variables that predict outcome such as age. It is also very important to remember that these post-hoc analyses involving subgroups without sufficient statistical power to answer the meaningful question should be considered as “hypothesis generating” and providing a rationale for further study only (9). The facts are that the baseline imbalance of stroke severity DOES NOT explain the entire results of the trial. The imbalance explains some of the difference, but there is no question that there is still benefit from TPA 91-180 minutes after stroke onset. Until now we had thought that this benefit had been made clear in our previous publications (5,6,8).

Table 1 Rankin (Good outcome = 0, 1) at 3 months for patients treated 91-180 minutes from stroke onset

<table>
<thead>
<tr>
<th>Baseline NIHSS (Patients treated 91 to 180 minutes)</th>
<th>Patients with Rankin Good Outcome (0,1) at Three months</th>
<th>Relative Risk (95%CI)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>TPA</td>
<td>Placebo</td>
</tr>
<tr>
<td>1-5</td>
<td>24/29 (83%)</td>
<td>6/7 (86%)</td>
</tr>
<tr>
<td>6-10</td>
<td>23/37 (62%)</td>
<td>23/46 (50%)</td>
</tr>
<tr>
<td>11-15</td>
<td>10/26 (38%)</td>
<td>5/35 (14%)</td>
</tr>
<tr>
<td>16-20</td>
<td>9/33 (27%)</td>
<td>6/33 (18%)</td>
</tr>
<tr>
<td>&gt;20</td>
<td>4/28 (14%)</td>
<td>2/46 (4%)</td>
</tr>
<tr>
<td>All Patients</td>
<td>70/153 (46%)</td>
<td>42/167 (25%)</td>
</tr>
</tbody>
</table>
Table 2 Symptomatic hemorrhage within 36 hours of treatment for patients treated 91-180 minutes from stroke onset

<table>
<thead>
<tr>
<th>Baseline NIHSS</th>
<th>Patients with Symptomatic Hemorrhage</th>
<th>Relative Risk (95%CI)</th>
</tr>
</thead>
<tbody>
<tr>
<td>(Patients treated 91 to 180 minutes)</td>
<td>TPA</td>
<td>Placebo</td>
</tr>
<tr>
<td>1-5</td>
<td>0/29 (0%)</td>
<td>0/7 (0%)</td>
</tr>
<tr>
<td>6-10</td>
<td>2/37 (5%)</td>
<td>1/46 (2%)</td>
</tr>
<tr>
<td>11-15</td>
<td>2/26 (8%)</td>
<td>0/35 (0%)</td>
</tr>
<tr>
<td>16-20</td>
<td>2/32 (6%)</td>
<td>1/33 (3%)</td>
</tr>
<tr>
<td>&gt;20</td>
<td>4/28 (14%)</td>
<td>0/46 (0%)</td>
</tr>
<tr>
<td>All Patients</td>
<td>10/152 (7%)</td>
<td>2/167 (1%)</td>
</tr>
<tr>
<td>&gt;5 (All, excluding 1-5)</td>
<td>10/123 (8%)</td>
<td>2/160 (1%)</td>
</tr>
</tbody>
</table>

Table 3 Number of patients treated 91-180 minutes after onset who died within 90 days post-treatment

<table>
<thead>
<tr>
<th>Baseline NIHSS</th>
<th>Patients with Rankin Good Outcome (0,1) at Three months</th>
<th>Relative Risk (95%CI)</th>
</tr>
</thead>
<tbody>
<tr>
<td>(Patients treated 91 to 180 minutes)</td>
<td>TPA</td>
<td>Placebo</td>
</tr>
<tr>
<td>1-5</td>
<td>0/29 (0%)</td>
<td>0/7 (0%)</td>
</tr>
<tr>
<td>6-10</td>
<td>0/37 (0%)</td>
<td>4/46 (9%)</td>
</tr>
<tr>
<td>11-15</td>
<td>5/26 (19%)</td>
<td>7/35 (20%)</td>
</tr>
<tr>
<td>16-20</td>
<td>7/33 (21%)</td>
<td>8/33 (24%)</td>
</tr>
</tbody>
</table>
These data suggest that, while the numbers are too small in each subgroup to reach significant conclusions, in each subgroup of NIHSS scores the trends are all in the direction of more favorable outcome (relative risk > 1.0 of 3 month Rankin score = 0 or 1) in the TPA group compared to placebo in all those except the NIHSS 0-5 group (the “mild” strokes that were, in fact, more prevalent in the TPA group treated from 91-180 minutes). In the total group, and in the subgroup that most concerns Dr. Mann, those with NIHSS > 5 treated 91-180 minutes post stroke, the data significantly favor TPA. As found in our previous publications (5), the risk of hemorrhage is higher in the TPA groups, but nevertheless there is no difference in mortality between TPA and placebo in any subgroup based on baseline NIHSS or treatment interval.

We have tried to briefly address Dr. Mann’s charges. There were numerous other technical errors and incorrect assumptions that he made. We will be submitting a more detailed analysis to a peer-reviewed journal shortly. Dr. Mann does not question the benefit of TPA less than 91 minutes after stroke onset. We are confident that benefit remains for patients treated 91-180 minutes, but that the benefit from treatment is less as time from onset increases. We would like to reemphasize the importance of the one hour door to needle time and seek the advice of the community of physicians on how to deliver this beneficial treatment to appropriate patients as soon as possible within the three hour limit.

