Report of the t-PA Review Committee

W. Michael O'Fallon, Ph.D., Chair

Kjell Asplund, Ph.D., M.D.

Lewis R. Goldfrank, M.D.

Vicki Stover Hertzberg, Ph.D.

Timothy J. Ingall, MB, BS, Ph.D.

Thomas A. Louis, Ph.D.

August 25, 2004

t-PA Review Committee Roster

Chair

W. Michael O'Fallon, PhD
Professor of Biostatistics
Division of Biostatistics
Mayo Clinic Rochester
Rochester, Minnesota

Committee Members

Kjell Asplund, MD, PhD
Professor of Medicine
Head, Department of Medicine
University Hospital
Umeå, Sweden

Lewis R. Goldfrank, MD
Director, Emergency Medicine
Bellevue Hospital Center
Professor, Clinical Medicine and Surgery
New York University School of Medicine
New York, New York

Vicki Stover Hertzberg, PhD
Associate Professor
Department of Biostatistics
Emory University
Atlanta, Georgia

Timothy J Ingall MB BS, PhD
Associate Professor of Neurology
Cerebrovascular Diseases Center
Mayo Clinic Scottsdale
Scottsdale, Arizona

Thomas A. Louis, PhD
Professor, Department of Biostatistics
Johns Hopkins Bloomberg School of Public Health
Baltimore, Maryland

- **Data Analyst**

  Teresa J. H. Christianson, BS
  Data Analyst II
  Division of Biostatistics
  Mayo Clinic Rochester
  Rochester, Minnesota

Table of Contents:

1. Executive Summary 1
   - 1.1 The NINDS Charge
   - 1.2 Principal Findings
   - 1.3 Secondary Analyses
     - 1.3.1 Blood Pressure Assessment and Management
     - 1.3.2 Intracerebral Hemorrhage
     - 1.3.3 Baseline NIHSS Imbalance
     - 1.3.4 Baseline Stroke Severity and Age
     - 1.3.5 Onset to Treatment Time
     - 1.3.6 Clinical Centers
     - 1.3.7 Stroke Subtype
     - 1.3.8 Preexisting Disability
     - 1.3.9 Diabetes Mellitus
   - 1.4 Issues in Need of Further Investigation
   - 1.5 Conclusion

2. Introduction 6
   - 2.1 Background
   - 2.2 Announcement
   - 2.3 Charge
   - 2.4 Membership

3. Committee Processes 9
   - 3.1 NINDS' Independent Contractor
   - 3.2 Communications
   - 3.3 Guiding Principles
   - 3.4 Timeline

4. Methods 11
   - 4.1 Data Management
     - 4.1.1 Data from NINDS' Independent Contractor
     - 4.1.2 Variable Identification and Definition
     - 4.1.3 Result Replication
   - 4.2 Study Design
     - 4.2.1 Stratification Factors
     - 4.2.2 Outcome Measures
     - 4.2.3 Intent to Treat
   - 4.3 Analytic Methods
     - 4.3.1 Treatment Group Balance
     - 4.3.2 Missing Data Imputation
     - 4.3.3 Analytic Models
     - 4.3.4 Subgroup Analysis and Interaction Detection
     - 4.3.5 Logistic Models
     - 4.3.6 Global Analysis

4.3.7 Analyses in the Probability Scale
4.3.8 Covariate Determination (Stepwise Models)

5. Results 19
5.1 Replication of Published Results
5.2 Baseline Balance between t-PA and Placebo Groups
5.3 Observed Outcomes
5.4 Covariate Selection
5.4.1 Logistic Analysis of Favorable Individual Outcome
5.4.2 Global Model of Favorable Outcome
5.4.3 Primary Covariate Model
5.5 Model-Based Treatment Comparisons
5.5.1 Logistic and Global Model Results
5.5.2 Treatment by Covariate Interactions
5.5.2.1 Results of t-PA by Covariate Interaction Tests
5.5.2.2 Power of Interaction Tests
5.6 Absolute Risk Differences
5.6.1 Differences between Success Rates
5.6.2 Attributable Fraction
5.7 Public Health Consequences and Exploratory Subgroup Analyses

6. Blood Pressure Assessment and Management 28
6.1 Stated Methodology
6.1.1 Original Protocol
6.1.2 Blood Pressure Manuscript
6.1.3 Systems Approach
6.1.4 FDA Submission
6.1.5 Exclusion Characteristics
6.2 Stated Data Sets
6.2.1 From the Original Publication
6.2.2 From the Manuscript on BP
6.2.3 Hypertension on Admission and Antihypertensive Therapy Received before Randomization
6.3 Stated Goals, Discussion and Conclusions
6.4 The NINDS t-PA Review Committee's Evaluation of the BP Issue
6.4.1 Review Data sets
6.4.2 Investigator Queries
6.4.3 Review of Study Datasets
6.5 Review Committee's Areas of Concern
6.5.1 Definitions
6.5.2 Protocol Applications
6.5.3 Therapeutic Interventions
6.6 Review Committee's Findings
6.6.1 Definition of Hypertension
6.6.2 History of Hypertension
6.6.3 Blood Pressure on Admission and at Baseline

6.6.4 Blood Pressure Exclusion Criteria
6.6.5 Antihypertensive Therapy before Randomization
6.7 Use of Blood Pressure Data in Review Committee Outcome Models
6.7.1 Blood Pressure Variables as Favorable Outcome Predictors
6.7.2 Influence of BP Variables on OR Estimates
6.8 Summary and Conclusions

7. Intracerebral Hemorrhage 42
7.1 Introduction
7.2 Review Committee Analyses
7.2.1 Missing Values
7.2.2 ICH Analyses
7.2.2.1 ICH Risk Increases with t-PA
7.2.2.2 Favorable Outcome Chance Decreases with ICH
7.2.2.3 Favorable Outcome Chance Increases with t-PA and no ICH
7.2.2.4 ICH Related Morbidity and Mortality
7.2.3 Net t-PA Effect
7.2.3.1 Favorable Outcome Chance Increases with t-PA Among All Patients
7.2.3.2 Modeled Likelihood ORs Significantly > 1 Among All Patients
7.2.3.3 Conclusion Regarding Net Effect
7.2.4 Identification of Variables Predicting ICH
7.2.4.1 Methodological Issues
7.2.4.2 Results of ICH Risk Factor Identification
7.2.4.3 Risk Score Sensitivity and Specificity for any ICH
7.2.4.4 A Simplified Risk Function
7.3 Summary and Conclusions

8 Special Topics 52
8.1 Age, Baseline Stroke Severity, and Baseline Stroke Severity Imbalance
8.1.1 Introduction
8.1.2 Baseline NIHSS Imbalance
8.1.3 Outcomes in NIHSS Quintiles
8.1.4 Outcomes in Age Quintiles
8.1.5 Age by Baseline NIHSS Interaction
8.1.6 Model-Based Assessment of Baseline NIHSS Imbalance
8.1.6.1 Baseline NIHSS Analysis
8.1.6.2 BsNIHSS Quintile Specific Odds Ratios
8.1.7 An Alternative Variable
8.1.8 Influence of the Age by Baseline NIHSS Interaction on the t-PA Treatment Effect
8.1.9 Summary and Conclusions
8.2 Onset to Treatment Time

8.2.1 Restricted Randomization
8.2.2 Distribution of OTT
8.2.3 Does t-PA Effectiveness Decrease with Increasing OTT?
8.2.3.1 OTT by t-PA Interactions (1)
8.2.3.2 OTT by t-PA Interactions (2)
8.2.3.3 OTT by t-PA Interactions (3)
8.2.4 Summary and Conclusions
8.3 Clinical Centers
8.3.1 Center Comparisons of Favorable Outcome Rates
8.3.2 Center Comparisons of t-PA Effect
8.3.3 Center by t-PA Interaction
8.3.4 Estimates of Differences in Favorable Outcome Percentages
8.3.5 Summary and Conclusions
8.4 Stroke Subtype
8.4.1 Introduction
8.4.2 Analyses
8.4.3 Summary and Conclusions
8.5 Preexisting Disability
8.5.1 Summary and Conclusions
8.6 Diabetes Mellitus
8.6.1 Analyses
8.6.2 Summary and Conclusions

9 Conclusion 75

10 References 76

1. EXECUTIVE SUMMARY

1.1 The NINDS Charge

In May 2002, in response to concerns about the results of the NINDS rt-PA Stroke Study, the independent t-PA Review Committee was established at the request of NINDS. The main charge given to the committee was:

> "to address whether there is concern that eligible stroke patients may not benefit from rt-PA given according to the protocol used in the trials and, whether the subgroup imbalance (in baseline stroke severity) invalidates the entire trial as claimed by some of the critics."

The committee was also asked, as a secondary issue, to explore if "pharmaceutical company participation biased the results of the trial". The committee declined to consider this charge on the grounds that it was in no position to assess whether financial arrangements biased any of the parties involved in the study, approval and endorsement of t-PA

1.2 Principal Findings

The principal findings of the Review Committee are as follows:

1. Using the global statistic devised by the NINDS investigators and the GEE, we found that, despite an increased incidence of symptomatic intracerebral hemorrhage in t-PA treated patients and subgroup imbalances in baseline stroke severity, when the drug was administered according to the study protocol, there was a statistically significant, and clinically important, benefit of t-PA treatment measured by an adjusted t-PA to placebo odds ratio of 2.1 (95% CI: 1.5-2.9) for a favorable outcome at three months. The analysis was adjusted for center, time to treatment (0-90 minutes and 91-180 minutes), study part, age, baseline NIHSS, diabetes, and preexisting disability.
2. We examined all of the adjusting variables to determine if they modified the treatment effect of t-PA as measured by the adjusted t-PA to placebo OR. Our analyses found no evidence that any variable modified the t-PA treatment effect. In particular, neither baseline NIHSS, nor time from symptom onset to treatment, modified the t-PA treatment effect. Baseline NIHSS was analyzed both as a continuous and categorical variable, while time from symptom onset to treatment was analyzed as a dichotomous variable (0-90 minutes and 91-180 minutes) reflecting its role as a stratification factor in the design of the study.

1.3 Secondary Analyses

The Review Committee considered the following issues in their evaluation of the NINDS t-PA study:

1.3.1 Blood Pressure Assessment and Management

Our analysis identified a number of problems regarding pre- and post-randomization blood pressure measurement and management:

- Non-compliance with the defined protocol was substantial, and persistent, throughout the study with regard to both the documentation of blood pressure readings, and adherence to the treatment regimen for hypertension.
- There was limited rigor with regard to the pharmacologic characteristics of antihypertensive regimens. In some instances pharmacologic monitoring was performed

by representatives (nurses) of the sponsoring pharmaceutical firm. Medications employed were listed by date, but not by time, eliminating consequential interpretive utility.
- The exact number of patients who received medication to lower blood pressure either prior to, or after, receiving study treatment is unknown.
- The confusion regarding blood pressure documentation, and the lack of knowledge of treatment of hypertension either prior to, or after, receiving study treatment, could have led to an unknown number of patients receiving treatment in violation of the nominal study protocol.

Based on these observations, we reached the following conclusions:

- It was not possible to assess the effect of hypertension management on clinical outcome in acute ischemic stroke patients treated in the NINDS study.
- The blood pressure variables should not be included in the statistical models. However, we also found that inclusion of the blood pressure variables in the statistical models would have been inconsequential with regards to altering the t-PA treatment effect.

Finally, the inconsistent documentation of both blood pressure readings and hypertension management seriously undermines the NINDS investigators statement that blood pressure management was a significant part of the protocol that contributed to the success of the study. Nonetheless, we concur with the NINDS investigators premise that blood pressure management should be included in the protocol for treating acute ischemic stroke patients with t-PA. It is biologically plausible that hypertension management could affect clinical outcome in acute ischemic stroke patients treated with t-PA, and data from the cardiology literature has already demonstrated that in acute myocardial infarct patients, the risk of having an intracerebral hemorrhage is related to pre-treatment blood pressure. However, further clinical studies will be needed to assess whether blood pressure management is related to better clinical outcomes in acute ischemic stroke patients treated with t-PA.

1.3.2 Intracerebral Hemorrhage

In the NINDS trial, the overall risk of symptomatic ICH was 6.5% in t-PA treated patients vs. 0.6% in patients receiving placebo. When a symptomatic ICH occurred after treatment with t-PA, there were significant clinical consequences. Only a small minority had a favorable outcome (e.g., for the Barthel index, the favorable outcome in patients with symptomatic ICH was 10% vs. 55% in patients without ICH) and the three month mortality rate was very high (75%).

A number of putative risk factors for ICH were identified, with many of them being interrelated. Our exploratory analysis found four risk factors, age >70 years, baseline NIHSS >20 points, plasma/serum glucose >300 mg/L and edema and/or mass effect on the initial CT scan, that were associated with both an increased risk of having an SICH and a lower likelihood of having a favorable outcome. For patients with either no risk factors or only one risk factor, the likelihood of having a favorable outcome favored the t-PA treatment group, while for the group at highest risk (> 1 risk factor), there was essentially no difference between the t-PA and placebo groups with regards to the likelihood of having a favorable outcome. However, the analysis also found that the adjusted t-PA to placebo odds ratios for favorable outcome in the

three subgroups with different numbers of risk factors were not significantly different, and were consistently in favor of the t-PA treatment group.

We conclude that there was no statistically significant evidence of the existence of any subgroup of acute ischemic stroke patients in whom the risk, and consequences, of having a symptomatic ICH clearly outweighed the beneficial effects of t-PA. However, it is important to keep in mind that because of the study design and the small number of patients who had an SICH, this trial was not powered to identify risk factors related to having either an SICH or a decreased likelihood of a favorable outcome. Risk factors for ICH acute ischemic stroke patients treated with t-PA should be evaluated in future studies that are designed, and powered, to evaluate this question.

How the findings of this exploratory analysis are used in the management of the individual patient with acute ischemic stroke, balancing risks and benefits based on very limited scientific information, is for the patient and the attending physician to decide.

1.3.3 Baseline NIHSS Imbalance

After a thorough evaluation of this issue, we found no evidence that the imbalance in the distribution of baseline NIHSS between the treatment groups had a either a statistically or clinically significant effect on the study results, We further believe that the original models using both Age and baseline NIHSS as continuous variables properly adjust for the complex role played by these two variables, both strongly (negatively) related to the likelihood of a favorable outcome. There was a strong interaction between age and baseline NIHSS with respect to both the global analysis and the analysis of each of the four outcome measures. The likelihood of a favorable outcome was particularly low in patients older than 70 who had a baseline NIHSS score above 20. However, there was no evidence of any Age by baseline NIHSS subgroup responding significantly differently to t-PA treatment than the study group at large.

1.3.4 Baseline Stroke Severity and Age

This analysis found evidence that age, baseline stroke severity as assessed by the baseline NIHSS score, and the interaction between age and baseline NIHSS, were related significantly in a negative manner to the likelihood of a favorable outcome. We believe that the original models using both Age and baseline NIHSS as continuous variables properly adjust for the complex role played by these two variables. There was a strong interaction between age and baseline NIHSS with respect to both the global analysis and the analysis of each of the four outcome measures. The likelihood of a favorable outcome was particularly low in patients older than 70 who had a baseline NIHSS above 20. Patients with minor symptoms at baseline (NIHSS 0-5) had similar high odds for favorable outcome whether or not they were treated with t-PA. However, there was no statistical evidence of any Age by baseline NIHSS subgroup responding significantly differently to t-PA treatment than the study group at large.

1.3.5 Onset to Treatment Time

Based on the substantially nonlinear nature of the distribution of time from symptom onset to treatment (OTT), and an idiosyncratic distribution of favorable response rates among the placebo patients, we conclude that the data provided by this study failed to support a conclusion that the effect of t-PA therapy diminished with increasing values of OTT within the protocol specified 3 hour time limit. However, this does not mean such a relationship does not exist, and further studies are needed to address the question of a differential t-PA treatment

effect related to time from symptom onset to treatment. It is also important to recognize that the results from this study provide no data on the effectiveness of thrombolytic therapy administered to acute ischemic stroke patients more than 180 minutes after symptom onset.

1.3.6 Clinical Centers

We found no significant difference between the centers in the baseline characteristics of the patients. The likelihood of having a favorable outcome differed considerably between the centers, those with fewer patients often having the worst outcome. However, the between-center variation in t-PA treatment effect for either the global outcome, or the individual outcome measures, was not statistically significant and did not invalidate the trial results. Nevertheless, it will be important in future studies to identify the factors that lead to good outcomes at institutions administering t-PA to treat acute ischemic stroke patients. This information will be very helpful to other institutions that are looking to develop the resources needed to administer t-PA safely to acute ischemic stroke patients.

1.3.7. Stroke Subtype

We conclude that it was appropriate that stroke subtype was not included as a covariate in the analytic models. Further, we conclude that the data of this trial do not support any claim regarding either the presence, or absence, of a differential t-PA treatment effect within stroke subtype.

1.3.8 Preexisting Disability

Although patients with a preexisting disability had a significantly reduced chance of experiencing a favorable outcome, there was no evidence that they responded any differently to t-PA therapy than those without a preexisting disability.

1.3.9 Diabetes Mellitus

The observed data, and the adjusted estimated t-PA effects, indicated a strong benefit for patients without diabetes mellitus (DM), but no benefit among patients with DM. However, this comparison must be treated cautiously because there was no statistical evidence of a t-PA*DM interaction. The trial found no statistically significant evidence that diabetic and non-diabetic acute ischemic stroke patients responded differently to t-PA therapy.

1.4 Issues in Need of Further Investigation

The NINDS t-PA trial was a prototype study of acute ischemic stroke that demonstrated a beneficial effect of thrombolytic treatment with t-PA when administered within three hours of the onset of stroke symptoms. The study was designed to show differences in the entire group of eligible patients and not in subgroups. The exploratory analyses conducted previously by the trial investigators, and now by us, found a number of issues that need to be explored further so that t-PA can be used confidently by a broad range of practitioners in routine clinical practice. Analysis of these issues could be done by either conducting new large-scale clinical trials, or combining primary data from all the t-PA in ischemic stroke trials that have already been conducted. Both strategies are ongoing.

Based on the findings of this review committee, some of the most critical questions that need to be addressed are:

- Is there a subgroup of patients with ischemic stroke in whom the risk for intracerebral hemorrhage is so high that the group as a whole has no benefit from t-PA treatment? Candidate high risk factors are; age > 70 years, baseline NIHSS > 20, high glucose levels, and signs of edema or mass effect on CT.

- Is there a subgroup of patients with only mild symptoms in whom t-PA provides no net benefit?

- Within the time frame of the NINDS trial (treatment within 180 min), is there evidence of a differential t-PA treatment effect related to time from symptom onset to treatment?

- What is the impact of elevated blood pressure, and its management, before and after t-PA treatment on clinical outcome?

- Can data from other trials be used to validate the cut-off for t-PA treatment used by the NINDS investigators (blood pressure <=185/110)?

- Can the exploratory analysis finding in the NINDS trial that stroke patients with diabetes do not benefit from t-PA treatment be confirmed?

1.5 Conclusion

The committee concluded that, despite an increased incidence of symptomatic intracerebral hemorrhage in t-PA treated patients and subgroup imbalances in baseline stroke severity, when t-PA was administered to acute ischemic stroke patients according to the study protocol, there was a statistically significant, and clinically important, benefit of t-PA treatment resulting in a higher likelihood of having a favorable clinical outcome at three months.

2. INTRODUCTION

2.1 Background: In 1995, a group of investigators, the NINDS rt-PA Stroke Study Group, published a seminal manuscript summarizing the results of two studies of t-PA as a therapy for acute ischemic stroke[1]. Prior to the publication of this manuscript, the study group had conducted several investigations in preparation for the performance of the two pivotal studies[2-4]. These investigations involved pilot studies of the use of t-PA, studies of the reliability of the NIH Stroke Scale[5] and the Barthel scale[6] in the setting of a clinical trial, and a study of the factors related to the risk of intracranial hematoma formation in patients being treated for ischemic stroke with t-PA[7].

Subsequent to the 1995 publication[1] the FDA considered and approved an application from Genentech for the approval of t-PA as a therapy for acute ischemic stroke when administered according to the NINDS protocol. In the meantime, the NINDS rt-PA Stroke Study Group published a series of manuscripts designed to: i) elucidate their methods[8], ii) refine their message regarding the therapeutic efficacy of t-PA[9], iii) examine the long-term consequences of therapy[10], v) consider the factors affecting the risk of ICH[11], and, v) describe the frequency of pre- and post-treatment hypertension and the effect of its management[12].

As t-PA was used in emergency departments around the country, results were not as universally successful as anticipated, and doubts began to arise. Eventually, these doubts were expressed in the form of publications[13-17], and a commentary[18]. As a result of the concerns raised by these publications, NINDS appointed this t-PA Review Committee.

2.2 Announcement: The following announcement of the creation of this Review Committee appeared in the October 2003, *NINDS Notes*

> "The NINDS recently invited an independent committee to review and consider the data from the five-year, multi-site "Tissue Plasminogen Activator for Acute Ischemic Stroke," published by the NINDS r-TPA Stroke Study Group in the New England Journal of Medicine, December 14, 1995. The study represents the first treatment for acute ischemic stroke, and the therapeutic agent t-PA was approved by the FDA for this usage in June of 1996.
>
> In recent months, public debate about the study findings has resulted in some discussion within the medical community about the appropriate use of this treatment for stroke. In answer to this, the NINDS asked that the committee "address whether there is concern that eligible stroke patients may not benefit from rt-PA given according to the protocol used in the (NINDS) trials, and whether any subgroup imbalances invalidate the trial as claimed by some of the critics."
>
> The committee is chaired by Dr. W. Michael O'Fallon, Ph.D., Professor of Biostatistics and former Chair of the Department of Health Sciences Research at Mayo Clinic, Rochester, Minnesota. Dr. O'Fallon chose the members of the committee, who represent an international cadre of physician-scientists with expertise in biostatistics, clinical medicine, cerebrovascular disease, neurology, and emergency medicine. None of the committee members has a connection with the previous published study or with the manufacturer of t-PA. (See attached sheet for a roster of the committee members.)
>
> The committee has full access to the study data, will re-analyze the study, and hopes to report its findings by early spring. The NINDS looks forward to the group's findings and the presentation of the data analysis at professional meetings and in the scientific literature."

2.3 Charge: The actual charge to the Review Committee, delivered by Dr. John Marler on May 24, 2002, read as follows:

> "As the effort to implement the acute stroke care guidelines resulting from the publication of the results of the NINDS rt-PA Stroke Study has proceeded, increasing scrutiny of the results has occurred. One group in particular has recently raised concerns about the implications of an imbalance in the severity of the baseline stroke between different subgroups for the two treatment arms.
>
> I would like the committee to address whether there is concern that eligible stroke patients may not benefit from rt-PA given according to the protocol used in the trials and, whether the subgroup imbalance invalidates the entire trial as claimed by some of the critics. The issue of whether pharmaceutical company participation biased the results of the trial is an important, but secondary issue for the group.
>
> The committee will have full cooperation and access to the data in any manner that they wish for their own independent analysis or for analysis by the statistician from the trial."
>
> John Marler
> NINDS/National Institutes of Health

2.4 Membership: As described above, in May of 2002, Dr. Marler of NINDS invited Dr. O'Fallon to appoint a committee that would be viewed as fair and objective to both advocates and critics of the NINDS t-PA trials. The NINDS did not participate in the selection of any member other than Dr. O'Fallon and did not review the credentials of the members he selected. Furthermore, neither NINDS staff nor original NINDS rt-PA Stroke Study Group investigators participated in any of the meetings of the committee. They communicated with the committee and/or Dr. O'Fallon only rarely and then at the committee's invitation. The committee members, who represent an international cadre of physician-scientists with expertise in biostatistics, clinical medicine, cerebrovascular disease, neurology, and emergency medicine, were paid as hired consultants to an independent contractor to NINDS. None of the committee members has a connection with the previous published study or with the manufacturer of t-PA.

The committee consists of three clinicians (Drs. Kjell Asplund, Lewis Goldfrank and Timothy Ingall) and three statisticians (Dr. O'Fallon and Drs. Vicki Hertzberg and Thomas Louis). Full titles and affiliations are listed on the Committee Roster, a component of the Title Page of this report. The three statisticians are well acquainted, but had not previously collaborated on research projects or worked at the same institutions. Dr. O'Fallon recruited Dr. Ingall with whom he had worked when Dr. Ingall was a Neurology Fellow at Mayo Clinic in Rochester, Minnesota, but otherwise the three statisticians did not know the clinicians. Drs. Ingall and Asplund, are acquainted, having collaborated in the analysis of the WHO MONICA project, but have not been colleagues and Dr. Goldfrank, was not known to any of the other committee members.

Although Dr. O'Fallon had investigated the epidemiology of stroke, he had not participated in any of the studies or trials leading up to the NINDS-supported investigations regarding the use of t-PA as a therapy for acute ischemic stroke. Dr. Hertzberg, has participated in stroke related research, but none involving the investigations being appraised. Dr. Louis, had no experience in stroke research but has an extensive background in clinical trials, most recently in HIV-Aids. The three MDs have active research careers and, importantly, are practicing physicians whose professional responsibilities necessitate an intimate understanding of

appropriate assessment and management workup and therapy for individuals experiencing an ischemic stroke.

3. COMMITTEE PROCESSES

3.1 NINDS' Independent Contractor: The NINDS has a contract with an independent contractor in the Washington DC area, to maintain data sets, monitor the activities of investigators and establish contractual relationships with ad hoc groups engaged in small studies. The review committee acted independently of NINDS through this contractor, which provided the financial support required by the committee. This independent organization has been responsible for archiving data from completed studies sponsored by NINDS and thus provide the data from the t-PA studies to Dr. O'Fallon at Mayo for the committee's analysis.

3.2 Communications: Except for one in-person meeting (March 22, 2003), the considerable communication necessary among committee members was conducted via telephone and e-mail. Regularly scheduled conference calls were established and documented by approved minutes. The first conference call was held on June 4, 2002 and calls were held every two or three weeks until late fall, 2002, since which time weekly calls were held. A member of the contractor's staff joins these calls, records them and prepares draft minutes. The minutes are circulated electronically, reviewed and approved at subsequent calls. The contractor maintains and archives the conference call minutes as well as the exchanges of analyses among the reviewers.

3.3 Guiding Principles: The Committee was established within the framework of the following principles:

- In all interactions, openness and candor were encouraged and respected.
- The committee's work must be performed independently of NINDS, Genentech and the investigators involved in the original studies.
- The committee must have unhindered and complete access to the original data upon which the published manuscripts and the FDA approval were based. The evaluation required analysis of the original data; it could not depend solely on reading the literature and arriving at a conclusion.
- The committee must be in control of its data analyses. To this end it was necessary to arrange for the data to be made available to a data analyst from the Division of Biostatistics at Mayo Rochester, who worked under the guidance of Dr. O'Fallon.
- The committee requested that the scientific community be made aware of its existence and charge.
- The committee declined to consider the "secondary issue" in the charge on the grounds that it was in no position to assess whether financial arrangements biased any of the parties involved in the study, approval and endorsement of t-PA

3.4 Timeline:

~May 1, 2002	O'Fallon appointed as Chair & asked to form a committee
May 20, 2002	Committee of 5 completed
May 24, 2002	Charge to Committee issued by Dr. Marler of NINDS

June 4, 2002	First Conference call
June 18, 2002	Committee budget proposed to NINDS
August 6, 2002	Final member joins the committee
August 27, 2002	Discussions regarding an alternative analytic support plan
Sept. 5, 2002	Arrangements for Mayo to provide analytic support
Sept. 15, 2002	Data made available to Mayo
Sept. 26, 2002	First Mayo replication of results
Oct. 25, 2002	NINDS Notes announces the committee's charge and composition
Nov. 15, 2002	Committee decides not to participate in NINDS sponsored "Stroke Symposium" (12/12/02)
Nov. 16, 2002	BMJ Reporter & Article
Nov. 21, 2002	NINDS Investigators Tilley & Brott join committee conference call regarding blood pressure management issues and data
Jan. 2003	Additional blood pressure data made available
Jan. 2003	Exchanges with Drs. Marler & Penn regarding publications and presentations
Feb. 6, 2003	NINDS investigators participate in another conference call regarding blood pressure management issues and data
~March 1, 2003	Committee sends abstracts to European Stroke Conference (ESC) and Society for Academic Emergency Medicine (SAEM)
March 22, 2003	Committee meets to complete the framing of its final report.
May 24, 2003	Presentation at ESC, Valencia, Spain
May 29, 2003	Presentation and Panel discussion, SAEM, Boston, Massachusetts
June 10, 2003	BMJ Reports on SAEM presentation
July 31, 2003	Final Report submitted to Dr. Marler
Subsequently	Manuscript prepared for publication and presentations made in Europe and Australia

4. METHODS

4.1 Data Management

4.1.1 Data from NINDS' Independent Contractor: In September 2002, the committee received a CD from the Institute's independent contractor labeled "NINDS t-PA Stroke Study Data Collection." It contained a 43 page descriptive document, a main directory with 114 SAS datasets and 81 SAS programs. There was also an ancillary directory with 112 additional files. We examined all 114 SAS datasets containing a total of 4,795 variables (with many variables being in multiple datasets) to discern the variables of interest. Many of these datasets contained 624 observations, but several had more than 10,000 observations. All the variables we used in these analyses were found in one of five main datasets. In January 2003, we received 8 additional SAS datasets upon request pertaining to blood pressure. These datasets were not used in our analyses.

4.1.2 Variable Identification and Definition: As described in 4.1.1, data were obtained from an independent contractor. Definitions of the primary variables were obtained and are presented in Table 4.1. In general, this was a straightforward process, but where issues arose we contacted the contractor and, on occasion, the original study statisticians, programmers, or investigators for clarification.

In Table 4.2, which will be used to assess balance between the t-PA and Placebo groups in Section 5.1, we summarize the 64 variables assessed at or prior to the time of randomization. We have adopted a standard nomenclature that is as transparent as possible. In addition to the variable name, we distinguish timing by the three prefixes: i) "Pr" to indicate a determination (often a diagnosis) made prior to the stroke, ii) "Ad" to indicate measurements/determinations made at admission to the Emergency Department for treatment of the stroke, and iii) "Bs" for baseline to indicate measurements/observations made between admission and randomization. Time constant variables (e.g., age, sex, race) do not require a prefix and some prefixes are somewhat arbitrary. For example, "AdAspirin" indicates whether or not the patient had been taking aspirin as a regimen up to the time of the stroke. A person who had discontinued such a regimen before stroke would be coded "no." We coded diabetes "PrDM," indicating a prior diagnosis of DM rather than an indicator of elevated glucose at arrival to the ED.

Many variables are dichotomous, indicating the presence or absence of some characteristic. Usually, we use "1" for presence and 0 for absence. If the coding might be unclear, we indicate the coding rule in ( ) after the variable name. Thus, sex (male) means that we coded males as 1. All dichotomous or polychotomous variables are summarized as percents rounded to the nearest one decimal. If a variable is continuous, we indicate the units of the measurement and in Table 4.2 the variable is summarized by its median.

4.1.3 Result Replication: We undertook to replicate results reported in several of the published manuscripts. Results of this replication process will be discussed in Section 5.1, Specifically, we replicated:

1995[1]: Tables 1, 2, 3, & 4

1997[11]: Table 1

1999[10]: Table 3

2000[19]: Table 3

4.2 Study Design

4.2.1 Stratification Factors: Evaluation of design, conduct, analysis, and interpretation issues was restricted to the committee's charge. The original investigators published multiple manuscripts and it was not the purview of the committee to assess and judge all aspects of each of these manuscripts.

The primary manuscript[1] describes two studies, referred to as Parts 1 and 2. The investigators describe Part 1 essentially as a Phase 2 study that seeks to determine whether the agent had activity. However, it was a randomized, placebo controlled, study designed to address the evidence for t-PA activity with respect to outcomes assessed 24 hours after stroke onset. Part 2, designed to assess results 90 days after stroke onset, was designed exactly as Part 1. The investigators, essentially the same as for Part 1, were blinded as to the results of Part 1 until Part 2 was complete. The studies were placebo controlled, randomized clinical trials (RCTs). Both were conducted at approximately 8 clinical centers with independent randomization at each. Randomization was stratified and balanced at each center according to whether the patient was randomized within the first 90 minutes or in the 91-180 minute interval after stroke onset. Patients whose time since onset had exceeded 180 minutes were ineligible. Furthermore, the investigators conducted assessments in both studies at 24 hours, 90 days and 1 year after stroke onset.

Consequently, as proposed by the investigators, Parts I and II can be treated as independent, replicate studies. For analytic purposes, we treated the two studies as a single, large RCT with three stratification factors, Part (1 or 2), Center, and onset to treatment time, (OTT: 0-90 or 91-180 minutes). After detailed examination of outcome data at 90 days and 1 year, the review committee decided to restrict its analysis to the outcome assessment at 90 days after stroke onset of all the patients in the two studies.

4.2.2 Outcome Measures: The NINDS investigators used four outcome measures. The Barthel index, modified Rankin scale, and Glasgow outcome scale are accepted as measures of functional status. The NIH Stroke Scale (NIHSS) is accepted as a measure of neurologic deficit. The primary response variable for each measure was a dichotomous indication of whether the outcome (at 90 days) was "favorable" or "not favorable." The definitions of "favorable" were: Barthel; 95 or 100, Rankin; 0 or 1, Glasgow; 1, and NIHSS; 0 or 1. For each measure death was treated as an unfavorable outcome. Since the measures assess different aspects of the consequences of stroke, they are neither completely congruent nor statistically independent. Therefore, the committee evaluated comparisons of the Placebo and t-PA treatments for each of the measures individually.

In addition, the NINDS investigators constructed what they refer to as a "Global" indicator of a favorable response[1, 20]. This Global indicator is a 4-dimensional vector of the favorable/unfavorable indicators, for each of the 4 indices. Thus, each patient had a global response vector consisting of 4 elements, each either zero or one, with zero indicating an unfavorable outcome and one a favorable outcome. Those dead by 90 days had a global

response vector of the form (0, 0, 0, 0) while those whose outcome was favorable on all the measures had a global response vector of the form (1, 1, 1, 1). The review committee replicated the results of the global analyses reported by the investigators and carried out additional analyses deemed appropriate.

4.2.3 Intent to Treat: The study investigators used the principle of "intent-to-treat" in analyzing all patients randomized in the study. Thus, they attempted formal follow up at 24 hours, 90 days and one year on all randomized patients. Patients "lost" in the sense that they were known to be alive but did not provide data permitting the determination of favorable/unfavorable status were assigned the least favorable known level for each index (4, 11) with its consequent favorability status. With two exceptions, the review committee used the same approach. The two exceptions involved individuals mistakenly randomized into the study at a point more than 180 minutes after onset. Since an essential component of our charge was to determine whether there were groups of patients who should not be treated with t-PA according to the study protocol, we excluded these two patients from subsequent analyses.