Dr. Trotter is misinformed (3). The NINDS trial was not sponsored by Genentech. The greatest possible distance from their influence was maintained during the conduct of the trial and immediately following the publication of the results. Genentech did provide the drug and did do the extra documentation required to apply for approval for use in treatment of acute stroke. They did not control the data from the trial and were not provided the data until after the final analysis had been completed. None of the NINDS investigators were employees or consultants to Genentech either during the conduct of the trial or at the time of the FDA hearings that occurred 6 months after the publication of the results. We cannot speak to the actions of the AHA and the financial relationship of that organization to Genentech, but the original guidelines recommending TPA(10) were formulated not by the AHA, but by independent “clinician- interpreters” empanelled by the AHA, 8/13 of whom had nothing to do with the NINDS trial, and none of whom personally were receiving funds from Genentech at the time the guidelines were formulated. The guidelines that seem to bother Drs. Mann and Trotter recommending the use of TPA for stroke as “standard of care” were formulated much later and for Emergency Physicians (11). We cannot deny that we are convinced by the data from the NINDS TPA stroke study. The results of the NINDS study have been confirmed by numerous independent reports from both academic and community hospitals. There is one report from Cleveland of bad experience in a community hospital setting (12). A recent abstract (13) reported outcomes in the same Cleveland hospitals consistent with those from the NINDS trial after educational efforts resulted in better protocol adherence. TPA is presently the only available effective drug treatment for stroke. We hope community physicians will join the effort to find better treatments in the future.

The NINDS investigators agree with Dr. Trotter on one important point—that sensationalistic journalism often presents a biased viewpoint. An excellent example of this is Jeanne Lenzer’s article in the BMJ (1). In this article, Ms. Lenzer raises many of the same issues as does Dr. Mann and that we have just addressed in the preceding paragraphs, but she makes no effort to balance her “investigative journalism” with any of the abundant evidence supporting the use of TPA. She is also inaccurate in the evidence she presents. For example, in her first paragraphs, what is her evidence that the treatment
recommendations to use TPA “could cost more lives than the disease itself”? This is a wild exaggeration, considering that of the over 400,000 acute ischemic strokes occurring in the U.S. yearly, about 15% or 60,000 will die in the first month without treatment, and that there was no excess mortality in TPA treated patients in the NINDS trial. Even in the worst case scenario as published in the “Cleveland study” (12), where there was an excess mortality of 10% in treated patients, this would amount to 800 excess deaths nationwide considering that such poor results occurred in a setting where less than 2% of stroke patients were treated.

In response to Dr. Wardlaw (4), we appreciate her assertion that the NINDS trial results are valid and that the results are supported by findings in other trials involving over 2400 patients. We would like correct her statement that the patients in the trial were treated at tertiary referral centers. Most of the patients in the NINDS trial were treated at community hospitals. Dr. Wardlaw also questions the blinding of the investigators in determining the primary 3-month outcome measures. She may not be aware of the extra efforts made in the NINDS trials. By protocol, the persons ascertaining the 3-month outcomes were persons who were not present at the time the patients were randomized. As we are, she is concerned that few people are receiving the treatment, but please note that she expresses no doubt that TPA has a beneficial effect. Her solution to the problem of limited use of TPA for stroke in practice is to be satisfied with an even smaller benefit for the larger group of patients in the 0-6 hour time window. Since she would be satisfied with a smaller effect, a much larger trial will be needed. If she wishes to stratify her trial by NIHSS, then she may do so, however, the need for stratification would seem to be even less important in a larger trial. The NINDS investigators would hope that convincing evidence of the benefit of TPA beyond three hours is eventually proven in a prospective trial. However, in the meantime, let’s focus on maintaining the shortest possible door to needle time and encouraging more patients to recognize stroke and come immediately to the hospital for emergency care. The recent delayed diagnosis of President Gerald Ford’s stroke should stand as a reminder to all of us how far we have to go in the treatment of this major disabling disease.

The NINDS rt-PA Stroke Trial was an unbiased study. Its results are highly significant and have been replicated in practice. These two consecutive randomized trials each showed there was less than 1% chance that the positive result was due to chance so that the likelihood that the positive results in both trials were due to chance is vanishingly small. We still think efforts to make this treatment available to as many patients as possible are warranted. Not learning how to carry out this therapy for appropriate patients at appropriate centers where acute stroke patients are brought, because of the guise that the NINDS study was biased, cannot be defended. Dr. Wardlaw is correct that “patients with stroke….are unable to fight for their rights”. It is up to those entrusted with their care to put away hyperbole and resistance to change. We are open to suggestions on how we can work with the larger community of physicians to make the promise of thrombolytic therapy a reality for more of the millions of patients who have a stroke each year.

References

I appreciate Dr Grotta's willingness to submit a detailed response on behalf of the NINDS study group, and welcome his provision of more detailed subgroup data for further analysis. The NINDS study group's rapid response letter is a testament to the social value of the bmj's rapid response section, because it demonstrates that the rapid response section can serve as a valuable forum for serious scientific discussion. The readers of the bmj (and the wider public) are much better served if they can study and analyse the arguments of multiple discussants, who obviously have different points-of-view. All the discussants obviously "spin" their analysis of the data, and the independently-thinking reader is hopefully able to determine the likely "scientific truth" by carefully perusing the different points-of-view.

Dr. Grotta stated "Dr. Mann writes that in order to obtain valid results, critical prognostic variables have to be prespecified, and corrected for, in the design of any randomized controlled trial. This is true, but Dr. Mann’s choice of prognostic variables is different than ours. What we think is important is predicting who will respond to TPA. Dr. Mann is most concerned with who will do well despite therapy." Dr Grotta is entirely mistaken if he presumes that I am more concerned with who will do well despite therapy. Surely, both factors have to be carefully considered when designing a tPA-for-stroke trial? I can appreciate the fact that the NINDS investigators mainly focused their attention on predicting who will most likely respond to tPA when designing their stroke trial -- along the practical lines suggested by David Sackett [1] who stated that confidence in a trial's results is greater when the signal/noise ratio of the trial is enhanced. According to Sackett the "signal describes the differences between the effects of the experimental and control
By deliberatedly choosing patients who would most likely have a substantial response to tPA, the NINDS trialists would obviously maximise the "signal", which would subsequently increase one's confidence in tPA's efficacy if the trial's results turned out to be positive. However, surely it is equally important to decrease the "noise" in order to be confident in the validity of the NINDS trial's results? Consider Sackett's definition of "noise", which he defines as "noise (or uncertainty) in an RCT is the sum of all the factors ('sources of variation') that can affect the absolute risk reduction or absolute difference". In the case of tPA- for-stroke trials that are poorly balanced for baseline stroke severity, variations in the expected rate of a favorable stroke outcome (due to the natural course of the disease) vary according to the degree of imbalance in baseline stroke severity. Significant imbalances in baseline stroke severity between treated and placebo patients create considerable "noise" -- because they produce a significant chance-variability in the rate of a favorable stroke outcome that may obscure (magnify/diminish) the "true" efficacy of tPA and cause the "apparent" efficacy of tPA to be greater-or- less than the "true" efficacy of tPA.

Now that Dr. Grotta has published the favorable stroke outcome results for each subgroup, it is much easier to demonstrate the degree of "noise" caused by stroke severity imbalances in the NINDS trial by simply reviewing the results presented in table 1 in Dr. Grotta's rapid response letter. The NINDS investigators' own data-presentation shows that the "apparent" efficacy of tPA for patients treated between 91-180 minutes is 21% (46% minus 25%), and only 14% (37% minus 23%) when the NIHSS 0-5 subgroup's results are eliminated from consideration. That represents a one-third reduction in tPA's "apparent" efficacy due to the elimination of "noise" from the biggest single source of confounding due to stroke severity imbalances between treated and placebo patients in the NINDS trial -- the stroke outcome results from the NIHSS 0-5 subgroups. The 7% absolute difference (due to the recruitment of such a large percentage of very mild stroke patients) was chalked up as being due to tPA therapy, when it was obviously due to the natural course of the disease. It is ironic that such a high proportion of very mild stroke patients were recruited into the NINDS trial. Patients with very mild strokes (NIHSS 0-5) represented about 20% of the total number of tPA patients treated between 91-180 minutes. Recruiting patients with very mild strokes is contrary to Sackett's basic principle of mainly recruiting high risk patients, who would more likely show a substantial response to tPA therapy. It was also contrary to the NINDS investigators' own policy of discouraging the recruitment of patients with a NIHSS score of < 4 [2].

There may be another significant source of "noise" due to stroke severity imbalances that may cause the "apparent" efficacy of tPA to be different from the "true" efficacy of tPA -- "noise" that would be generated if the placebo and tPA patients within EACH subgroup were not near-perfectly balanced for baseline stroke severity (even though the total number of placebo and tPA patients in each subgroup was near-equal). I explored that particular issue at great length in a letter to the CMAJ [3] and I wonder to what degree stroke severity imbalances within each/all of the NINDS subgroup's could be a confounding factor. The true answer to that question will only become fully apparent when the NINDS investigators make all the raw data from the NINDS trial publically available -- so that the public can much more accurately determine how well-balanced EACH of the subgroups were for baseline stroke severity. According to the TOAST graph [4] a two-point difference in the "average" baseline NIHSS score between treated and placebo patients in the NIHSS range of 16-20 can cause a 3-5% absolute difference in the rate of a favorable stroke outcome, which could alter the "apparent" efficacy of tPA for those patients by a factor of 30-50%. The figure of 30-50% is obtained by simply examining the NIHSS 16-20 subgroups' results in table 1, which showed that the apparent" absolute efficacy of tPA for those patients was 9% (27% minus 18%).

Another probable example of "noise" due to stroke severity imbalances within the NINDS trial's subgroups can be ascertained by looking at the NIHSS 11-15 and NIHSS 16-20 placebo subgroup's rate of favorable stroke outcome results in table 1 in Dr. Grotta's letter. The rate of favorable stroke outcome for the NIHSS 11-15 placebo subgroup was 14%,
which was less than the figure of 18% for the NIHSS 16-20 placebo subgroup. That result is obviously surprising because patients with a baseline NIHSS stroke severity score of 11-15 are naturally expected to have a much better stroke outcome result than patients with a baseline stroke severity NIHSS score of 16-20. The figure of 14% seems to be extraordinarily low for untreated patients with a baseline NIHSS score in that stroke severity range and it is much less than would be predicted. It would be very informative if the NINDS trialists would publish the rate of favorable stroke outcome results for the NIHSS 11-15 placebo subgroups from the 0-90 minute arm of the NINDS trial, ECASS trial, ECASS II trial and ATLANTIS trial (including their "average" baseline stroke severity scores). It would be extremely useful to know whether the "average" rate of favorable stroke outcome of the NIHSS 11-15 placebo patients from those other trials are closer to the 34% figure predicted by the TOAST graph [4], and whether the comparable NINDS placebo results from the 91-180 minute cohort is a statistical outlier that artefactually inflates the "apparent" efficacy of tPA in that subgroup of patients. How much of an effect could this particular imbalance have if the "true" rate of a favorable stroke outcome for placebo patients in the NIHSS 11-15 subgroup was 30%? The answer is that an additional 5 patients would have a favorable stroke outcome.