4.3 Analytic Methods

4.3.1 Treatment Group Balance: In theory, the process of the completely random assignment of patients into one of two treatments within strata should produce nearly equivalent distributions of observed covariables (not treatment effect variables) in the two treatment groups. We will examine whether or not that happened for the variables listed in Table 2 using Chi-Squared tests for the dichotomous and polychotomous variables and rank sum tests for the continuous variables.

4.3.2 Missing Data Imputation: As is seen in Table 4.2, there were patients whose values of several of the variables were missing. Before any in depth analyses could be undertaken we elected to take the following actions:

1. We eliminated 5 variables from all subsequent analyses because they were each missing for more than 40 patients. From Table 4.2 it is seen that these five variables are: BMI, Prior Atherosclerosis, Prior Hyperlipidemia, Baseline fibrinogen and Prior TIA.
2. We imputed the other missing values essentially by sampling at random from the existing data. In this imputation, if a variable was categorical with some categories being observed less than 10% of the time we used 10% as the probability for that category and adjusted the most common category appropriately to assure that the sum of the several percents was 100.

For the logistic regression analyses, we performed "best case-worst case" imputation and replicated the random imputation process several independent times, running the regression models for each resulting data set. The distributions of parameter estimates were then examined to determine if any aberrations were observed. Seeing none considered critical, we ran one final random imputation thus creating an analysis data set including 622 patients each with a complete set of values for the variables to be considered in future analyses. It should be noted that this use of a single imputation sample will result in underestimated standard errors. However, the number of patients and variables for which imputation was necessary was small so the bias should be negligible. There was no specific evidence in the published material as to what, if any, actions were taken by the NINDS investigators in reaction to the

missing values, although there is an allusion in the FDA application in which the sponsor states: "An Intent-to-treat analysis was the performed, and the data imputation plan for missing values as devised by the NINDS Investigator group would be utilized"[21]. It is possible that our analyses might differ from theirs in minor ways as a consequence of the use of different imputation strategies.

4.3.3 Analytic Models: With the primary analyses focused on the "favorability" outcome, statistical models must be appropriate to the analysis of proportions (equivalently, probabilities or percents). Such data can be analyzed on one of three scales: the original scale examining the difference between two proportions, the ratio (log) scale analyzing the ratio (Relative Risk) of two proportions, or the odds (logit) scale analyzing the ratio of two odds (Odds Ratio). Each measurement scale has its advantages and disadvantages. The original scale is most clinically relevant, but the log and logit scales generally produce more parsimonious models with better understood statistical properties. The investigators reported relative risks (RR) and odds ratios (OR) when possible, but performed their most extensive analyses in the logit scale using univariate and multiple logistic regression models and reported the resulting odds ratio estimates.

Validated approaches and software are available to implement each approach. A Generalized Linear Model (GLiM) with the identity, log or logit "links" and "binomial" variation unified the approach. The GLiM can be used to compare two treatments with respect to an outcome measured on the probability scale while adjusting for stratification factors, confounding factors and even effect modifying factors, with model specification being essentially identical to that for standard, linear regression.

In analyzing the Global outcome measure, the investigators used the Generalized Estimating Equation model (an extension of a GLiM) with the logit link function and with the correlation structure estimated by the empirical, observed, correlations among the four indices[1, 20]. This analysis yields a general odds ratio estimate comparing the odds of a (global) favorable outcome in the t-PA treated group to that in the placebo group while adjusting for stratification and baseline factors. After determining that this was appropriate the review committee used these same models in its analyses.

4.3.4 Subgroup Analysis and Interaction Detection: In addition to evaluating overall results, the t-PA Review Committee was charged with considering the question of whether subgroups of patients might actually be harmed by the use of t-PA therapy. The investigators addressed that issue[9]. Possibly the FDA Advisory Committee considered the subgroup issue, however FDA approval was without conditions other than the restriction that the therapy be administered according to the protocol.

The sample sizes in Parts I and II were determined so that the study would have sufficient statistical power to detect a clinically relevant difference between t-PA and placebo. Neither study was powered to detect clinically important subgroup effects or treatment interaction effects. The combined studies still have low power for these investigations. Even though the power is low, a large number of evaluations are likely to generate some statistically (and apparently clinically) significant results even when the underlying truth is that no such treatment/subgroup relations are operating. Consequently, subgroup analyses and evaluations of interactions operate in a low power, exploratory context.

The NINDS investigators attempted to address the low power issue by performing the tests at a very generous p-value (0.2 & 0.1)[9]. While this certainly increases the power (decreases the chances of a Type II error), it does so at the price of an increase in the chances of a Type I error and may result in spurious "findings." The investigators quite correctly point out (and we concur) that such findings are best used as motivation for further studies designed specifically to address the issue raised by the identification of these interesting groups. The review committee has examined some of the potentially interesting subgroups in considerable detail, reports results and emphasizes the caveats and cautions.

4.3.5 Logistic Models:

The term "Odds" refers to the ratio of a probability to it's complement (e.g., P/[1-P]). In this report the term "odds ratio (OR)" always refers to the ratios of the odds of a favorable outcome in one group of patients to the odds of a favorable outcome in another group. Since favorable outcome is defined on 4 scales, it is essential that the appropriate scale be kept in mind, but do not incorporate references to Barthel, Rankin, Glasgow, or NIHSS in the "OR" notation.

Define

P[F | T] = probability of a favorable outcome on the treatment

and P[F | P] = probability of a favorable outcome on the placebo.

In this notation the odds ratio is:

$$\mathrm{OR} = \frac{\mathrm{P[F|T]/\{1\text{-}P[F|T]\}}}{\mathrm{P[F|P]/\{1\text{-}P[F|T]\}}}.$$

With "log" indicating the natural logarithm, the basic logistic model takes the form,

$$\text{logit} = \log(OR) = \beta_0 + \beta_1 X + \underline{\gamma}^t \underline{Z},$$

where X is a 0/1 indicator with "0" the comparison group, $\underline{Z}$ is a vector of covariates with the corresponding coefficient vector $\underline{\gamma}$. The OR adjusted for the covariates $\underline{Z}$ is estimated by $e^{\hat{\beta}_1}$. We use SAS Proc Logit to estimate the parameters β_0, β_1, and $\underline{\gamma}$.

Effect modification, for example by component Z_1 of $\underline{Z}$ (i.e., the influence of X on the OR is dependent on Z_1) is modeled by an interaction term:

$$\text{logit (favorable outcome)} = \beta_0 + \beta_1 X + \beta_2 X * Z_1 + \underline{\gamma}^t \underline{Z}.$$

We test the null hypothesis that $\beta_2 = 0$ to asses whether or not there is sufficient evidence to declare Z_1 an effect modifier. This type of question and test is critical to the question of whether baseline imbalances have influenced results.

In our logistic models, the vector $\underline{Z}$ of covariates must include all stratification variables and will include variables that are statistically significantly related to the likelihood of a favorable outcome. Note that imbalance of a variable between the t-PA and placebo groups does not necessarily imply that it must (will) be included in the model.

4.3.6 Global Analysis

The Global analysis, described in detail by the NINDS investigators[20] provides a powerful, multi-outcome approach to assessing the relation of baseline variables and treatment to the probability of a favorable outcome. This approach treats the four binary outcomes as a four-dimensional outcome vector which is then related to covariates much as in the basic logistic regression models. Correlation among the outcome measures must be taken into account. Software for estimating the parameters in such a comprehensive model was limited, when the original investigators conducted their analyses. Now, SAS Genmod and other implementations of the Generalized Estimating Equation (GEE) approach facilitate such analyses.

4.3.7 Analyses in the Probability Scale

As mentioned earlier, logistic regression methods are sometimes criticized because they are based on the odds scale. The fundamental results of logistic regression on a clinical trial analysis is an estimate of an odds ratio. This odds ratio (OR) relates the odds of a "success" (in this study, the odds of a favorable outcome at 90 days) among those on the active therapy (t-PA in this study) to the odds of a success in the comparison (placebo) group. Estimates of the <u>difference</u> between the probability of success in two groups rather than the odds ratio of probabilities does provide a more clinically relevant comparison.

We estimate the difference $\Delta = P[F \mid T] - P[F \mid P]$ rather than the logistic regression based odds ratio. Using the odds of a favorable outcome among the placebo treated patients to "represent" the status in the general stroke patient population, we estimate the odds ratio as described and then can estimate the difference between the two percents as follows: Define
K = odds of favorable outcome among placebo treated patients
OR = estimated t-PA to placebo odds ratio
Then the estimate of Δ is

$$\hat{\Delta} = \frac{K}{1+K}\left[\frac{OR-1}{1+ORgK}\right] \qquad \text{(Equation 1)}$$

In a more general context if patients are stratified into M groups, $G_1, \ldots, G_M$, we can represent the data by the following table.

	G_1		G_2		…	G_M		
	F	UF	F	UF		F	UF	
Placebo t-PA								n_1 n_2

In such a context, it is possible to extend the above formula to estimate the difference between the two proportions (probabilities) taking into account this sub-grouping and all other covariates. This formula is based on a logistic regression model which includes M-1 indicator variables distinguishing the M groups as well as all other covariates and, if necessary, M-1 variables accounting for the differential effects of t-PA in the M groups.

Define

1) $P[G_i] = n_i / n \quad i = 1, ..., M$

where n = total number of randomized patients

n_i = total number of randomized patients in group G_i.

2) Define the variable

$$T = \begin{cases} 1 \text{ for those randomized to t-PA} \\ 0 \text{ for those randomized to placebo} \end{cases}$$

3) For i=1, …, M-1, define variables:

$$X_i = \begin{cases} 1 \text{ for all patients in } G_i \\ 0 \text{ for all patients not in } G_i \end{cases}$$

Note: This designation of M-1 variables, X_i, specifies group G_M as the comparison group. That is completely arbitrary and in practice we will tend to use one of the groups containing a large number of patients.

4) θ = odds of a favorable outcome in the placebo treated patients from the group chosen to be the comparison group.

5) $\underline{Z}$ a vector of covariates with corresponding vector $\underline{\delta}$ of coefficients.

The most general logistic regression model takes the form:
logit (Favorable Outcome)

$$= \beta_0 + \beta_1 T + \sum_{j=1}^{M-1} \gamma_i G_i + \sum_{j=1}^{M-1} \gamma_{M+j} T * G_i + \underline{\delta}^t \underline{Z}$$

and the generalized formula for the weighted difference in the likelihood of a favorable outcome between the t-PA and placebo groups is

$$\hat{\Delta} = \theta\left[\exp(\beta_1) - 1\right] \sum_{j=1}^{M} \left\{ \frac{\exp(\gamma_j)}{\left[1 + \theta \exp(\beta_1 + \gamma_j + \gamma_{M+j})\right]\left[1 + \theta \exp(\gamma_j)\right]} \right\} P(G_j) \qquad \text{(Equation 2)}$$

The $\gamma_M = \gamma_{2M} = 0$ and if the interaction terms are not included $\gamma_{M+1} = \gamma_{M+2} = ... = \gamma_{2M} = 0$.

The standard errors of the difference estimators defined by equations 1 and 2 above were estimated using the Jackknife method described by Efron and Gong[22].

4.3.8 Covariate Determination (Stepwise Models): All baseline covariates available to the NINDS investigators (with the exception of baseline/admission blood pressure measurements, as explained in Section 6) were considered for initial inclusion in the models A forward stepwise selection process ($p < 0.05$ to enter and remain) was used to derive the final covariates for inclusion, after constraining the model to include the design stratification variables, CENTER, OTT, and PART. This modeling process was performed for each of the

outcome measures to derive a candidate list of covariates for consideration. A covariate was considered to be in the candidate list if it entered the stepwise process for at least one of the four outcome measures. In addition, all covariates in the candidate list were reviewed for clinical relevance, i.e., did the relationship make sense biologically. After arriving at this candidate list of covariates, these covariates were then screened for pairwise interactions, again using the forward stepwise selection process. From this process, any covariate or interaction (and any contributing lower order effects) was included in the final list of covariates if it remained in the model after this second stepwise screening process for any of the four outcome measures. A similar process was employed for the analyses described in Section 7 where the occurrence of an intracerebral hemorrhage was the endpoint.

5. RESULTS

5.1 Replication of Published Results: As indicated in Section 4.1.3, we selected tables from several of the NINDS investigators' publications[1, 10, 11, 19] and attempted to replicate them. In Tables 5.1 through 5.7 with the matched Tables 5.1a through 5.7a, respectively, we summarize our replication. There are nothing but trivial differences between our results (Tables 5.1 through 5.7) and the corresponding published results (Tables 5.1a through 5.7a), respectively. Thus, we concluded that we had access to the correct data and had defined the variables correctly so we continued with our planned analyses.

5.2 Baseline Balance Between t-PA and Placebo Groups: Table 4.2 was constructed to facilitate an examination of the balance at randomization of the distribution of the 64 variables which may be used in the analysis but were not specifically balanced by the randomization process. In general, randomization should yield a balance-on-average between the t-PA and Placebo groups. However, when many variables are assessed, chance alone will result in a statistically significant imbalance in some variables. The primary questions faced by the original investigators and the review committee must be whether any imbalances noted represent some excess above that expected by chance alone, whether such imbalances suggest an inherent flaw in the randomization process and whether observed imbalances confound treatment comparisons.

Many of the 64 variables included in Tables 4.1 and 4.2 are constructed from a common set of inputs and, consequently, the multiple "significant" p-values observed need to be considered with some care. We observe, as did the original investigators[1, 19], that there were imbalances in the following areas:

Age: - Placebo group somewhat younger than the t-PA group
Weight: - Placebo group somewhat heavier than the t-PA group
Aspirin - Fewer in the Placebo group on a daily regimen of aspirin than in the t-PA group
Baseline NIHSS - While the Placebo and t-PA group medians of baseline NIHSS (BsNIHSS) were not significantly different; when BsNIHSS was categorized as: 0-5, 6-10, 11-15, 16-20, and > 20, a significant imbalance was identified. Primarily, among patients in the 0-5 range, there was a greater proportion of patients randomized to t-PA than to placebo. It is with respect to this latter imbalance that much controversy regarding the study results has arisen.

Of course, the most critical question is whether or not an imbalance is so severe that any observed treatment effect could be explained by the imbalance (false positive effect) or any lack of observed effect could have been the result of the imbalance obscuring the effect (false negative). The NINDS investigators concluded that, regarding the above noted imbalances, neither instance seemed likely. We describe our investigations of this issue in the following sections.

5.3 Observed Outcomes: Table 5.8 contains the observed data regarding favorable outcomes for each of the 4 outcome measurements. The 622 patients are divided into the 310

treated with t-PA and the 312 treated with the Placebo, and are classified further as to whether they had had a favorable 90-day outcome according to each of the outcome measures. Results are summarized in 3 ways. For each outcome measure the difference between the percents favorable for the t-PA and Placebo groups, the ratio of these percents, and the corresponding odds ratios are all presented. For all of these comparison scales for each of the outcome measures, these data summaries show that the t-PA treatment is significantly more likely to produce a favorable 90-day outcome than the Placebo and the estimated treatment effect is clinically important.

In evaluating our covariate adjusted analyses (Section 5.5) it will be important to refer back to these unadjusted results. Assuming that randomization was properly conducted, these results are valid. For the four outcome measures, the proportion expressing a favorable outcome in the t-PA treated group exceeds that proportion in the Placebo group by between 13.7% and 16.3%. These differences indicate that if 1000 ischemic stroke patients received t-PA therapy according to the NINDS protocol, about 150 more of them would experience a favorable outcome than if t-PA had not been available or used. The four odds ratios, ranging from 1.78 to 2.07, are all significantly different from one, again suggesting that t-PA is more likely to produce a positive outcome than Placebo. It will be informative to consider the effect of the adjustments relative to these basic estimates.

Table 5.9 contains three sub-tables showing the joint distribution of the 4 dichotomous outcome variables so that their interrelationships can be examined. The first subtable shows the entire cohort of 622 randomized patients classified into the 16 possible categories. Here we see that 325 (52%) of the patients failed to have a favorable outcome on any of the 4 outcome measures which means that 48% had a favorable outcome on at least one of the outcome measures. At the other extreme, there were 151 (24%) of the randomized patients who had a favorable outcome on all of the 4 measures.

These two extremes suggest two straightforward ways to combine the 4 outcome measures into two consolidated outcome scales. In the lower two tables we see that among the patients treated with t-PA, 169 (54.5%) had at least one favorable outcome while among the Placebo patients 128 (41.0%) had at least one. So, in the “at least one” scale, t-PA is better than Placebo by 13.5% with an OR of 1.72. For the more rigorous condition of having a favorable response on all 4 of the outcome measures, 98 (31.6%) of the t-PA treated patients achieved that level while only 53 (17.0%) of the Placebo patients did. This represents a difference of 14.6% in favor of the t-PA treated patients with a corresponding OR of 2.26.

The global analysis described by the NINDS investigators[20] is a more sophisticated way of combining the four outcome measures. Because the four measures are correlated, combining them is not equivalent to simply increasing the sample size by a factor of 4. However, because they are not perfectly correlated, combining them brings more information through the global analysis than is contained in any analysis of an individual outcome measure. As a consequence, the global analysis is more powerful than the individual analyses, as emphasized by the NINDS investigators[20].

5.4 Covariate Selection: In this section we describe the process of developing the models that account for the study design and covariates used to adjust estimates of the t-PA to

Placebo odds ratios and the corresponding differences in the probability of a favorable outcome. In subsequent sections we use these tools to address several critical issues among which will be the following.