Finally, there is another "noise" element factor due to stroke severity imbalances in the NINDS trial that should be considered. Note that there were 18 more patients in the placebo group in the NIHSS >20 subgroup (compared to the tPA group), and that those patients only had a 4% chance likelihood of a favorable stroke outcome. If those 18 patients were equally distributed between the NIHSS 5-10, 10-15 and 16-20 subgroups, then an additional 6 patients would have a favorable stroke outcome. Adding that figure of 6 patients to the 5 additional patients from the NIHSS 11-15 subgroup means that an additional 11 placebo patients would have a favorable stroke outcome. Then the computed figure for the placebo group (excluding the NIHSS 0-5 subgroup) would be 47/160 and not 36/160, which translates to 29% and not 23%. That means that the "apparent" efficacy of tPA would be reduced by another 6%, and it would only be 8% and not 14%.

How useful is this post hoc conjecturing about the NINDS trial's subgroup data? Dr. Grotta stated "It is also very important to remember that these post-hoc analyses involving subgroups without sufficient statistical power to answer the meaningful question should be considered as "hypothesis generating" and providing a rationale for further study only." I agree with Dr. Grotta, and I think that my post hoc conjecturing about the "noise" influence of stroke severity imbalances within the NINDS trial's subgroups is simply a "hypothetical" explanation, which only becomes a "realistic" explanation if the raw data supports my theory's basic tenets. That is why I have requested that the NINDS investigators make the pooled raw data from all the tPA-for-stroke trials publically available [5], so that the raw data can be independently examined. The pooled results from all the tPA-for-stroke trials should be examined for EACH level of baseline stroke severity (from a baseline NIHSS score of 1-25) for different times-to-treatment, so that the "noise" due to stroke severity imbalances can be eliminated as a confounding factor. By also examining the favorable stroke outcome results for different times -to-treatment, it will immediately become clear to what degree delays in time-to-treatment affect the "apparent" efficacy of tPA -- without having to depend on the hypothetical model constructed by the NINDS investigators [6].

A number of other statements made by Dr. Grotta deserve further commentary. He stated "Contrary to Dr. Mann's assertion, NIH Stroke Scale was a prespecified variable that was known to predict outcome and it was corrected for in the usual way in the original publication." I have parsed that original publication countless times and I have never read any statement that implied that the NINDS investigators had corrected for imbalances in baseline stroke severity between treated and placebo patients. Hopefully, the NINDS investigators, or other bmj readers, could point out the particular "statement/statements" in that original publication that I must have missed. Dr. Grotta also stated "The facts are that the baseline imbalance of stroke severity DOES NOT explain the entire results of the trial. The imbalance explains some of the difference, but there is no question that there is still
benefit from TPA 91-180 minutes after stroke onset.” I wholeheartedly agree with Dr. Grotta -- the imbalance only explains some of the difference. However, the critical question is how much of the difference is due to stroke severity imbalances and how much is due to the "true" efficacy of tPA? That question has presently not been answered, and I strongly suspect that an accurate answer will only become apparent when all of the NINDS trial's raw data is made available to the public. Do the NINDS investigators have a valid reason for not making the raw data available -- considering that the study was funded with public money through the NIH? There is a disturbing dissonance between the refusal of the trial's investigators to make patient-level data publicly available, and the NIH's traditional stance on the dissemination of the results of NIH-sponsored research. Indeed, this is unequivocally explicated in a draft policy statement concerning data sharing released on March 1, 2002 [7]. The statement states "There are many reasons to share data from NIH-supported studies. Sharing data reinforces open scientific inquiry, encourages diversity of analysis and opinion, promotes new research, makes possible the testing of new or alternative hypotheses and methods of analysis -----". In fact, the NIH draft statement makes some definite recommendations and it explicitly states "The NIH will expect investigators supported by NIH funding to make their research data available to the scientific community for subsequent analyses."

Finally, Dr. Grotta also stated "The results of the NINDS study have been confirmed by numerous independent reports from both academic and community hospitals." How is that possible if those studies did not have a placebo arm? By what means could post-marketing tPA-for-stroke studies determine the "true" efficacy of tPA if they did not have an absolute or relative comparator? In the absence of an absolute comparator (equally balanced group of placebo patients in a RCT), one could theoretically only determine that the "other" study had a similar degree of efficacy as the NINDS trial if the "other" study had tPA patients with an identical stroke severity distribution as the original NINDS trial. Does anyone know of such a study? In the absence of knowledge of such a study, I took the Multicentre Stroke Survey's group of >1,000 tPA patients, who had an "average" rate of a favorable stroke outcome of 33%, and I calculated the likelihood of a similar group of untreated stroke patients having a favorable stroke outcome due to the natural course of the disease (using data from the NINDS trial and not the TOAST study). The calculated results were reported in my rapid response letter [3] and the estimated "average" figure was 31.7%. That figure suggests that tPA was probably not significantly efficacious in those patients. Although the results are only based on a relative comparison, which is not universally regarded as being statistically valid, I would be interested in knowing if anyone has a better means of demonstrating how the results of post-marketing studies (which are not RCTs) can accurately confirm-or-refute the positive results of the NINDS trial.

Jeffrey Mann.

References:

1. Sackett, David L. Why randomized controlled trials fail but needn't: 2. Failure to employ physiological statistics, or the only formula a clinician-trialist is ever likely to need (or understand!) CMAJ: Canadian Medical Association Journal. 165(9):1226-1237, October 30, 2001.

Available online at http://www.cmaj.ca/cgi/content/full/165/9/1226

2. Comment by Patrick Lyden at the FDA Advisory Committee meeting - June 6th, 1996. From the meeting’s transcripts - lines 15-16 on page 183.


Available online at http://bmj.com/cgi/eletters/324/7339/723#22326

6. Representative copy of figure 2 from the Marler article.

Available online at http://emguidemaps.homestead.com/files/marlergraph.html

7. NIH announces draft statement on sharing research data. Release Date: March 1, 2002.


The difference between the "apparent" and "true" efficacy of tPA in the NINDS trial

Jeffrey Mann, Physician
Salt Lake City, UT 84103

Send response to journal: 
Re: The difference between the "apparent" and "true" efficacy of tPA in the NINDS trial

As a result of recent e-mail discussions with a number of tPA- proponents, I have come to realize that they cannot appreciate the validity of my criticism of the NINDS trialists' interpretation of the NINDS trial, and I have thought of another method of making my critical points more vividly real and also easier to understand.

If one reviews Dr. Grotta's presentation of data in table 1 in his rapid response letter to the bmj [1], one will note that the efficacy of tPA for patients treated between 91-180 minutes in the NINDS trial was calculated by dividing the total number of patients having a favorable response by the total number of patients in that arm of the trial. The figure for the treated patients was 70/153 (46%) and the figure for the placebo patients was 42/167 (25%) so that the calculated efficacy of tPA was 21%. I have argued that the calculated efficacy figure is flawed because there were 22 more tPA patients than placebo patients in the NIHSS 1-5 subgroup and 18 more placebo patients than tPA patients in the NIHSS>20 subgroup. Many people apparently have difficulty understanding how that imbalance could affect the validity of the NINDS trial's efficacy figure of 21%, and the following theoretical demonstration may make it much easier to understand how that imbalance affected the calculation of tPA's efficacy.

Was it fair to have 18 more placebo patients than tPA patients in the most severe stroke severity subgroup (NIHSS >20), and 22 more treated patients than placebo patients in the mildest stroke severity group (NIHSS 1-5) in the NINDS trial? The answer may become much clearer if one simply corrected the situation and ensured that there were an equal number of patients in the NIHSS 1-5 and NIHSS>20 subgroups, while leaving all the other patient numbers and response rates the same. Then a "hypothetically fairer" trial would have the following results.

The results from Dr. Grotta's table 1 would look like this: -

Baseline NIHSS subgroup--tPA patients with favorable outcome--placebo patients with favorable outcome

1-5 subgroup -- 24/29 (83%) -- 25/29 (86%)
6-10 subgroup -- 23/37 (62%) -- 23/46 (50%)
11-15 subgroup -- 10/26 (38%) -- 5/35 (14%)
16-20 subgroup -- 9/33 (27%) -- 6/33 (18%)
> 20 subgroup -- 4/28 (14%) -- 1/28 (4%)

http://bmj.bmjournals.com/cgi/eletters/324/7339/723
ALL patients -- 70/153 (46%) -- 60/171 (35%)

Calculated efficacy of tPA = 11%.

The calculated efficacy of tPA would be 11% (and not 21% as was originally calculated in the NINDS trial). In other words, the "apparent" therapeutic benefit of tPA would only be 50% of the figure calculated by the NINDS trialists. Also, note that the placebo patients would have an overall rate of favorable stroke outcome of 35%, and that figure is much closer to the results obtained by the placebo patients in the other tPA-for-stroke RCTs, than the NINDS trialists' placebo figure of 25%.

Is there a tangential way to prove that the "apparent" efficacy figure of 11% is more likely to be accurate than the "apparent" efficacy figure of 21%?

Consider the calculated efficacy results of tPA for each subgroup from table 1 in Dr. Grotta's letter.

NIHSS 1-5 subgroup = minus 3% (83% - 86%)
NIHSS 6-10 subgroup = 12% (62%-50%)
NIHSS 11-15 subgroup = 24% (38%-14%)
NIHSS 16-20 subgroup = 9% (27%-18%)
NIHSS >20 subgroup = 10% (14-4%)

Note that if one removed the NIHSS 1-5 subgroup from consideration (because patients with very mild strokes regularly have an excellent recovery rate due to the natural course of the disease), then there would be 4 subgroups remaining with the following efficacy results.

NIHSS 6-10 subgroup = 12%
NIHSS 11-15 subgroup = 24%
NIHSS 16-20 subgroup = 9%
NIHSS >20 subgroup = 10%

Note that the NIHSS 11-15 subgroup is a statistical outlier, and that if one removed its results from consideration, then the "average" efficacy result would be between 10-11%. How does one explain the wayward efficacy result of 24% for the NIHSS 11-15 subgroup? Salim Yusuf in his article on the analysis and interpretation of subgroup results [2] states "Even when treatments have similar effects in different subgroups, the play of chance may well exaggerate, dilute, or occasionally reverse the results in a particular subgroup." In other words, the wayward results of that particular subgroup could be due to chance. Another method of gauging that chance probably affected the results of the NIHSS 11-15 subgroup is to examine the efficacy results of the subgroups immediately above and below that particular subgroup's results. Physiological expectations, based on the expectation that tPA should have similar therapeutic effects in all subgroups in the middle of the stroke severity range, suggest that the NIHSS 11-15 subgroup should have efficacy results between 9-12%. In fact, the NIHSS 11-15 subgroup's results would be between 9-12% if the placebo patients' rate of favorable stroke outcome result was in the range of 26- 29% (rather than 14%). I have previously argued [3] that the TOAST graph predicted that the NIHSS 11-15 placebo group's rate of favorable stroke results should be around 30%. Therefore, if one corrected for that fact, then the 91-180 minute subgroup results would appear as follows:-

NIHSS 6-10 subgroup = 12%
NIHSS 11-15 subgroup = 9-12%
NIHSS 16-20 subgroup = 9%
NIHSS >20 subgroup = 10%

Then the calculated "average" efficacy of tPA would be 10-11%, which is identical to the figure obtained by the standard methodology (as used by the NINDS trialists) -- after correcting for stroke severity imbalances.