1. Did the BsNIHSS imbalance bias the treatment comparison in a critical way?
2. Does the increased risk of ICH among t-PA treated patients put in question the value of t-PA as therapy for acute ischemic stroke patients? In particular, are there subsets of patients in whom the risk and consequences of ICH outweigh the benefits of t-PA?
3. Do the data support an informative analysis of the impact of the time from onset to treatment on the efficacy of t-PA therapy?
4. Is the t-PA benefit consistent among the several centers involved in the study?

5.4.1 Logistic Analysis of Favorable Individual Outcome: The results of the first stage in the process of selecting the covariates to be included in the outcome models are summarized in Table 5.10. All of the variables in Table 4.2, except for those with a large number of missing values (Section 4.3.2) and for blood pressure measurements reported as made at admission or baseline (for reasons described in Section 6), were considered as potential covariates. For each of the four outcome measures, each variable was considered separately in a logistic model of favorable outcome constrained to include the stratification variables of PART, CENTER, and OTT. The top part of Table 5.10 lists the stratification variables and those variables that had a p-value <0.20 for association with a favorable outcome for at least one of the outcome measures. The variables are ranked in order of their level of significance within the Barthel model. Thus, in these analyses of one potential covariate at a time, 18 of the potential covariates have p-values <0.20 for at least one of the outcome measures and 9 have p <0.20 for all four of the outcome measures. Not surprisingly, baseline NIHSS (BsNIHSS) in either of two constructs, AGE, and evidence of a preexisting disability (PrDISAB) are all highly (negatively) related to a favorable outcome for all 4 outcome measures.

The lower half of Table 5.10 illustrates the results of a forward stepwise process (p <0.05 to enter and remain). The three variables, BsNIHSS (as a continuous variable) AGE and PrDISAB enter, in that order, for all four of the outcome measures. Seven other variables enter for at least one but not all of the four models. Some of these variables – most notably weight – entered these models even though their univariate p-values (at the onset of the stepwise process) were not <0.20.

The next stage of the process of identifying covariates to be included in the outcome models is illustrated in Table 5.11. Here, the top part of the table illustrates the four separate models, including all of the variables that entered in the aforementioned stepwise process. All of the potential interactions among those variables included in the models were made available as candidates to enter the model in another stepwise process (with p <0.05 to enter and remain). For each of the outcome variables the interaction between AGE and BsNIHSS (AGE*BsNIHSS) was highly significant and was included in all subsequent models. The role of this interaction between stroke severity and age will be discussed in Section 8.1.5.

Two other interactions entered with the Rankin score and one of them also with the Glasgow score. The resulting models are presented in Table 5.12. The most interesting aspect of Table 5.12 is the different impact of the inclusion of the AGE*BsNIHSS interaction within each model. Understandably, with the inclusion of AGE*BsNIHSS in a model containing AGE and BsNIHSS some impact on the two "main effect" terms is expected. What is seen is that for the Barthel index, and to a lesser degree for NIHSS, nearly all of the impact of AGE and BsNIHSS is found in the interaction term. In contrast, for the Rankin and Glasgow scores both of these main effect terms retain significance in the presence of the interaction term. Although not visible in this table (see Tables 5.17. through 5.21), another interesting aspect of this interaction term is that its coefficient is negative. Since increasing values of the two variables decreases the chance of a favorable outcome, this negative coefficient indicates that they are synergistic in their interaction with each other. This is actually a somewhat uncommon phenomenon since advancing age often overwhelms other factors regarding the effect of a disease. This will be discussed further in Section 8.1.

Following an argument described in Section 5.5.2, we decided that all of our models of treatment effect will include as covariates the three stratification factors, four main effects (BsNIHSS, AGE, PrDISAB and PrDM) and the AGE*BsNIHSS interaction. However, the next stage of the process is to use the information gleaned from the analyses of the individual outcome measures to develop a Global model.

5.4.2 Global Model of Favorable Outcome: The first stages of the process of building a Global model are illustrated in Table 5.13. To the left, with OTT, PART and CENTER fixed in all models, is a summary of the independent impacts in a global model of each of the variables that either had a p-value less than 0.20 or were potentially interesting for other reasons. The stepwise process, which is performed automatically for the logistic models, is, of necessity, performed one-variable-at-a-time in the Global model. The top part of the right side of Table 5.13 summarizes the order in which seven variables "entered" the model in this process. The first three of these seven variables are the same as the three that entered all of the models for the individual outcome measures. The bottom part of the right side of Table 5.13 shows what would be the final model if no interactions entered this global model.

Since the AGE*BsNIHSS interaction seemed certain to enter the global model, we began the process of looking for interactions among the covariates in the global model by entering that interaction into the model. The results are described in Table 5.14. Here, the introduction of that interaction term changes the p-value of one of the existing variables, BsED/ME, to be greater than 0.05. We thus applied a backwards removal process, removing that variable and, subsequently, two more of the original seven variables. This left a Global covariate model – Table 5.14 – with three stratification factors, four main effects (BsNIHSS, AGE, PrDISAB and PrDM) and the AGE*BsNIHSS interaction. The coefficient estimates and their standard errors for these covariates in the global model are also included in a small subtable of Table 5.14.

5.4.3 Primary Covariate Model: For the sake of uniformity, we declared the covariates described above to be the covariates to be included in all models used in subsequent treatment comparisons, those for each of the outcome measures as well as for the Global analysis. Having so declared our "final" covariate model, we again examined all interactions among them for each outcome measure, as well as for the Global model. We found no other

interactions of consequence so all further analyses are based on the comprehensive covariate model described above. The first of these analyses are discussed in Section (5.5).

5.5 Model-Based Treatment Comparisons: In Section 5.3 (Tables 5.8 & 5.9), we described the data regarding the comparison of t-PA therapy to Placebo in the most fundamental terms. For the 4 outcome measures of Barthel, Rankin, Glasgow and NIHSS, the actual counts of patients experiencing a favorable outcome resulted in odds ratio estimates of 1.78, 2.07, 1.85 & 2.01 respectively. The corresponding, unadjusted Global estimate is 1.88. These were all highly significantly different than 1, ($p < 0.0001$) indicating that t-PA is superior to Placebo insofar as the likelihood of a 90-day favorable outcome was concerned.

5.5.1 Logistic and Global Model Results: Subsequent to our examination of the fundamental data, we examined variables potentially related, either positively or negatively, to the prospects of a favorable outcome. As discussed in Section 5.4, we have identified several variables that significantly influence outcome and, consequently, should be included, along with the stratification variables, as covariates in any model-based estimates of the OR.

Tables 5.15 & 5.16 summarize the evolution of the process of estimating a t-PA to Placebo odds ratio (OR) for each outcome measure (Table 5.15) as well as for the global analysis (Table 5.16). As the estimation process became more sophisticated and complete through the use of models that "adjusted" the OR estimates first for the stratification factors alone and ultimately for the stratification factors and the covariates of BsNIHSS, Age, PrDiabetes Mellitus, and PrDisability, the adjusted OR estimates became 2.19, 2.43, 2.13, 2.19, & 2.13, respectively. These adjusted OR estimates are numerically larger and statistically more significant than their unadjusted counterparts.

On the basis of similar analyses, the NINDS investigators concluded that t-PA, when administered according to the NINDS protocol is significantly superior to Placebo[1, 11]. **The review committee concurs with this conclusion.**

5.5.2 Treatment by Covariate Interactions: Before the general conclusion stated above can be considered valid, we must examine for each of the outcome measures as well as in the global analysis whether any of the covariates in the model directly moderated the effect of t-PA. Such moderation could be synergistic (enhancing the t-PA effect) or antagonistic (depressing the effect of t-PA). As indicated in Section 4.3.3, such "effect modification" is assessed by the inclusion of appropriate interaction terms in the logistic and Global models. Only in the absence of terms that are large relative to the main effects, are we in a position to report, without qualification, a universal statement of the evidence regarding the effectiveness of t-PA after "adjusting" for the presence of a number of covariates. If the interaction effects are such that all estimates of treatment comparisons are in the same direction a general statement might be possible, but, even then, care in interpretation is essential (Section 4.3.6).

5.5.2.1 Results of t-PA by Covariate Interaction Tests: In a series of five tables (Tables 5.17, through 5.21),for the analyses of the four outcome measures; Barthel, Rankin, Glasgow, NIHSS and the Global analysis respectively, we provide extensive summaries of the analytic models. In each of these tables, the first two results columns, labeled "Estimate" & "std. Error", provide the estimates of the coefficients of each of the covariates within a strictly covariate model. Here we see the negative coefficient on the Age*BsNIHSS interaction term alluded to

earlier. No p-values are provided here because they have been presented in Tables 5.12 & 5.14.

In the next set of three columns we see the primary adjusting models leading to the adjusted Odds Ratio estimates seen in Tables 5.15 and 5.16 which is obtained by inserting the treatment variable (t-PA) into the covariate model. These estimates are obtained by exponentiating the t-PA coefficient (e.g. for the Barthel model OR = exp(.7837)). In these models we also note that the coefficients of the covariates are changed only a little by the addition of the treatment variable t-PA into the models.

The remainder of the sets of columns summarizes the testing of interactions between t-PA & Covariates (including the stratification factors) within each model. There are 4 dichotomous covariates, each with a single degree of freedom, and 2 polychotomous covariates with multiple degrees of freedom. The interaction between t-PA and the dichotomous covariate have a very direct interpretation so we elect to discuss them in more detail and, specifically, to examine the power of the tests that are performed within the models to determine if the interactions are significant and need to be included in the models.

If there is a dichotomous variable that interacts with the treatment variable, the treatment by placebo odds ratio, our basic indicator of a treatment effect, is different depending on whether the covariate is absent (coded 0) or present (coded 1). In such a situation it is not possible to refer simply to "a treatment effect" because there are two different ones. In the modeling process, the interaction between t-PA & a Covariate is tested by inserting the product of the t-PA indicator and the covariate, with regression slope θ, into the model and testing whether $\theta = 0$. This test involves estimating θ and its standard error. In Tables 5.17 through 5.21, we summarize 20 such tests by reporting the estimates of θ, the standard errors and the corresponding p-values. None of the p-values are <0.05, so we report that these interactions are not significant and do not retain them in the models when we summarize the treatment effect. However, tests of no interaction have notoriously low power, a point we will examine in a moment.

The interpretation of θ is summarized by:

$$e^{\theta} = \frac{OR(\text{cov} = 1)}{OR(\text{cov} = 0)},$$

where OR(cov = 1) is the t-PA versus placebo OR in the presence of the covariate, and OR(cov = 0) is the OR in the absence of the covariate.

Only when $\theta = 0$ is the ratio of ORs equal to 1, indicating that the effect of t-PA is the same whether the covariate is present or not.

5.5.2.2 Power of Interaction Tests: As mentioned, we report all 20 tests of no interaction of t-PA with dichotomous covariates as being not significant (p > 0.05). But, this study, as is the case for most clinical trials, was not designed to have much power to assess such interactions. In the tables below, we report how large the ratio of the two ORs would have to be before our tests would have had an 80% chance of being significant.

These “minimally detectable ORs” are based on the level of significance being set at 0.05 (2-sided in the first table and 1-sided in the second), the power set at 0.80 and using the empirically observed estimates of the standard error of the various estimates of θ.

	Barthel	Rankin	Glasgow	NIHSS	GLOBAL
OTT	3.6	3.8	3.6	4.0	2.9
PART	3.6	3.7	3.6	3.9	2.9
PrDisab	28.1	33.4	33.1	79.7	16.5
PrDM	4.8	5.2	5.1	6,4	3.7
OTT	2.7	2.8	2.7	2.9	2.3
PART	2.7	2.7	2.7	2.9	2.3
PrDisab	12.9	14.8	14.7	28.8	8.6
PrDM	3.3	3.6	3.5	4.1	2.8

Table entries are the ratio of odds ratios that would have to actually exist for the interaction tests just performed to have an 80% probability of being statistically significant. For example, the t-PA effect, as measured by the t-PA vs. Placebo OR, would have to be about 4 times higher (lower) in those with DM than in those without DM in order for these tests to have a reasonable likelihood of detecting the interaction.

The interactions involving polychotomous covariates with multiple degrees of freedom have even less power than indicated by these tables because they involve the spreading of patients over multiple classes, with smaller numbers per class.

Thus, while we, and the NINDS investigators, examined these interactions and report that they are not statistically significant, lack of significance does not imply the absence of interactions. Indeed, lack of evidence of an effect is not equivalent to evidence of the lack of an effect. Caution is needed in evaluating subgroup effects.

5.6 Absolute Risk Differences: Sensitive to the several concerns raised by many subsequent to the NINDS publications, the FDA approval, and the American Heart Association endorsement, the committee continued with further analyses of the data. Some of those analyses will be discussed in the subsequent sections; here we will discuss estimating the difference between the probabilities that a t-PA treated patient and a Placebo treated patient will experience a favorable outcome.

5.6.1 Differences Between Success Rates: Recall that in Table 5.8 the observed differences between success (favorable outcome) rates (%) for the t-PA and Placebo treatment groups were: 14.1%, 16.3%, 14.4%, & 13.7% for the Barthel, Rankin, Glasgow, and NIHSS outcome measures respectively. These estimates translate more directly than odds ratios into interpretations of the impact of treating a population of acute ischemic stroke

patients. If 1000 patients were treated according to the NINDS protocol, these numbers suggest that between 140 and 160 more patients would experience a favorable outcome than if t-PA therapy was not available or was not used on the whole population.

As outlined in Section 4.3.7, these differences between the t–PA and Placebo success rates can be estimated using the adjusted OR estimates. The resulting estimated differences and their 95% Confidence Intervals are: 19.3% (9.6-29.0%), 20.2% (10.6-29.8%), 17.9% (8.3-27.5%), and 15.6% (6.8-24.4%) respectively for the Barthel, Rankin, Glasgow and NIHSS outcome measures. Thus, after taking into account the modeling process, which led to slightly larger OR estimates, the estimated differences are also greater than the actual observed differences. In light of these differences, it seems reasonable to suggest that between 150 & 200 more of the hypothetical population of 1000 acute ischemic stroke patients would experience a favorable outcome if all 1000 are treated with t-PA according to the NINDS protocol than if none are.

5.6.2 Attributable Fraction: Computation of "attributable risk" or "attributable fraction" sheds additional light on the role of t-PA as a treatment for acute ischemic stroke. We can use this concept, which originated in the field of epidemiology to estimate what fraction of a disease in a population might be reasonably "attributed" to the presence of a risk factor in the population, to estimate the proportion of the unfavorable outcomes that can be "attributed" to exposure to the Placebo. Such estimates which are based on our OR estimates and the fraction of placebo patients with an unfavorable outcome can be interpreted as that fraction of the unfavorable outcomes that could be eliminated if all Placebo treatment could be eliminated in favor of the t-PA treatment. From the raw data, the estimates of the "attributable fractions are; 24.8%, 20.8%, 25.7% and 27.6% respectively for the Barthel, Rankin, Glasgow, and NIHSS outcome measures. The corresponding numbers using the model based OR estimates instead of the raw data are: 30.8%, 26.1%, 29.7%, and 29.8%. Thus, if we take the Barthel index as an example, 61.9% of the placebo treated patients had an unfavorable outcome. In our hypothetical 1000 acute ischemic stroke patients we thus expect ~620 to have an unfavorable outcome if all were treated with placebo. If, in contrast, all were treated with t-PA we expect a reduction in this number of unfavorable outcomes by between 25% (unadjusted) and 30% (adjusted). That is we expect between 435 and 465 unfavorable outcomes rather than 620 or a reduction of between 155 and 185 unfavorable outcomes achieved through the use of t-PA therapy. These numbers are obviously very similar to the figures quoted above corresponding to the increase in the number experiencing a favorable outcome.

5.7 Public Health Consequences and Exploratory Subgroup Analyses: The results of a clinical trial lead to population-based decisions rather than to patient-specific decisions. In this public health context, we (and the NINDS investigators) conclude that the use of t-PA in accord with the NINDS protocol will result in an increase in the total number of favorable responses among those acute ischemic stroke patients who satisfy the conditions of the protocol. However, physicians and patients face patient-specific decisions, even among patients who meet the conditions of the protocol, and further refinement of the results would be helpful in making these decisions. In subsequent chapters we examine subgroups to determine if there are groups of acute ischemic stroke patients who satisfy the conditions of the protocol but might not fare as well on t-PA therapy as the overall evidence suggests. In interpreting these subgroup analyses it is important to keep in mind both that the study was not designed to have substantial power to assess subgroup differences so these tests may fail to detect real

differences (see Sections 4.3.6 and 5.5.2) and that performing many exploratory analyses may deliver spuriously “significant” findings.

6. Blood Pressure Assessment and Management

6.1 Stated Methodology

6.1.1 Original Protocol: The NINDS investigators, in their first publication[1], made the following statements regarding patient eligibility for the clinical trial; (a) patients did not undergo randomization if they had "... a systolic blood pressure above 185 mmHg or diastolic blood pressure above 110 mmHg;...." (p 1582 Col 1 Para 3), (b) patients were also excluded if aggressive treatment was required to reduce their blood pressure to the specified limits. (p 1582 Col 1 Para 3), and (c) the protocol required that ... "blood pressure be maintained within prespecified values." (p 1582 Col 1 Para 5)

6.1.2 Blood Pressure Manuscript: In a subsequent publication[12], the NINDS investigators stated:" All patients had BP measurements at the time of admission to the emergency department and at the time of randomization (equivalent to the time of study-drug initiation) those with a systolic BP of ≤185 mm Hg and a diastolic BP of ≤110 mm/Hg were eligible for randomization." Patients with higher BP readings at the time of admission but who met BP criteria by the time of randomization were defined as hypertensive before randomization.