Is this type of statistical correction for wayward subgroup results acceptable? Yusuf in his article on the analysis of subgroup results [2] stated that one should "interpret the results in the context of similar data from other trials, from the architecture of the entire set of data on all patients, and from principles of biological coherence". Therefore, I think that it is entirely appropriate to make that type of correction, which is both physiologically coherent and consistent with the data from the other NIHSS subgroups and other stroke trials. Some people would still argue that post hoc subgroup analyses are fraught with potential inaccuracies because of small sample sizes. That argument is entirely valid. However, isn't it interesting that the "average" efficacy figure of 10-11%, which was determined by the "averaging" methodology of post hoc subgroup analysis, is identical to the efficacy figure obtained using the standard methodology -- after making an appropriate correction for stroke severity imbalances? As an aside -- the figures are not really identical, because one would also need to make the same statistical correction for the NIHSS 11-15 placebo group when using the standard methodology, and that would decrease the "apparent" efficacy results by 3% [3], so that the "apparent" efficacy of tPA using the standard methodology (after correction for stroke severity imbalances) would only be 8% (46%-38%).

If the "true" efficacy of tPA is approximately 8-11%, then it is less than 50% of the calculated "apparent" efficacy of tPA (21%) as determined by the NINDS trialists using their standard methodology (without correcting for imbalances in baseline stroke severity).

If the "true" efficacy of tPA is only 8-11% for patients treated between 91-180 minutes, and the NINDS trialists hypothesize that tPA's efficacy wanes throughout the 91-180 minute time period [4], then what is the likely efficacy of tPA for stroke patients treated between 150-180 minutes? Would the likely efficacy rate be greater-or-less than the likely rate of a symptomatic ICH (~7%)? The "estimated" answer may markedly affect one's calculation of the risk:benefit ratio of tPA, and one's decision whether to utilize the drug for stroke patients who can only be treated >150 minutes from the time of stroke onset.

Jeffrey Mann.

References:

Available online at http://bmj.com/cgi/eletters/324/7339/723#23369


3. Mann J. The raw data of the NINDS trial should be made public. bmj rapid response letter. 8 July 2002.
Available online at http://bmj.com/cgi/eletters/324/7339/723#23369

4. Representative copy of figure 2 from the Marler article.
Available online at http://emguidemaps.homestead.com/files/marlergraph.html
Conflict of interest important at all levels

Veronica M Wilkie, GP Postgraduate Tutor The Charles Hastings Education Centre, Charles Hastings Way Worcester WR5 1DD

Send response to journal: Re: Conflict of interest important at all levels

Many educational events are sponsored by pharmaceutical companies; postgraduate tutors in general practice have an uneasy relationship because of the absence of any funding for their meetings. There seems to be an increasing tendency for Primary Care Organisations to use Drug Company Sponsorship for "educational meetings"; some of which will be more about policy changes than "pure education" eg A County diabetic group changing its policy on the use of newer oral hypoglycaemics, discussing this over dinner at a hotel, all paid for by both manufacturers of the said drugs. The need for HAs and PCOs to be aware of this and perhaps change their policy of industry sponsorship should be tackled in the same way as larger national organisations.

Doubts and ethics

Mark H Miller, Emergency Physician Seattle, Washington 98133

Send response to journal: Re: Doubts and ethics

I have read the above discussion with great interest. While I cannot pretend to sort out the statistical and procedural arguments I think one thing is clear: further discussion alone is not going to resolve the various questions and doubts presented. It would appear to me that there is at least a reasonable possibility that we are doing our patients more harm than good using thrombolysis for stroke. It is clear that some patients are harmed by thrombolysis, a statement that can be made about many therapeutic endeavors. However, as opposed to most therapies the number needed to harm here is quite low, not even one order of magnitude less than the number needed to treat.

For me this brings up a serious ethical dilemma. Were it the case that those patients with the most severe strokes and terrible outcomes were the only ones who might suffer intracerebral hemorrhage from thrombolysis one might justify the low treat to harm ratio as their outcome would have been terrible in any situation.

But are we not taking patients, some of whom would otherwise have had a reasonable outcome, and converting them to patients with terrible or lethal outcomes? Do we really have the right to do so? This is usually not an ethical dilemma since either the risk of doing harm is so small compared to the benefit that it is not an ethical question or the disease process is so lethal that it is worth the risk (e.g. bone marrow transplant for leukemia.) But in this case I think we need to be able to fall back on a whole lot more than a P value. We each need to know that if that patient who right now is a little dysarthric with some lateralizing weakness is suddenly aphasic, hemianopic, hemiparetic and bedridden, or possibly dead due to the therapy we advised, that we know that it was still the right therapy to have advised. I don't think many of us feel that secure about the data. Let's repeat the studies and examine the raw date from the NINDS study as well.

Mark Miller MD Diplomate, ABIM, ABEM Seattle

Competing interests: None declared

The NINDS trial was too small in size to be confident in the trial’s results

Jeffrey Mann, Retired physician Salt Lake

It is interesting to discover how the NINDS investigators deal with my criticism of the NINDS trial's result-analysis. I have repeatedly questioned the validity of the results of the 91-180 minute arm of the NINDS trial, because of the imbalance in baseline stroke
severity between the tPA-treated patients and placebo patients in that arm of the trial.

In a talk called "Accomplishments in Stroke Care" presented at the NINDS stroke symposium in December 2002 (freely available online at reference number 1), Patrick Lyden presents a slide (slide number 17) that demonstrates that the OR is still 1.7 if one leaves out patients with a baseline NIHSS score of <5 and >20. However, look at the wide confidence intervals which cross an OR figure of 1.0 (unity). That "fact" demonstrates that the NINDS trial was too small in size to allow a person to be confident in the validity of the results for those stroke patients treated between 91-180 minutes.

In another slide (slide number 7) Patrick Lyden suggests that one doesn't need a larger study sample size than 600 patients for tPA-for- stroke trials because the ARR is 12% (compared to 2% for cardiac tPA trials). I think that his argument is without merit, because the 12% ARR figure is only based on one controversial study -- the NINDS trial. That would be like arguing that one only needs a sample size of a few hundred patients for clinical trials of PG2b3a inhibitors in ACS patients, because the EPISTENT trial showed a significant mortality reduction at one year. However, the EPISTENT study was an "outlier", and all the other trials of PG2b3a inhibitors in ACS patients showed no mortality reduction (Read pages 69-73 in reference number 2 to see how all the experts agreed that there was no mortality reduction in all those trials).

Another issue that Patrick Lyden doesn't deal with are the "strange" results in the NIHSS 11-15 subgroup (91-180 minutes arm). If you read my rapid response letter to the bmj [3], you will note that the placebo group only had a 14% excellent stroke outcome rate. That 14% figure is very low, and it makes the overall OR results of 1.7 for the 91-180 minute arm (excluding stroke patients with a NIHSS score <5 and >20) appear much better than it probably really is -- compared to a more realistic situation where the placebo group is more representative of community patients with a stroke severity in the NIHSS 11-15 range (who can be expected to have an excellent stroke outcome rate of ~30%). That is why I would like to get the NINDS study's patient-level data to examine that particular subgroup's results.

I have previously suggested that the NINDS investigators should make all the patient-level raw data from the NINDS study publically available. However, they have apparently failed to so, despite repeated requests from multiple people over an extended period of time. How can they ethically justify not making their patient-level raw data publically available?

References:


2. FDA Advisory meeting transcripts. Available from the FDA's website => Go to the list of Cardiovascular and Renal Drugs Advisory Committee meetings for 1999 => Go to October 14th 1999 => Choose Microsoft word document number 3555t1.rtf.

Also available at the following webpage: http://www.fda.gov/ohrms/dockets/ac/cder99t.htm

3. Rapid response letter -- Mann J. The difference between the "apparent" and "true" efficacy of tPA in the NINDS trial.

Available online at http://bmj.com/cgi/eletters/324/7339/723#23927.

Competing interests: None declared
Jeffrey Mann,
Retired physician
Salt Lake City, UT
84103

Send response to journal:
Re: A personal analysis of the NINDS study using patient-level raw data

I have written a number of rapid response letters to bmj.com questioning the validity of the NINDS trialists' interpretation of the results of the 91-180 minutes arm of the NINDS study -- on the basis of an imbalance in baseline stroke severity between the treated and placebo groups. I could never precisely quantify the "degree" to which this imbalance problem affected the "correct" interpretation of the NINDS study's results, because I did not have access to the study's patient-level raw data.

I recently received the NINDS study's patient-level raw data. The raw data enabled me to more thoroughly test my theory that imbalances in baseline stroke severity between the treatment and placebo groups of the 91-180 minute arm of the NINDS study invalidated the "official" interpretation of the NINDS study's results. After studying the raw data, I became increasingly convinced of that fact. I therefore decided to write a new manuscript that deeply analyses the NINDS study's results.

The manuscript is called "A personal analysis of the NINDS study using patient-level data".

It is available at

If you cannot access that website file, go to
http://www.homestead.com/emguidemaps/JeffMannEMguidemaps.html, and click on the manuscript's title in the soapbox section.

My general conclusion is that the imbalance in baseline stroke severity between the treated and placebo groups in the 91-180 minutes arm of the NINDS study accounted for approximately 50% of the estimated "apparent" efficacy of tPA therapy for stroke patients treated after 90 minutes. What do you think of my "theoretical" estimation?

Jeff Mann. MD.

Competing interests: None declared

Is an adjusted OR more accurately reflective of reality than an unadjusted OR

Jeffrey Mann,
Retired physician
Salt Lake City, UT
84103

Send response to journal:
Re: Is an adjusted OR more accurately reflective of reality than an unadjusted OR

The NINDS investigators have done it again. They have reinterpreted the pooled tPA-for-stroke data using obscure statistical techniques that <1:1,000 medical journal readers could possibly understand, in an attempt to demonstrate that tPA is significantly effective even beyond the 3 hour point in time. They first used this obscure statistical technique to create a hypothetical model, which they first published in an article in December 2000 [1]. I never met ayone who could understand the statistical technique used to create that hypothetical model. Now they have repeated their performance with a new version of that hypothetical model that predicts an even greater time-to-treatment benefit than their original hypothetical model predicted.

In an article published in the March 6, 2004 issue of the Lancet [2], the ATLANTIS-ECASS-NINDS rt-PA study group investigators published an article demonstrating that tPA therapy had a much better therapeutic effect if given early rather than later. In particular, they claimed that the adjusted OR for patients treated in <90 minutes was 2.81 compared to an adjusted OR of 1.55 for patients treated between 91-180 minutes, an adjusted OR of 1.40 for patients treated between 181-270 minutes, and an adjusted OR of 1.15 for patients treated between 271-360 minutes.