Between admission and randomization, aggressive antihypertensive therapy, defined as use of intravenous nitroprusside or repeated intravenous infusions of other medications, could not be used to meet eligibility criteria. After randomization, BP measurements were collected prospectively on a scheduled basis (Appendix 2)[12]. Patients with elevations of systolic BP >180 mm Hg or of diastolic BP >105 mm Hg in the 24 hours after randomization were defined as hypertensive after randomization. For such elevations, repeat BP determinations were recommended every 5 to 10 minutes but were not recorded in the trial. Prespecified antihypertensive treatment guidelines were given (Appendix 2)[12]. The date of administration of any antihypertensive treatment was recorded but not the time of administration. Acute antihypertensive therapy was defined as administration of intravenous nitroprusside, nicardipine, labetalol, or hydralazine; sublingual nifedipine; and sustained-released or topical nitroglycerin.

"To explore the relationship among BP reduction, thrombolytic therapy, and antihypertensive therapy, the severity of hypertension and declines in BPs were calculated at various time frames from randomization. To evaluate severity of hypertension, for each patient in the study the maximum mean arterial pressure during the first 24 hours after baseline was calculated. " To identify precipitous drops in BP soon after initiation of placebo or t-PA, the maximum abrupt decline, defined as the maximum decline between two consecutive mean arterial pressures during the first 8 hours, was calculated (measurements were hourly after the first 8 hours)."

6.1.3 Systems Approach: In another manuscript[8], the investigators offer additional guidelines (ibid. Table 5, p. 1539); (a) Patient Selection: contraindications: "On repeated measurements, systolic BP>185 mm Hg or diastolic BP>110 mm Hg at the time treatment is to begin, and patient requires aggressive treatment to reduce BP to within these limits:" and (b) BP control: pretreatment " Monitor BP every 15 minutes (should be <185/110 mm Hg), if >185/110, BP may be treated with one to two 10- to 20-mg doses of labetalol given IV push within 1 hour and/or nitroglycerin paste. If these measures do not reduce BP < 185/110 and keep it down, the patient should not be treated with rt-PA."

6.1.4 FDA Submission: PLA supplement 96-0350 Submitted by Genentech to the FDA

3/19/96[21]: (a.) (p23-24) It is stated that 17,367 patients with strokes were screened, but not enrolled; but none appear to have been excluded for reasons due to blood pressure. Unless BP is included in other serious illness 490/17367 this certainly would imply rapid spontaneous or therapeutic control of blood pressure. On further review the reasons for exclusion on p24 of the Genentech submission document only 95% of all patients. Possibly the investigators did not include patients excluded because of elevated BP. If so, then the 5% would be a reasonable estimate of the number potentially excluded for BP.

Exclusion reason	Number	%
Time from onset too long	8708	51.6
Symptoms rapidly improving	1749	10.4
Intracranial Hemorrhage	1306	7.8
Symptoms too minor	1106	6.6
Outside age range	1021	6.1
Other serious illness	490	2.9
Seizure at stroke onset	391	2.3
Stroke not present	373	2.2
Time from onset 90-180 min[1]	267	1.6
Recent prior stroke	219	1.3
Oral anticoagulants	210	1.2
Subarachnoid Hemorrhage	169	1.0

6.1.5 Exclusion Characteristics: In an attempt to define the exclusion characteristics of the study these data were compared with the investigators' report of 17,324 patients in the their 1997 manuscript[23]. The data sets in these two documents provide inconsistent tallies. The allocated percentages are inconsistent. The investigators suggest that these discrepancies are primary exclusionary criteria, but these also do not achieve 100% of the population. The graphic represents 93% of the population. Figure 3 in the manuscript rounds up/down inconsistently with relation to the Table above.

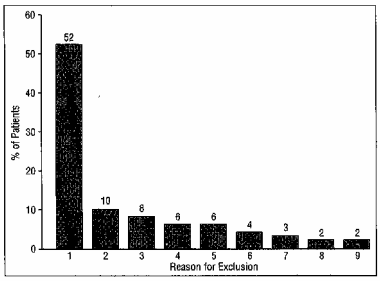


Figure 3. *Primary reasons for exclusion from the National Institute of Neurological Disorders and Stroke t-PA Stroke Trial (17,324 patients screened; 624 patients randomized). Reasons are coded as follows: 1, time; 2, improving rapidly; 3, intracranial hemorrhage on baseline by computed tomographic scan; 4, minor symptoms; 5, age; 6, serious coexisting illness; 7, seizure at stroke onset; 8, stroke ruled out; and 9, other reason.*

On p 18 of the clinical review it is stated that the most common protocol violation involved blood pressure criteria represented by 29/54 patients with violations of the 624 study cases. No details are offered with regard to the 29 patients who had a BP eligibility violation. There are no details in the FDA submission, or details in any of the manuscripts.

Via teleconference calls and email communications with Drs. Tilley and Brott, the following information was obtained related to BP and its management:

a) Figure 3 from the original manuscript displayed causes for exclusion of the 16,741 patients out of the total 17,363 who were screened[1]. The Table showed that 2 percent were excluded for "other reasons," and of that group, 162 were excluded because of high blood pressure. Seven of those patients were trial patients. Patients whose primary reason for exclusion was something other than blood pressure might have had blood pressure issues as well.
b) The exact number of patients given BP lowering medication prior to receiving treatment with the study medication was unknown. No information was available on patients who were treated by non-study physicians before the study physician's arrival in the ED.
c) While information was available as to which cardioactive drugs were given to the study patients, no information was available regarding the indications for giving the medications. Thus, it was not known if the cardioactive drugs were given for reasons other than lowering BP such as the treatment of chest pain or managing the ventricular response rate in patients with atrial fibrillation. The investigators reported that they reviewed the medication list (prior to identifying any patient characteristics) to make suggestions as to which medications their reviewer should consider as antihypertensive therapy. Genentech nurses determined which specific medications recorded on forms were to be considered antihypertensive therapies and so coded the agent.

6.2 Stated Data Sets

6.2.1: From the Original Publication[1]

Table 1. The Medical Histories of the Patients in the Study. (p 1582)

Variable	**Part 1**		**Part 2**	
	t-PA N=144	Placebo N=147	t-PA N=168	Placebo N=165
		percent		
Stroke	17	17	12	9
Transient ischemic Attack	22	14	13	19
Aspirin therapy	41	31	40	26
Diabetes	24	21	20	20
Hypertension	66	64	67	67
Myocardial infarction	25	21	22	20
Atrial fibrillation	18	20	20	16
Angina pectoris	18	22	24	24
Congestive heart failure	14	17	16	19
Valvular heart disease	11	7	6	6
Smoking in year before stroke	43	37	27	35
No preexisting disability	90	91	95	93

Table 2. Base-line characteristics of the patients in the two parts of the study, according to treatment group. (p1583)[1]

Characteristic	Part 1		Part 2	
	t-PA N=144	Placebo N=147	t-PA N=168	Placebo N=165
Blood Pressure (mm Hg)				
Systolic	155±22	153±20	153±22	152±21
Diastolic	85±12	85±13	85±14	86±15

6.2.2 From the Manuscript on BP[12]: "Hypertension was present on admission for 121 (19%) of the 624 patients eventually randomized into the NINDS rt-PA Stroke Trial; 65 were placebo-treated patients and 56 were t-PA-treated patients (Table 1). Postrandomization hypertension was detected during the first 24 hours in 372 patients (60%); 195 were placebo-treated and 177 were t-PA patients. For all patients, the frequency of antihypertensive therapy was similar for both the placebo- and t-PA-randomized patients. Before randomization, 28 (9%) of the 312 placebo patients and 28 (9%) of the 312 t-PA patients received antihypertensive treatment, whether or not they were hypertensive as defined above; 1 patient in the t-PA treated group, included in our analysis, received aggressive antihypertensive therapy (i.e., intravenous nitroprusside, a protocol violation). After randomization, 92 placebo patients (29%) and 75 t-PA patients (24%) received antihypertensive therapy. Antihypertensive therapy was administered either before or after randomization to 110 placebo patients (35%) and 96 t-PA (31%) patients." (p 1506)

6.2.3 Hypertension on Admission and Antihypertensive Therapy Received Before Randomization (p1506)[12]: Of the 121 patients who were hypertensive on admission, slightly more placebo patients received antihypertensive therapy (22 of 65, 34%) before randomization than did t-PA patients (11 of 56, 20%), but the difference was not significant (Table 1). The effects of antihypertensive therapy before randomization were similar in the groups randomized to t-PA and placebo for all clinical outcomes except death at 3 months (Table 2).

Table 1. Antihypertensive Therapy by t-PA-Treated and Placebo-Treated Groups

Hypertension Recorded	Received Anti-Hypertensive Therapy Placebo n	Placebo %	t-PA n	t-PA %	P*
Admission **, }	65	34	56	20	0.17
Within 24 hours after randomization **, ω	195	41	177	37	0.33
Not hypertensive by definitions	109	9	127	11	0.81

*Mantel-Haenszel test adjusting for centers and time strata.
**Groups are not mutually exclusive.

6.3 Stated Goals, Discussion And Conclusions: In the first t-PA manuscript[1], the investigators stated; "In our trial treating physicians used an algorithm to manage blood pressure after treatment began." (p 1586. Col 2 Para 2). In the hypertension manuscript[12] they stated;

"BP eligibility criteria more stringent than those used for t-PA-treatment of acute myocardial infarction were instituted, but aggressive measures to lower BP to allow enrollment were prohibited to prevent precipitous falls in BP. After initiation of t-PA therapy, a BP management

algorithm was followed, adapted from a similar algorithm designed for treatment of stroke patients in general (9). Recommended drugs were selected because of their rapid onset of action and because of their predictable effects with low potential for overshoot. Adjustments in the algorithm were made in response to experience during the course of the trial."

In the NINDS t-PA Stroke Trial[1] the investigators chose BP eligibility criteria similar to those used in the dose-finding trial (p 1505 Col 1 Para 2). The authors focus on a systolic BP of > 185 mmHg and a diastolic BP of > 110 mmHg at admission. Tables 4 and 5, which focus on blood pressure deal solely with severity and rate of reduction of the mean arterial blood pressure. Abrupt decline is analyzed in two ways as per Appendix 2; q 15 min in first 2 hours and q 30 min in hrs 2-8 following randomization.

The authors emphasized "gentle management" (p 1504) in those patients "who were hypertensive"[12]. In the last paragraph the authors state that "after initiation of t-PA therapy a BP management algorithm adapted from an American Academy of Neurology guideline[24] was followed.

In Subjects and Methods, aggressive therapy was defined as intravenous Nitroprusside or *repeated* infusions of other medications. Their chosen antihypertensive intravenous medications were stated to be nicardipine, labetalol, or hydralazine or sublingual nifedipine and sustained release or topical nitroglycerin. Based on our other data set, furosemide and diltiazem were also utilized as therapeutic agents for reasons determined by study monitors.

It is not clear how the Appendix 2[12] relates to the Subjects and Methods section. As the authors did not initially use the mean blood pressures for study entry their emphasis on mean vs. systolic or diastolic does not describe individual abnormalities.

In their discussion they state; "The antihypertensive therapy used in the NINDS study was modest in its effects and had little potential for overshoot. Hypertensive placebo patients who received the antihypertensive therapy after randomization did not have a greater maximum decline in mean arterial pressure over the first 24 hours compared with hypertensive patients who did not receive antihypertensive therapy. In addition, abrupt declines in BP were not more pronounced among placebo patients who were treated with antihypertensive therapy compared with those who were not, reflecting the careful use and gentle effects of the antihypertensive therapy administered in this study (Appendix 2)[12]."

The interaction of antihypertensive therapy with intravenous t-PA in this exploratory analysis is intriguing, but interpretations should be cautious. For the patients randomized to receive t-PA, antihypertensive therapy administered before t-PA was not associated with differences in early or late outcomes. However, hypertensive t-PA patients who received antihypertensive therapy had a more pronounced abrupt decline in mean arterial BP. Hypertensive t-PA patients who received antihypertensive therapy after randomization were less likely to have a favorable outcome at 3 months than hypertensive t-PA patients who did not. One possible explanation is the nonrandomized administration of antihypertensive therapy at the bedside. Investigators could have been more likely to treat hypertensive patients they judged to be sicker. " (p 1508 Col 1 Para 2, 3)[12]

"A randomized trial would be necessary to address adequately the effects of antihypertensive therapy on BP and on clinical outcome." (p1508 Col 2 Para 2)

"In summary, hypertension was a common phenomenon in the NINDS trial. BP eligibility criteria were applied in a balanced fashion. The antihypertensive therapy was designed for, and resulted in, modest effects on BP with low potential for overshoot. The results do not suggest that use of antihypertensive therapy adversely affected BPs or clinical outcomes of placebo-randomized patients. The effects of antihypertensive therapy following treatment with t-PA are complex and merit further study. Careful attention to BP and gentle management remain warranted for stroke patients treated with t-PA." (p 1508, Col 2, Para 2, 3)

In the investigators' manuscript on ICH[11] we find (p.2111, last Para.) under the heading Baseline and Time Dependent Covariates a first citation for admission diastolic blood pressure > 100 mmHg as associated with increased risk of symptomatic ICH. Later, p. 2114, Para 1, the authors suggest high correlations between systolic BP and mean BP and between systolic BP and pulse pressure. On several occasions such as the next to last paragraph of the Methods in the last sentence the authors state "prerandomization and postrandomization antihypertensive therapies were evaluated with patients who were hypertensive." Under Results: In the next paragraph the authors state: "patients received antihypertensive treatment whether or not they were hypertensive as defined above." In the last paragraph under Results (maximum BPs and declines in BPs) they state that the "more severe BPs were more likely to be treated". The last sentence in that same paragraph with regard to BP decline states that "abrupt declines were noted more frequently in treated patients."

6.4 The NINDS t-PA Review Committee's Evaluation of the BP issue

6.4.1 Review Data sets: As described in Section 4.1 of this report, the review committee had access to extensive data and we sought strict definitions of the following variables, their names, and their locations. When necessary, more information was requested and some clarification was obtained.

(i) Hypertension: Prestroke, Post stroke – Prerandomization, and Post Randomization

(ii) Hypertension Therapy: Prestroke, Post stroke – Prerandomization, and Post Randomization

(iii) Blood Pressure: At Admission, At Randomization, and Subsequent to Randomization

6.4.2 Investigator Queries

(i) When comparing the mean and SD of baseline systolic (BsSYS) and diastolic (BsDIA) blood pressures with admission: systolic (AdSYS) and diastolic (AdDIA) blood pressures, the admission values were higher.

	Mean	Std Dev	Max
BsSYS:	153.12	21.27	227
AdSYS:	158.92	21.33	254
BsDIA:	85.32	13.53	134
AdDIA:	89.24	15.81	180

(ii) The investigators' study form 10 section C item 2 asks if the patient has a history of hypertension with the options of answering yes, no, or unknown. Item 2b followed up by asking "If yes was medication prescribed?" and again the answers are yes, no or unknown.

(iii) A review of the descriptive characteristics of the submitted variables and, the range of blood pressures in the dataset indicates that some of these readings would have placed a substantial number of the patients in an exclusionary status. Inclusion would have been in violation of the upper systolic and/or diastolic blood pressures limits established.

(iv) Drs. Tilley and Brott stated (personal communication) that blood pressure variables for readings at admission (on arrival at the ED) and baseline (time of randomization) should be available.

The investigators stated (personal communication) that they used the randomization blood pressure variables in their analyses, but they stated that they may have used the terms baseline and randomization interchangeably.

Although patients may have had additional blood pressure readings prior to randomization and after randomization only the randomization blood pressure was recorded.

The investigators stated that admission blood pressure could be quite high, but if an antihypertensive regimen could lower the blood pressure to 185/110 or below at any time before randomization, the patient could be randomized into the trial. Post-admission prerandomization medication information was not collected.

It was uncertain whether the authors restricted lowering BP to those who were hypertensive by their exclusion criteria or at any other specific levels. The authors do not define precisely their therapeutic goals: How far below the cut off post randomization values of 180/105 did they wish to go?

The investigators did not offer information defining specific data as to what antihypertensive therapy was employed. The protocol reviewer utilized the list of medications given during hospitalization with the times administered to determine what prerandomization drugs had been used to treatment hypertension. Labetalol was considered antihypertensive every time, whereas calcium channel blockers and diuretics were reviewed and judged to be antihypertensive or not depending on a retrospective chart analysis. This could have been a source of error, but the investigators believe that it was an error that affected t-PA and control patients uniformly. The investigators' goal was to look at the effects of antihypertensive therapy given after admission to hospital.

The investigators' determination of pre-randomization treatment referred to the history of hypertension question ("yes/no/unknown" from form 10, item 2.) The assessment of post randomization hypertensive therapy and its influence on outcome was based on post-hoc evaluations of the medications given and an ad hoc decision determining whether this represented antihypertensive treatment or not.

Dr. Brott stated that in Cincinnati they did not treat high blood pressure to permit entry into the

trial, but that other centers engaged in this practice at the time of the study. Throughout the documents it is suggested that it is acceptable to treat hypertension so that patients can be treated with t-PA providing that the treatment is not 'aggressive'.

6.4.3 Review of Study Datasets: Our examination of the blood pressure data led to the following observations:

(i) Nineteen individuals were found with abnormally elevated (BP >185 mmHg or >110 mmHg) at both admission and baseline.
(ii) Ten individuals were found to be hypertensive at baseline who had not been so at admission. Although this table states >185/110 it actually means either >185 mmHg or >110 mmHg.

	Baseline BP	
≤185/110	572	95.17%
>185/110	29	4.83%

Twenty-one patients were missing baseline BP.

	Admission BP	
≤185/110	501	80.68%
>185/110	120	19.32%

One patient was missing admission BP.

		Admission BP	
		≤185/110	>185/110
Baseline BP	≤185/110	474 (79.00)	10 (1.67)
	>185/110	97 (16.17)	19 (3.17)

Twenty-two patients were missing admission or baseline BP.

(iii) Admission BP readings were missing in 1 patient, and randomization BP readings were missing in 21 patients.

(iv) A pair wise comparison of the recorded admission and baseline blood pressures for the entire cohort was performed. Restricting our attention to the 622 patients who were randomized into the study within 180 minutes of onset it was noted that 112 (18%) of them had identical admission and baseline blood pressures.

6.5 Review Committee's Areas of Concern:

6.5.1 Definitions The NINDS trial[1] had no specific definitions for the 'Prior Medical History' conditions, including a 'History of hypertension'. It was left to the discretion of the investigators at each site to determine if a patient had one of these conditions. With regard to the history of hypertension, we were unable to determine those patients who had a previous history, a

current history, whether treatment was current or whether treatment had occurred in the ambulance prior to hospitalization.

6.5.2 Protocol Applications There appeared to be some patients included in the study whose blood pressures at randomization exceeded either the systolic or diastolic maximum permissible values.

6.5.3 Therapeutic Interventions Dr. Tilley stated (11.11.02) that the data set included blood pressure at admission and blood pressure at baseline plus post randomization blood pressures.

6.6 Review Committee's Findings:

6.6.1 Definition of Hypertension: Publications from the t-PA studies and written and oral communication with Drs. Tilley and Brott document confusing and inconsistent information with respect to nominal and actual procedures for BP recording and management, and confusing and inconsistent nominal and actual eligibility and exclusion criteria. It was never defined as to what was precisely meant when the term hypertensive was used. Was it always based on their exclusion values, standard terms, or history of treatment?

6.6.2 History of Hypertension: Item 6.7.1 creates confusion with regard to the numbers of patients considered to have a history of hypertension throughout diverse comments and manuscripts. This resultant variability stems from the uncertainty of the definitions of hypertension and of the term history of hypertension. Although an analysis of current, recent or past use of hypertensive medications could be of interest, the data definitions are not sufficiently precise to support these exploratory analyses.

6.6.3 Blood Pressure on Admission and at Baseline: There appears to be a persistent uncertainty of the investigators in their written and stated use of terms. We confirmed that the terms baseline and at randomization values have been used interchangeably. In various manuscripts this confusion seems to be represented periodically. We demonstrated that admission and baseline blood pressure were identical across all centers 18% of the time which suggests that the interpretation for each term was confused at various times. At one center 50% of the blood pressure values were identical at admission and baseline.

Teleconference and email communications with Dr.'s Brott and Tilley revealed that there was variability between centers in the interpretation of the definition of admission BP. This led to some centers using the BP reading taken at the time of randomization as both the admission and randomization BP measurements. There were also some patients where the admission BP reading was the BP reading taken at the time of admission into the ICU after receiving the study drug.

6.6.4 Blood Pressure Exclusion Criteria: There was substantial inconsistency in the presentation of the exclusionary blood pressure criteria for entry into this study. It appears that the investigator's intent was to exclude patients whose blood pressure at the time of randomization:

(1) Exceeded either 185 mm Hg systolic or 110 mmHg diastolic on repeated measures.

Or

(2) received “aggressive antihypertensive therapy” to fall within these limits.

Drs. Tilley and Brott explained that a patient whose pre-randomization BP readings were persistently ≤ 185/110, could be included in the study even if the BP reading at the time of randomization exceeded 185/110.

While the Stroke Trial Guide manuscript stated clearly that the exclusionary BP criteria were based on repeated BP readings, throughout the relevant papers these criteria are written as

(1) Exceed 185 mm Hg systolic and 110 mm Hg diastolic.

Or

(2) > 185/110

These are not equivalent exclusion criteria, confuse the reader, and may have confused the investigators at various sites as 29 patients may have been included in the study with randomization blood pressures that would have required exclusion.

6.6.5 Antihypertensive Therapy Before Randomization: The concept of aggressive therapy is uncertain in written terms to the investigators and probably to the site practitioners. Prehospital therapy by EMS and other pre-randomization therapy could have included numerous diverse exceptionally efficacious rapid acting agents without being termed aggressive. The effects of these interventions in addition to all else that was done between admission and baseline makes the blood pressure determinations of limited value from an analytic perspective.

The caveats with regard to what was or was not considered antihypertensive treatment remains a concern. What agents? What doses? At what time? To whom? What results? Patient charts were evaluated retrospectively. In addition some patients would have been treated pre-randomization with delayed effects post-randomization.

Did giving an antihypertensive medication result in a lowering of BP? The investigators theoretically had a large number of patients who became hypertensive (again) post randomization. This is an interesting question; however the quality of the existing dataset may not allow for a proper analysis.

It is stated that 9% of all patients enrolled in the study received prerandomization treatment[12]. It would appear that the investigators substantially underestimated the number of patients who were treated with antihypertensive medication prior to randomization in view of the neglected or unidentified treatment regimens utilized outside the study protocol.

Other patients would have had blood pressure return to normalcy because stress, pain, hypoxia and other clinical issues were treated. Although managing BP was an important part of the protocol, the q15 minute BP readings that were required prior to giving t-PA, were not recorded neither was it recorded whether medication was given specifically for treating elevated BP. There was no formal protocol of sequential BP measurement that might allow for analysis of peak effect or duration of drug effect.

6.7 Use of Blood Pressure Data in Review Committee Outcome Models: Because of the concerns expressed above, the review committee decided not to incorporate information regarding blood measurement or management obtained during the prerandomization workup in our principal models of treatment effect. However, some have questioned this decision and so we include a summary of the assessment of baseline (randomization) blood pressures as predictors of favorable outcome using the same methods as were used to derive the best covariate model as described in the analysis section of this report. Recall that the variables fixed in all models – for each of the four outcome measures as well as for the global analysis – were the three stratification variables (Center, Part, and OTT (±90 mm)) and, ultimately, the covariates BsNIHSS, Age, BsDisab, BsDM, and the interaction between Age and BsNIHSS that were selected as described. No blood pressure variables were included in the process of determining which covariates to include in the adjusting models

6.7.1 Blood Pressure Variables as Favorable Outcome Predictors There were seven baseline blood pressure variables as described and defined in Tables 4.1 and 4.2. As seen in Table 4.2, there were 22 patients with missing values for these measurements. Values were imputed for these patients as described in Section 4.3.2. In the table below for each of the seven variables and each of the four outcome measures as well as for the global analysis we provide the chi-square and p-values which these variables would have carried into the stepwise process had they been included.

	Barthel		Rankin		Glasgow		NIHSS		Global	
	χ^2	p-v	χ^2	p-v	χ^2	p-v	χ^2	p-v	χ^2	p-v
BsSYSbp	2.36	0.12	2.51	0.11	2.20	0.13	0.43	0.50	1.53	0.21
BsSYSbp>190	0.25	0.61	1.96	0.16	0.26	0.60	0.91	0.33	0.94	0.33
BsDIAbp	0.07	0.77	3.05	0.08	1.30	0.25	0.93	0.33	0.36	0.54
BsDIAbp>100	1.94	0.16	0.69	0.40	0.43	0.50	0.26	0.60	1.53	0.21
BsMBP	0.94	0.32	3.74	0.05	2.23	0.13	0.92	0.33	1.04	0.30
BsMBP>130	0.34	0.55	0.74	0.38	0.12	0.72	0.35	0.55	0.27	0.60
BsPulseP	2.40	0.12	0.32	0.56	0.78	0.37	0.00	0.94	0.96	0.32

These χ^2 and p-values should be compared to the values for the variables that are summarized in Table 5.10 of Section 5.4. Note particularly the p-values for BsNIHSS, Age, and Pr Disability, all of which were <.0001. While some of these BP variables were "borderline significant", none were even remotely as important as predictors of favorable outcome as the variables ultimately included in the models.

The next table shows, for the four outcome measures and the global analysis, the χ^2 and p-values corresponding to each of the seven BP variables if they were each individually added

to the models including the 3 stratification factors and the 5 covariates selected for our analyses.

	Barthel		Rankin		Glasgow		NIHSS		Global	
	χ^2	p	χ^2	p	χ^2	p	χ^2	p	χ^2	p
BsSYSbp	1.97	0.16	4.39	0.03	3.92	0.04	0.64	0.42	1.74	0.18
BsSYSbp>190	1.15	0.28	4.26	0.03	1.53	0.21	1.93	0.16	2.68	0.10
BsDIAbp	1.78	0.18	7.00	0.00	3.97	0.04	2.57	0.10	3.38	0.06
BsDIAbp>100	0.18	0.66	0.00	0.94	0.08	0.77	0.12	0.72	0.04	0.84
BsMBP	2.45	0.11	7.43	0.00	5.11	0.02	2.00	0.15	3.53	0.06
BsMBP>130	1.25	0.26	1.91	0.16	0.68	0.40	0.97	0.32	1.38	0.24
BsPulseP	0.45	0.49	0.30	0.57	0.78	0.37	0.04	0.82	0.05	0.81

From this table we see that only in a few instances would any of these variables be selected for inclusion in the models predicting favorable outcome. For all but one of the outcome variables, the blood pressure variable with the smallest p-value and therefore at the top of the list to be added was BsMBP. So, we entered that variable into each model and the following table illustrates the impact that had on the remaining six BP variables.

	Barthel		Rankin		Glasgow		NIHSS		Global	
	χ^2	p	χ^2	p	χ^2	p	χ^2	p	χ^2	p
BsSYSbp	0.02	0.87	0.19	0.65	0.01	0.91	0.58	0.44	0.14	0.70
BsSYSbp>190	0.34	0.55	1.90	0.16	0.25	0.61	1.08	0.29	1.06	0.30
BsDIAbp	0.02	0.87	0.19	0.65	0.01	0.91	0.58	0.44	0.14	0.70
BsDIAbp>100	2.29	0.13	3.09	0.07	1.33	0.24	0.27	0.60	2.10	0.14
BsMBP>130	0.21	0.64	0.03	0.86	0.07	0.80	0.17	0.68	0.10	0.75
BsPulseP	0.03	0.87	0.19	0.66	0.01	0.91	0.59	0.44	0.14	0.70

Clearly, only if we were very generous regarding the qualifications necessary for a variable to enter the model as a covariate would any of these variables enter.

6.7.2 Influence of BP Variables on OR Estimates We now assess the impact that the addition of these variables on the estimate of the t-PA effect. To set the stage, recall, (Table 5.15) that for the Barthel, Rankin, Glasgow and NIHSS outcome variables, the raw (unadjusted) odds ratio estimates were, respectively, 1.78, 2.07, 1.85, and 2.01. Following adjustment by the extremely significant covariates included in the model, these estimates

became, respectively, 2.19, 2.43, 2.13, and 2.19. In other words, adjusting for extremely significant covariates increased the odds ratio estimate by a relatively small amount.

Following the addition of BsMBP and any others with a p <0.10, the corresponding odds ratio estimates became: 2.20, 2.51, 2.16, and 2.20, respectively. Thus, inclusion of the blood pressure variables, which were only marginally related to the outcome, had, predictably, almost no influence on the odds ratio measure of treatment effect.

In the global analysis, the BsMBP variable had a p-value to enter the model of 0.06. After it was allowed to enter, the next most "significant" variable was BsDIAbp>100 with a p-value of 0.15 so no other blood pressure variable other than BsMBP was entered into the global model. The Global odds ratio estimates were:

Unadjusted:	1.88
Adjusted (w.o. bp variables):	2.13
Adjusted including BsMBP:	2.14

The addition of the blood pressure variables had no impact on the estimate of the t-PA to Placebo odds ratio estimate.

6.8 Summary and Conclusions

Our analysis identified a number of problems regarding pre- and post-randomization blood pressure measurement and management:

- Non-compliance with the defined protocol was substantial, and persistent, throughout the study with regard to both the documentation of blood pressure readings, and adherence to the treatment regimen for hypertension.

- There was limited rigor with regard to the pharmacologic characteristics of antihypertensive regimens. In some instances pharmacologic monitoring was performed by representatives (nurses) of the sponsoring pharmaceutical firm. Medications employed were listed by date, but not by time, eliminating consequential interpretive utility.

- The exact number of patients who received medication to lower blood pressure either prior to, or after, receiving study treatment is unknown.

- The confusion regarding blood pressure documentation, and the lack of knowledge of treatment of hypertension either prior to, or after, receiving study treatment, could have led to an unknown number of patients receiving treatment in violation of the nominal study protocol.

Based on these observations, we reached the following conclusions:

- It was not possible to assess the effect of hypertension management on clinical outcome in acute ischemic stroke patients treated in the NINDS study.

- The blood pressure variables should not be included in the statistical models. However, we also found that inclusion of the blood pressure variables in the statistical models would have been inconsequential with regards to altering the t-PA treatment effect.

Finally, the inconsistent documentation of both blood pressure readings and hypertension management seriously undermines the NINDS investigators statement that blood pressure management was a significant part of the protocol that contributed to the success of the study. Nonetheless, we concur with the NINDS investigators premise that blood pressure management should be included in the protocol for treating acute ischemic stroke patients with t-PA. It is biologically plausible that hypertension management could affect clinical outcome in acute ischemic stroke patients treated with t-PA, and data from the cardiology literature has already demonstrated that in acute myocardial infarct patients, the risk of having an intracerebral hemorrhage is related to pre-treatment blood pressure[25-27]. However, further clinical studies will be needed to assess whether blood pressure management is related to better clinical outcomes in acute ischemic stroke patients treated with t-PA.

7. INTRACEREBRAL HEMORRHAGE

7.1 Introduction: Prior to the initiation of the t-PA trials, there was concern that t-PA therapy for ischemic stroke might increase the risk of an intracerebral hemorrhage (ICH) to an unacceptably high level. Indeed, the NINDS investigators specifically stated, "… the use of rt-PA for cerebral arterial thrombolysis requires a careful evaluation of both the risks and potential benefits" (p. 1581)[1]. In each of the two primary studies, ICH was considered a serious adverse event and, consequently, the protocols required that a CT scan be performed at 24 hours and between 7 to 10 days after randomization and whenever symptoms suggested an intracerebral hemorrhage. A symptomatic intracerebral hemorrhage was defined as "a CT-documented hemorrhage that was temporally related to deterioration in the patient's clinical condition in the judgment of the clinical investigator"[1]. Asymptomatic hemorrhages were defined as those confirmed by the protocol designated CT, in the absence of symptoms[1].

The investigators' protocol stated that, "Interim analyses were required after every three symptomatic ICHs and after every 10 deaths" and that the rate of occurrence of symptomatic ICH among t-PA treated patients was "compared with the rate of 8% estimated from pilot studies using similar doses and times of treatment" (p. 1584)[1]. The NINDS investigators reported a total of 22 symptomatic and 23 asymptomatic ICHs within 36 hours of treatment (p.1586)[1]. Of the symptomatic ICHs, 20 occurring among patients treated with t-PA and two among those receiving placebo ($p < 0.001$) whereas, of the 23 asymptomatic ICHs, 14 occurred in patients treated with t-PA and 9 in those receiving placebo ($p=0.23$). These data are summarized in the following table.

ICH	rt-PA	Placebo	Total
Symptomatic	20	2	22
Asymptomatic	14	9	23
None	278	301	579
	312	312	624

Subsequent to the publication of the primary analyses, the NINDS investigators published a manuscript focused on ICH[11]. The stated purpose of that manuscript was the identification of "variables associated with intracerebral hemorrhage in patients with acute stroke who receive t-PA". In this manuscript they again report 22 symptomatic ICHs (20 from the t-PA treated group and two from the placebo group) but report only 21 asymptomatic ICHs (13 from the t-PA treated group and 8 from the placebo group) in contrast to the 23 reported in the first manuscript[1, 11]. One of these exclusions is explained as a post-surgical ICH and the other apparently occurred more than 36 hours after treatment. Hemorrhages occurring more than 36 hours after t-PA therapy (there were 5 symptomatic) were deemed unrelated to therapy.

The manuscript describes complex statistical analyses designed to identify patients at a high risk of experiencing an ICH. These analyses utilized both prerandomization (baseline) and time dependent data collected during the 36 hours subsequent to the initiation of therapy. The investigators started with 45 variables with a stated goal of identifying risk factors for ICH in four different scenarios.

1. Symptomatic ICH; t-PA treated only (n=312, ICH=20)
2. Symptomatic ICH; t-PA & placebo patients (n=624, ICH=22)
3. Symptomatic & asymptomatic ICH; t-PA treated only (n=312, ICH=33)
4. Symptomatic & asymptomatic ICH; t-PA & placebo (n=624, ICH=43)

The variables remaining in their "final model" for each of the above scenarios were:

Scenario 1
- Baseline NIHSS Score (categorized into 5 levels)
- Edema/mass effect on baseline CT (yes/no)
- No time dependent covariates

Scenario 2
- Baseline NIHSS Score (categorized into 5 levels)
- Edema/mass effect on baseline CT (yes/no)
- A treatment indicator variable (t-PA/Placebo)
- No interaction of treatment with the covariates

Scenario 3
- Baseline NIHSS Score (categorized into 5 levels)
- Edema/mass effect on baseline CT (yes/no)
- Time dependent covariates: external bleeding/oozing and pulse pressure

Scenario 4
- Baseline NIHSS Score (categorized into 5 levels)
- Edema/mass effect on baseline CT (yes/no)
- A treatment indicator variable (t-PA/Placebo)
- A treatment effect interaction with current smoking.
- No mention of time dependent covariates

It was stated in their Methods Section (p. 2111)[11] that these models would be used to define high-risk subgroups for the development of ICH within which t-PA treatment effect could be assessed. However, they are never mentioned in the results or discussion sections.