How did they calculate the adjusted OR? They stated that they used a global statistic that incorporated three different stroke outcome scoring systems and
then made an additional adjustment to take into account certain variables - age, baseline glucose concentration, baseline NIHSS, baseline diastolic blood pressure, previous hypertension, and interaction between age and baseline NIHSS.

How many of you understand how the ATLANTIS-ECASS-NINDS rt-PA study group investigators could validly create a single global stroke outcome measure, which incorporated three different stroke outcome scoring systems (NIHSS 0-1, modified Rankin 0-1, Barthel Index 95-100) that do not have a linear correlation with each other? Do you believe that it is scientifically appropriate, and accurately reflective of reality, to create such an arbitrary stroke outcome system? Do you not think that the ATLANTIS-ECASS-NINDS rt-PA study group investigators should have explained how they created that arbitrary stroke outcome scoring system by posting the precise details online, together with the other pooled data details that they posted online at [http://image.thelancet.com/extras/03art1122webappendix2.pdf](http://image.thelancet.com/extras/03art1122webappendix2.pdf)? In the modern internet era, there is no limit to the amount of data that be published online in an appendix to an article, and the ATLANTIS-ECASS-NINDS rt-PA study group investigators could easily post complete details of their statistical methodology online.

Note that the ATLANTIS-ECASS-NINDS rt-PA study group investigators decided to adjust the OR for certain variables. Do you believe that an adjusted OR is more accurate and more reflective of tPA's "true" treatment effect than an unadjusted OR? On what basis can you make that decision without understanding how those variables influence tPA's treatment effect? Do you understand how the ATLANTIS-ECASS-NINDS study group investigators made that adjustment? Do you not think that the ATLANTIS-ECASS-NINDS rt-PA study group investigators should have posted precise details regarding their statistical methodology at [http://image.thelancet.com/extras/03art1122webappendix2.pdf](http://image.thelancet.com/extras/03art1122webappendix2.pdf), so that other statisticians could check their methodology to determine whether they agreed that an adjusted OR provides a more accurate reflection of tPA's "true" treatment effect than an unadjusted OR?

The following table contains an unadjusted OR for patients treated between 0-90 minutes from the NINDS study using a single stroke outcome scoring system (modified Rankin score <1) for each of the stroke severity subgroups. In the fifth column are the *adjusted ORs for a favorable stroke outcome for each of the five stroke severity groups, which I obtained from figure 2 in the Lancet article. The ATLANTIS-ECASS-NINDS rt-PA study group investigators stated that they adjusted the OR for age, baseline glucose concentration, baseline diastolic pressure, and previous hypertension. (There is a slight difference in the patient numbers between my table and figure 2, and I am not sure why there is that small discrepancy considering that all-or-most of the patients must have come from the NINDS study, but the difference is small and it shouldn't significantly affect the results of the calculations).

| NINDS study: 0-90 minute arm. Rate of a favorable stroke outcome at 3 months (mRS<1) |
|-----------------|-----------------|-----------------|-----------------|-----------------|
| Baseline NIHSS | Treated patients | Placebo patients | Unadjusted OR   | *Adjusted OR    |
| 0-5            | 9/13 (69%)       | 7/9 (78%)        | 0.64            | 0.7             |
| 6-10           | 23/30 (77%)      | 15/37 (40%)      | 4.8             | 4.8             |
Note the difference between the adjusted OR and the unadjusted OR for the NIHSS 11-15 and NIHSS 16-20 stroke severity subgroups. Do you believe that those two adjusted OR figures are credible and that they would differ from the unadjusted OR figures to that degree -- based on a few baseline variables that have not previously been shown to have a large prognostic influence on the rate of a favorable stroke outcome?

I could not compare the unadjusted OR to the adjusted OR for patients treated after 90 minutes, because I unfortunately do not have access to the patient-level raw data from the ECASS, ECASS II and ATLANTIS trials. I would love to obtain access to that data, so that I can plot the unadjusted rate of favorable stroke outcome at 3 months for each of those 2,775 patients against the baseline NIHSS score as I did in my personal analysis of the NINDS trial's raw data [3]. I would then be able to determine whether there is a time-to-treatment interaction without having to rely on any statistical manipulations. I personally believe that the latest statistical manipulation of the data by the ATLANTIS-ECASS-NINDS rt-PA study group investigators is a distortion of reality, and that it doesn't reflect the "true" treatment effect of tPA therapy in acute ischemic stroke. I think that my simple technique of plotting the rate of favorable stroke outcome at 3 months against the baseline NIHSS score would more accurately establish the likely "true" treatment effect of tPA without any statistical distortions.

In conclusion, do you have substantial reason, based on your knowledge of the evidence-based medical literature, to believe that an adjusted OR provides a more accurate perspective of tPA's "true" treatment effect than an unadjusted OR? Do you believe that in the interest of science, that the ATLANTIS-ECASS-NINDS rt-PA study group investigators should make the patient-level raw data publically available so that other independent trial interpreters can interpret the patient-level raw data using different interpretative approaches? Due to public pressure, the NINDS investigators eventually made the raw data of the NINDS study publically available. In the interest of science, I now think that it's time for the ATLANTIS-ECASS-NINDS rt-PA study group investigators to make the patient-level raw data from all those 2,775 pooled patients publically available.

Jeffrey Mann. MD.

References:


2. ATLANTIS-ECASS-NINDS rt-PA study group investigators. Association of


Competing interests: None declared