7.2 Review Committee Analyses: As specified in the committee charge, we conducted a "careful evaluation of both the risks and potential benefits" and followed the NINDS investigators' lead to see if baseline data can help define high-risk subgroups in which t-PA treatment might be contraindicated due to the level of elevated risk of ICH. Our analytical efforts involved three separate activities.

(1) Imputation of missing data,
(2) Assessment of the net effect of t-PA therapy in the face of the increased risk of ICH,
(3) An attempt to identify a group at high-risk for the development of ICH.

Among the 622 patients to whom we restricted our analyses (Section 4.1.3), we identified 22 who experienced a symptomatic ICH and 20 who experienced an asymptomatic ICH.

Consequently, all of our analyses and comments pertain to these 622 patients among whom at most 42 experienced an ICH.

7.2.1 Missing Values: We planned to use the same set of 45 variables that the NINDS investigators utilized to define groups of patients at high risk for ICH[11]. However, as described in Section 4.3.2, some variables had missing values which we imputed while others were not observed in enough patients to warrant their inclusion in the analysis. Further, we elected, as is described in Section 6, not to use admission or baseline blood pressure determinations in our analyses. Consequently, the analyses reported herein are restricted to 34 variables with observations on all of the 622 patients.

7.2.2 ICH Analyses: To state our questions precisely, to describe the data available to address the questions, and to put our analyses into perspective, we offer the following statements and observations. While these observations pertain specifically to the occurrence of Symptomatic Intracerebral Hemorrhages they also apply in essence to all ICHs, both symptomatic (SICH) and asymptomatic (ASICH).

7.2.2.1 ICH Risk Increases with t-PA: The chance of an ICH increases with the use of t-PA therapy.

a) 2 SICHs out of 312 placebo treated patients.
b) 20 SICHs out of 310 t-PA treated patients.

7.2.2.2 Favorable Outcome Chance Decreases with t-PA: The chance of a favorable outcome decreases in the presence of an SICH.

In the t-PA treated group, for the Barthel index (B), among those 290 patients not experiencing an SICH, 55% had a favorable outcome at 90 days. In contrast, among those 20 patients experiencing an SICH, only 10% had a favorable outcome. For the Rankin (R), Glasgow (G) and NIHSS (N) outcome measures, the corresponding percents are: 45% & 10%, 48% & 10%, and 36% & 15%, respectively.

In the placebo group, the favorability rates among those 310 patients not experiencing an SICH were 38%, 27%, 31% and 21% respectively for the B, R, G, & N outcome measures respectively. There were only 2 patients in the placebo group who experienced an SICH and they both had an unfavorable outcome. Thus, there are no data regarding the rate of a favorable outcome among ischemic stroke patients experiencing an SICH in the absence of t-PA therapy.

7.2.2.3 Favorable Outcome Chance Increases with t-PA and no ICH: The chance of a favorable outcome increases with t-PA therapy in patients without SICH.

In the comments above, we note that the percent of patients without an SICH who had a favorable outcome was higher in the t-PA treated group than in the placebo group. This can be summarized, for the 4 outcome measures, in the table.

t-PA	Placebo	rate diff	rate ratio	odds ratio
B	55%	38%	17%	1.44	1.98
R	45%	27%	18%	1.69	2.25
G	48%	31%	17%	1.53	2.02
N	36%	21%	15%	1.75	2.18

Thus, among those not experiencing an SICH, the favorable outcome rate is greater in the t-PA treated patients than in those on placebo by between 15 & 18 percentage points with the corresponding odds of a favorable outcome among t-PA treated patients essentially twice that among those on the placebo. Again, the lack of data on the favorable outcome of SICH patients in the placebo group presents an analytic problem.

The fundamental question that must be addressed is how to balance the evidence of the efficacy of t-PA therapy with the equally clear evidence that such therapy carries an associated increased risk of ICH, substantially decreasing the chances of a favorable outcome.

There are two components to this question. One pertains to the net effect of t-PA therapy while the other pertains to the issue of whether there are subgroups of patients who are particularly susceptible to ICH and, therefore, should not be treated with t-PA. While we intend to offer comments on both issues, we must express caution, as did the NINDS investigators[1, 9, 11], that the clinical trials whose data we are examining, were designed and powered to address the question of a net effect, not the question of an interaction or subgroup effect, and that we are performing exploratory subgroup analyses.

7.2.2.4 ICH Related Morbidity and Mortality: Among the 42 patients with either a symptomatic or an asymptomatic ICH, 10 were in the Placebo group and 32 in the t-PA group. Very few had a favorable outcome (7, 6, 6, 5 on the B, R, G, and N scales respectively with only one from among the 10 placebo treated patients). Fewer of the 22 SICH patients had a favorable outcome (2 each by B, R, and G and only 1 by N) with none of them from the 2 SICHs in the Placebo group. At 90 days, 22 (52%) of the 42 ICH patients were dead which included 16 (73%) of the 22 SICH patients

7.2.3 Net t-PA Effect: The following concerns the net t-PA effect, addressing the issue of whether a patient should be administered t-PA.

7.2.3.1 Favorable Outcome Chance Increases with t-PA Among All Patients: The chance of a favorable outcome **increases** with t-PA therapy even when those patients experiencing an SICH are included in the analysis.

*In the table below, the observed data **including all patients** randomized to the studies whether experiencing an SICH or not, are summarized in terms of the rates of favorable outcomes.*

	t-PA	Placebo	rate diff	rate ratio	odds ratio
B	52%	38%	14%	1.37	1.78
R	43%	27%	16%	1.61	2.07
G	45%	31%	14%	1.46	1.85
N	34%	21%	13%	1.67	2.01

These observed rates and comparisons, with the effects of t-PA, clearly diluted by the inclusion of the SICH patients, nearly all from the t-PA treatment group with their associated reduced chance for a favorable outcome, are still highly suggestive of a net positive effect associated with t-PA therapy.

7.2.3.2 Modeled Likelihood ORs Significantly > 1 Among All Patients: The chance of a favorable outcome increases with t-PA therapy, even when those patients experiencing an SICH are included in the analysis and formal models are created to adjust for stratification factors and other covariates associated with the chances of a favorable outcome.

The models developed are discussed in Section 5.5.1. They do not contain any interaction terms involving the t-PA indicator, as none was found to be significant. The adjusted OR estimates for Barthel, Rankin, Glasgow and NIHSS are, 2.19, 2.43, 2.13 and 2.19 respectively. The adjusted results, in terms of estimates of differences in favorable outcome rates, (Section 5.6.1) are 19.3%, 20.2%, 17.9% and 15.6% respectively, all somewhat larger than seen in the above table. The fundamental message in these last two analyses is that the net effect of t-PA therapy remains positive even though some patients are put at higher risk of an unfavorable result due to their increased risk of an SICH.

7.2.3.3 Conclusion Regarding Net Effect: In Section 5.6 of this report, the public health implication of these analyses is discussed. In summary, if 1000 acute ischemic stroke patients receive t-PA therapy according to the NINDS protocol, somewhere between 120 and 160 more of them will experience a favorable outcome at three months than if t-PA was not available. This even though about 65 of these 1000 patients would experience an SICH as a result of the t-PA with the resultant reduced chance of a favorable outcome.

7.2.4 Identification of Variables Predicting ICH: This analysis is based on the study of 622 patients, 310 randomized to t-PA and 312 to placebo. Of the 310, 20 experienced a symptomatic ICH within 36 hours of randomization (odds = 20/290 = 0.069) and an additional 12 were diagnosed as having an asymptomatic ICH within the same time period (odds of any ICH = 32/278 = 0.115). Among the 312 patients randomized to placebo, 2 experienced a symptomatic ICH and 8 an asymptomatic ICH with respective odds of 2/310 = 0.0065 and 10/302 = 0.033. Each of these placebo odds differs significantly from its counterpart in the t-PA group ($p = .0004$ & $p < .0001$ for all ICH & SICH respectively, Table 7.1).

To facilitate the remainder of this discussion we define some notation used in the tables. We used the same 4 "scenarios" defined by the NINDS investigators terms of types of ICH (symptomatic or all ICH) and treatment groups (t-PA treated or all patients)

I. Symptomatic ICH; t-PA treated only (n=310, SICH=20)

II. Symptomatic ICH; t-PA & placebo patients (n=622, SICH=22)

III. Symptomatic & asymptomatic ICH; t-PA treated only (n=310, ICH=32)

IV. Symptomatic & asymptomatic ICH; t-PA & placebo (n=622, ICH=42)

The odds of an ICH are, respectively: 0.069, 0.037, 0.115 and 0.072.

Within each scenario we used 34 prerandomization variables, including the treatment indicator variable, t-PA if appropriate, to examine the question of which of them, individually and collectively, might predict those at a higher risk for ICH. In Table 7.1, we have summarized the results of univariate logistic model analyses of the influence of each of the 34 variables within each of the four scenarios. The variable names are defined in Section 4.1.2. The bolded variable names indicate variables for which some imputation was necessary. The "DF" column indicates the number of degrees of freedom a variable requires in a model (i.e., one for continuous and dichotomous variables and greater than one for variables dividing the patients into more than two classes). The remaining columns are divided into four sets of two with the four sets corresponding to the four scenarios (indicated by Roman Numerals) and the 2

columns within each set providing the univariate p-values and odds ratio estimates (no ORs provided if DF>1). The variables are in ascending order according to their p-values in the scenario I analyses and those with p-values <0.20 are in bold. This facilitates observation that while there is a great deal of commonality among the scenarios as to which variables are "significant", there is also some diversity. Note that in scenarios II & IV the t-PA variable is included and its p-value indicates the significant difference between the t-PA treated patients and the placebo treated patients with regard to the risk of an ICH.

In Table 7.2**,** the next stage of the investigation, as carried out by the NINDS investigators, is summarized. In that stage they took all of the variables within each scenario whose univariate p-values were <0.20 and put them into a multivariate logistic model. The same structure is used so it can easily be seen which variables are not in the models. Note here that some of the 34 variables are literally constructs of others and they cannot be in a multivariate model together. In such situations where both were significant, we either made a considered judgment as to which variable to include or we used the variable selected by the NINDS investigators. We note that in these models there are frequently variables that have p-values >0.20 and they are no longer presented in boldface. This type of thing happens when variables are correlated.

In Table 7.3, the results of simple stepwise modeling processes for each scenario are summarized. As did the NINDS investigators, we required a variable to have a p-value < 0.20 to enter and remain in each model. Here there are columns labeled STP to indicate the order (step) in which the variables entered the models – an indication of "importance". There is much more diversity in these models although some variables appear nearly always, indicating some consistency, if not validity, in the process.

7.2.4.1 Methodological Issues Our primary analyses are based on BsNIHSS, the "continuous" version of the NIHSS variable, but we did investigate use of BsNIHSS(5), a partition of the NIHSS score into 5 categories used by the NIHSS investigators. Such categorization allows for non-linear relations, but our analyses did not indicate a sufficient degree of lack of linearity to warrant 5 categories. However, not surprisingly, BsNIHSS(5) did suggest that the major ICH risk was at the upper end of the NIHSS score. However, BsNIHSS was the most statistically significant and we based our analyses on it.

Glucose (GLU) presented a similar issue with a continuous version and a dichotomous version, dividing the glucose scale at 300 mg/dl, competing with each other. Because the continuous version was so statistically significant when the t-PA & placebo groups were combined we included it in further analyses.

There were two important and related concepts, edema and mass effect as assessed at the prerandomization CT scan. We identified three modeling options for these dichotomous variables: let each be a candidate for the model, with only one allowed in; combine them into a 2 degree of freedom variable; or create an "either/neither" indicator variable. The NINDS investigators used this latter option as did we, creating the "ED/ME" variable which equals 1 if a patient has either Edema or Mass Effect and 0 otherwise, which is highly significant in all our models.

In Table 7.4, the results of stepwise modeling with the constraint that certain variables must be in the models are summarized. For scenarios I, and III the variables constrained to be in the

models were AGE & BsNIHSS while in scenarios II & IV, t-PA was also constrained to be in the models. The variables constrained to be in the models regardless of their p-values are designated as having entered the models at step 0.

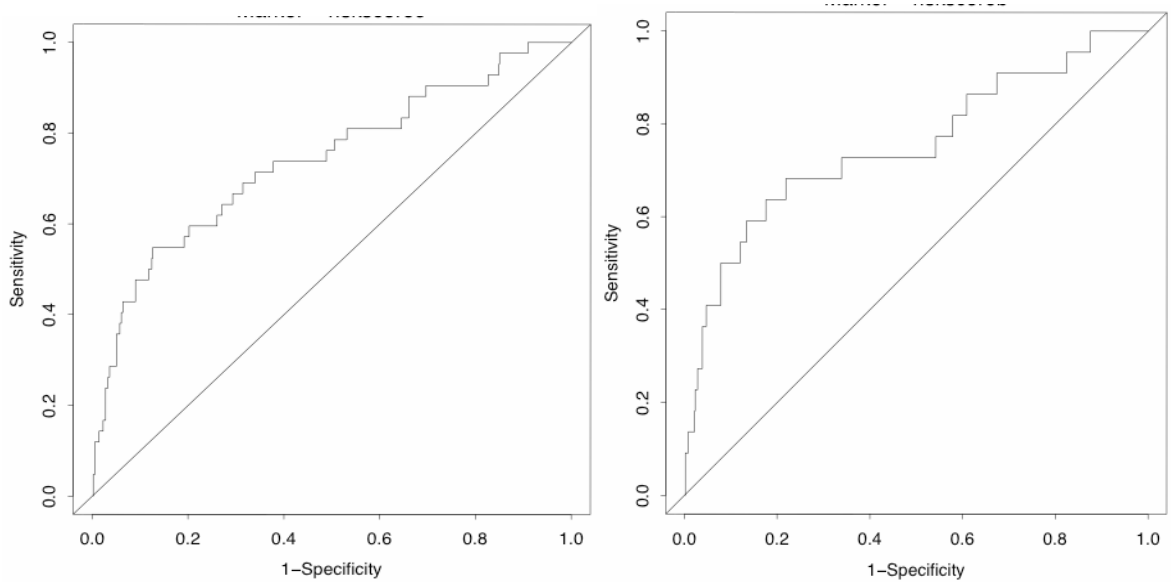
7.2.4.2 Results of ICH Risk Factor Identification: In all scenarios the variables that are important ($p < 0.20$)(in addition to t-PA) are associated with increased risk of ICH. In addition, in scenarios II & IV we can investigate whether any variables modify the t-PA effect in increasing the risk of ICH. To investigate this question we selected the two models in Tables 7.3 and 7.4 that were the most effective and tested whether any t-PA interactions with other variables were statistically significant. While the t-PA/current smoking (CSMK) interaction was "suggestive" ($p = .15$ in Scenario II and $p = .07$ in Scenario IV) in light of all the "data mining" taking place we elected not to consider it further.

In addition to checking for interactions, we reran the stepwise models for scenarios II & IV with the t-PA variable eliminated. The resulting models contained the same variables as when t-PA was available, indicating that the presence of the t-PA variable did not influence which other variables were associated with an increased risk for an ICH in this patient population.

Therefore, we created two risk scores (RS_{ICH} and RS_{SICH}) using in both the variables with $P < 0.1$ in the model developed for scenario IV, forcing AGE & BsNIHSS into the models. Each of these risk scores is based on a multivariate logistic/linear model using dependent variables AGE, BsNIHSS, ED/ME, GLU, CSMK and RACE to discriminate between patients with and patents without an Intracerebral Hemorrhage. For RS_{ICH} we discriminated between patients with any ICHs (either symptomatic or asymptomatic) and those with no ICH. For RS_{SICH} we discriminated between SICH patients and all other patients. Each patient then obtained a value for each of these risk scores based on the patient's values for the 6 variables and the estimated intercept and coefficients of the variables in the two separate models.

7.2.4.3 Risk Score Sensitivity and Specificity for any ICH: The Receiver Operating Characteristic (ROC) curves (sensitivity plotted against (1-specificity)) associated with each risk score are illustrated below. The curves show that neither RS is very effective in predicting ICH and that they are almost equally effective. Indeed, one indicator of the value of a risk score is the area under the ROC curve. These two ROC curves have almost identical areas (.74 for RS_{ICH} and .75 for RS_{SICH}).

The ROC Curve on the left corresponds to RS_{ICH} and the one on the right to $RS_{SICH.}$



When each risk score is inserted in the separate logistic models for the four outcome measures (Barthel, Rankin, Glasgow and NIHSS) and in the Global model predicting a favorable outcome (see Sections 5.4.3 and 5.3.4), it brings some new information into some of the models as is illustrated by the p-values in the table below.

	RS_{ICH}	RS_{SICH}
Barthel	0.02	0.02
Rankin	0.005	0.003
Glasgow	0.14	0.13
NIHSS	0.41	0.53
Global	0.04	0.05

Thus, in all but the Glasgow and NIHSS models the risk scores each bring some new information into the models. Considering that the risk scores are constructed using Age and BsNIHSS, both of which are included in the adjusting covariates, it might have been expected that they would add nothing to these models.

However, that is not the primary question at hand. What is really critical is whether the risk scores help in the identification of a subset of patients who, because of their risk for an intracerebral hemorrhage, might be at especially high risk of an unfavorable outcome should they be exposed to t-PA. Thus, in models with the adjusting covariates and an indicator for t-PA treatment (see Section 5.5) we also insert each risk score and the associated t-PA by Risk Score interaction term. The results of these interaction tests are summarized in terms of p-values, in the table below.

	RS_{ICH}	RS_{SICH}
Barthel	0.35	0.40
Rankin	0.32	0.25
Glasgow	0.22	0.27

NIHSS	0.77	0.64
Global	0.73	0.81

In each of these models, for each of the risk scores, the interaction between the t-PA group indicator and risk score is not statistically significant. Thus, while the concept of a risk score based on a careful statistical analysis comparing those with and without an intracerebral hemorrhage is appealing, this formal process led to risk scores which were not particularly sensitive of specific and did not identify a group of patients who would be placed at special risk if treated with t-PA.

7.2.4.4 A Simplified Risk Function: To simplify real-time implementation of the RS_{ICH} approach, we dichotomized the 4 most important variables used in computing the RS_{ICH} (AGE, NIHSS, ED/ME, & GLU) as indicated and subdivided the 622 patients into 3 categories according to those factors

- Age >70 years
- Glucose >300 mg/dl
- Baseline NIHSS >20
- Edema and/or Mass Effect on the CT scan

As seen in the table below, the risk of symptomatic ICH and of any ICH increases noticeably with the number of risk factors ($p < .0001$ in both instances). Clearly, this grouping based on these four factors does predict the occurrence of ICH.

Table: Simplified ICH Risk Function

No. of risk factors	No. of patients (%)	% with ICH		
		Symptomatic	Asymptomatic	Total
None	238 (38%)	1.3	2.1	3.4
1	278 (45%)	2.9	2.9	5.8
≥2	106 (17%)	10.4	6.6	17.0

In Table 7.5 we summarize a basic analysis of the observed data yielding rates of favorable outcomes by all four outcome measures, within each of the three groups among all patients. The bottom part of the table, pertaining to the 106 patients with one or more of the ICH risk factors, provides some interesting results. Of the 106 patients, the overall percents of favorable outcome for the B, R, G, & N outcome measures respectively, were 15%, 8%, 12% & 8.5%, much less than reported for the study overall. Most importantly, the rates of a favorable outcome for the placebo treated patients are slightly though not significantly larger than for the t-PA treated patients for three of the 4 outcome measures. When the three groups are compared in models containing only the stratification factors of Center, Part and OTT we found them to have significantly different odds of a favorable outcome for each of the 4 outcome measures and for the Global analysis ($p <0.0001$ for all models, models not shown). In the same context when we searched for t-PA by ICH group interaction none were significant ($p = 0,21$, $p = 0.16$, $p = 0.09$, $p = 0.15$, and $p = 0.41$ for the Barthel, Rankin, Glasgow, NIHSS and Global models respectively).

In Table 7.6, we summarize the results of inserting indicator variables separating these three groups into the individual outcome models and the global model with all of the adjustment

factors (Section 5.5) and the treatment indicator t-PA included. In such models, we find no evidence that the three groups have different rates of a favorable outcome because the variables BsNIHSS and AGE, which are key to forming the groups, are among the adjusting variables. Furthermore, and most importantly from the net effect standpoint, we find no evidence of a significant interaction between t-PA and these groups in any of the models (p = 0.57, 0.28, 0.24, 0.18 and 0.41 for the Barthel, Rankin, Glasgow, NIHSS and Global models respectively). Recall all of the caveats regarding the detection of significant subgroup effects.

7.3 Summary and Conclusions:

In the NINDS trial, the overall risk of symptomatic ICH was 6.5% in t-PA treated patients vs. 0.6% in patients receiving placebo. When a symptomatic ICH occurred after treatment with t-PA, there were significant clinical consequences. Only a small minority had a favorable outcome (e.g., for the Barthel index, the favorable outcome in patients with symptomatic ICH was 10% vs. 55% in patients without ICH) and the three month mortality rate was very high (75%).

A number of putative risk factors for ICH were identified, with many of them being interrelated. Our exploratory analysis found four risk factors, age >70 years, baseline NIHSS >20 points, plasma/serum glucose >300 mg/L and edema and/or mass effect on the initial CT scan, that were associated with both an increased risk of having an SICH and a lower likelihood of having a favorable outcome. For patients with either no risk factors or only one risk factor, the likelihood of having a favorable outcome favored the t-PA treatment group, while for the group at highest risk (> 1 risk factor), there was essentially no difference between the t-PA and placebo groups with regards to the likelihood of having a favorable outcome. However, the analysis also found that the adjusted t-PA to placebo odds ratios for favorable outcome in the three subgroups with different numbers of risk factors were not significantly different, and were consistently in favor of the t-PA treatment group.

We conclude that there was no statistically significant evidence of the existence of any subgroup of acute ischemic stroke patients in whom the risk, and consequences, of having a symptomatic ICH clearly outweighed the beneficial effects of t-PA. However, it is important to keep in mind that because of the study design and the small number of patients who had an SICH, this trial was not powered to identify risk factors related to having either an SICH or a decreased likelihood of a favorable outcome. Risk factors for ICH acute ischemic stroke patients treated with t-PA should be evaluated in future studies that are designed, and powered, to evaluate this question.

How the findings of this exploratory analysis are used in the management of the individual patient with acute ischemic stroke, balancing risks and benefits based on very limited scientific information, is for the patient and the attending physician to decide.

8. SPECIAL TOPICS

8.1 Age, Baseline Stroke Severity, and Baseline Stroke Severity Imbalance

8.1.1 Introduction: It is well documented that stroke severity at onset and age are major predictors of favorable outcome. For each of the four outcome measures, and for the global statistic (Section 5.4.3), baseline NIHSS (BsNIHSS) and Age were the two most significant indicators of outcome following stroke among all available covariates in this analysis. It was further demonstrated that they interacted with each other in a synergistic way so that the joint effect of increasing age and increasing stroke severity was greater than the "sum" of their individual effects. This is not an unexpected result since at an advanced age even a modest increase in stroke severity can have significant clinical consequences. In the development of the covariate model we used BsNIHSS and Age as continuous variables and included in all models the product of BsNIHSS and Age to account for the interaction (Section 5.4.3). The NINDS investigators also initially analyzed the baseline NIHSS score as a continuous variable to adjust their analyses[1]. In that format the two treatment groups are in balance, with the NINDS investigators reporting nearly equal median values as being not significantly different by a rank sum test ($p = 0.10$). We corroborated that result (Table 4.2), and also acknowledged and discussed an imbalance noted later by the NINDS investigators (4). This imbalance became obvious when patients were grouped into five classes (approximately quintiles) according to baseline NIHSS (Q_1: 0-5, Q_2: 6-10, Q_3: 11-15, Q_4: 16-20, Q_5: >20). As seen in Table 4.2, this categorical distribution of BsNIHSS differed quite significantly ($p = 0.005$) in the t-PA and Placebo groups. To facilitate discussion, we refer to the categorical variable as BsNIHSS(5) to distinguish it from BsNIHSS. Age demonstrated a smaller imbalance ($p = 0.02$) as is seen in Table 4.2, in the opposite direction as there were more younger patients randomized to the Placebo arm of the trial than to the t-PA arm.

The primary goal of this section is to investigate the impact of the BsNIHSS imbalance since it seems the most likely factor to have impacted results and it has received widespread attention as potentially invalidating the study results. However, because of the high synergy between Age and BsNIHSS, any analysis of one must involve and impact the other, seriously complicating this process.

8.1.2 Baseline NIHSS Imbalance: The table illustrating this imbalance, shown below, demonstrates that, in the first quintile (NIHSS 0-5), 72% of the 58 patients were randomized to t-PA therapy. This was an unexpected observation since, in a randomized trial, it would be expected that within each quintile, there would be approximately equal numbers of patients randomized into each treatment group. The imbalance in the first quintile is countered in the second and fifth quintiles where the corresponding percents are 45%. The third and fourth quintiles are balanced.

ALL PATIENTS*

Treatment Group	Baseline NIHSS Quintiles					TOTAL
	0 - 5	6 - 10	11 - 15	16 - 20	> 20	
Placebo	16 (28%)	83 (55%)	66 (50%)	70 (49%)	77 (55%)	312 (50.2%)
t-PA	42 (72%)	67 (45%)	65 (50%)	73 (51%)	63 (45%)	310 (49.8%)
Total	58	150	131	143	140	622

* p-value for test for imbalance = 0.005

This imbalance led critics of the NINDS study to suggest that it could have affected the overall study results. In this section we pursue this matter. As is illustrated in the two tables immediately following, the majority of the imbalance occurred among patients randomized in the stratum defined by the time from onset to treatment (OTT) being between 91 & 180 minutes. We could not establish that this fact contributed in any substantial way to our analysis and elected to proceed with our description of what is a rather complex analysis with no further reference to the relationship of the BsNIHSS(5) categorical variable to the OTT variable.

OTT ≤ 90 PATIENTS*

Treatment Group	Baseline NIHSS Quintiles					TOTAL
	0 - 5	6 - 10	11 - 15	16 - 20	> 20	
Placebo	9 (41%)	37 (55%)	31 (44%)	37 (48%)	31 (47%)	145 (48.0%)
t-PA	13 (59%)	30 (45%)	39 (56%)	40 (52%)	35 (53%)	157 (52.0%)
Total	22	67	70	77	66	302

* p-value for test for imbalance = 0.001

OTT > 90 PATIENTS*

Treatment Group	Baseline NIHSS Quintiles					TOTAL
	0 - 5	6 - 10	11 - 15	16 - 20	> 20	
Placebo	7 (19%)	46 (55%)	35 (57%)	33 (50%)	46 (62%)	167 (52.2%)
t-PA	29 (81%)	37 (45%)	26 (43%)	33 (50%)	28 (38%)	153 (47.8%)
Total	36	83	61	66	74	320

* p-value for test for imbalance = 0.7

8.1.3 Outcomes in NIHSS Quintiles: In Table 8.1.1, we illustrate for each of the 4 outcome variables (B:Barthel, R:Rankin, G:Glasgow and N:NIHSS) the numbers of unfavorable (UF) and favorable (F) responses within each of the five classes (hereinafter referred to as quintiles) as defined by BsNIHSS(5). Also, in this table we present three measures of comparison of t-PA to Placebo within each quintile for each outcome scale. These three measures are:

D% = difference in % favorable outcome (t-PA minus Placebo)
RR = ratio of these favorable outcome percents
OR = ratio (t-PA/Placebo) of the odds of a favorable outcome.

For each outcome variable, in the 1st quintile (Q_1), all patients, whether placebo or t-PA treated, had an excellent chance for a favorable outcome. It is also interesting to note that, for three of the four outcome variables, the placebo group does modestly better than the t-PA group in Q_1, although this is not statistically significant. For each outcome variable, the proportion of favorable outcomes for both treatment groups decreases with increasing NIHSS category. However, for categories Q_2 through Q_5, all indicators of treatment effectiveness favor t-PA therapy for all the outcome variables. At the upper end of the NIHSS score, indicating more severe strokes, the likelihood of a favorable outcome is quite poor and the absolute difference in favorable outcome (D%) is much smaller in Q_5 (3.3-5.6% for the four outcomes) than in Q_2 - Q_4. However, even in Q5, the RR and OR indicators of treatment effect show values in favor of t-PA not much different than those seen in Q_2, Q_3, and Q_4.

8.1.4 Outcomes in Age Quintiles: As we have noted, age was the second most significant variable related to favorable outcomes (Tables 5.10 through 5.13). In Table 8.1.2, which is similar to Table 8.1.1 but applies to Age Quintiles we saw that effect. Overall the favorable outcome percentages (regardless of treatment) decrease with increasing age although in the age decade 65 – 74, which we divided into two groups because of the number of patients in that age decade, we saw no evidence of a decrease. However, patients in the final group (age 75+) clearly have the smallest chance of a favorable outcome.

From the standpoint of treatment effect, the differences in the favorable outcome percentages are positive, in favor of t-PA, for all outcome measures for all quintiles. The odds ratios show a

decreasing trend with increasing age group for some outcome variables. However, in the light of the major interaction of Age with BsNIHSS we did not investigate this further.

8.1.5 Age by Baseline NIHSS Interaction: We stated in the introduction to this section that the relationship between Age and NIHSS was synergistic. From the models shown in Tables 5.17 through 5.21 we see that the coefficient of the interaction term is always negative. In a setting such as this where the influence of the two variables is also negative, this means that each gains more influence as the other increases in value. For example, consider the Barthel covariate model (Table 5.17). The estimated coefficients suggest that, for a patient aged 50, a one unit increase on the NIH Stroke Scale results in a 12% decrease in the odds of a favorable outcome while, for a patient age 80 a one unit increase in the NIH Severity Scale results in a 22% decrease in the odds of a favorable outcome.

8.1.6 Model-Based Assessment of Baseline NIHSS Imbalance: We will now review formal attempts to evaluate the role of BsNIHSS(5) on the assessment of treatment effect. In Section 5.4 we describe our process of identifying the variables significantly related to the occurrence of a favorable outcome that were included as adjusting covariates in the treatment effect models. The continuous version, BsNIHSS, with a single degree of freedom in the models, was always highly significant ($p<0.0001$) as a predictor of favorable outcome and was more significant than BsNIHSS(5), with four degrees of freedom in the models, for all but one of the analyses. For this reason, and to avoid using a grouping that was identified from data exploration, we used the continuous BsNIHSS, in our principal analyses. Thus, all analyses other than those to be discussed herein are based on the use of BsNIHSS, a single degree-of-freedom variable, in our logistic and global models to adjust for the impact of baseline stroke severity, as measured by NIHSS, on 90 day outcome.

8.1.6.1 Baseline NIHSS Analysis: In what follows we first show that the choice of which of these versions of NIHSS to use is irrelevant. The choice affects the estimated coefficients of other variables in the model by only a small amount and has very little effect on the estimated t-PA vs. Placebo favorable outcome odds ratios.

For each outcome variable, Table 8.1.3 shows parameter estimates with standard errors and/or p-values for four different models. For each outcome variable, the first two models include BsNIHSS and differ only as to inclusion of the t-PA indicator variable. The second two models include BsNIHSS(5) with four degrees of freedom. For each outcome measure, the t-PA vs. Placebo odds ratios estimated by models with the different versions of NIHSS are – for all practical purposes – identical. However, of greater importance is the fact that the coefficients or p-values of all of the other covariates are essentially the same whether BsNIHSS or BsNIHSS(5) is used.

In Section 5.5.2.1 we report that t-PA did not interact with the three degrees-of-freedom variable associated with the variables BsNIHSS, Age, and Age*BsNIHSS for any of the outcome variables individually or in the Global analysis. Because of the strong interaction between Age and BsNIHSS, it is necessary to treat these three variables collectively as a triumvirate. The presence of an interaction of t-PA with the BsNIHSS, Age, Age*BsNIHSS triumvirate, would mean that patients in some group(s), defined by the combination of stroke severity as measured by BsNIHSS and Age, responded differently to t-PA treatment than patients in other groups. That neither we, nor the NINDS investigators, found statistically

significant evidence of an interaction, does not imply its absence. Allowing for this caveat, the absence of a statistically significant interaction indicated that there was no evidence of a differential t-PA treatment effect related to baseline stroke severity. This finding indicates that the baseline stroke severity imbalance did not affect the study outcome.

8.1.6.2 BsNIHSS Quintile Specific Odds Ratios: The quintile-specific OR estimates for each of the outcome measures are documented in Table 8.1.1. The unadjusted global OR estimates for the five quintiles, Q_1 through Q_5, were: 0.9, 2.4, 1.9, 1.6, and 1.7 respectively. For each outcome measure, the ORs favor t-PA except in Q_1 where the values are all close to 1. Even in Q_5, the OR is in favor of t-PA therapy. Tests of the hypotheses that the odds ratios are equal across the quintiles, adjusting for the stratification factors, were not statistically significant. However, when adjusting for all covariates, these tests are complicated by the presence of a highly significant interaction between age and baseline NIHSS. In this context, such tests involve a complex interaction with 9 degrees of freedom. The table below documents the results of the chi-square tests for models including both the four and nine degrees of freedom tests. These analyses demonstrate that for each of the four outcome measures and the global analysis there was insufficient evidence to declare a difference in treatment effects (ORs) across the five quintiles.

Treatment Group	**Test for Equal ORs**			
	Adjusted for stratification factors*		**Adjusted for all covariates#**	
	Chi-square (4 DF)	**p-value**	**Chi-square (9 DF)**	**p-value**
Barthel index	4.27	0.37	5.41	0.80
Modified Rankin scale	2.69	0.61	5.54	0.78
Glasgow outcome scale	2.89	0.58	6.09	0.73
NIHSS	0.66	0.96	2.96	0.97
Global analysis	2.30	0.68	3.65	0.93

* Stratification factors: study part, center, OTT
# All covariates: stratification factors + history of diabetes, preexisting disability, age, baseline NIHSS, and age*baseline NIHSS

After a detailed examination of all of these models the two most important messages are; (i) with the exception of Q_1, the t-PA to placebo odds ratio estimates are uniformly greater than 1, indicating a superiority of t-PA over placebo in patients with a baseline NIHSS score of > 5, and (ii) If we focus on the age of 70 – essentially the median age of the study group – the t-PA to placebo OR estimates from the model containing the interactions are not much different from the estimates from the no interaction models. Thus, we conclude that there is no evidence that the baseline stroke severity grouping defined by BsNIHSS(5) has identified a group of patients who respond differently to t-PA therapy than the study cohort in general. All earlier caveats about the proper interpretation of non-significant tests of no interactions continue to apply.

8.1.7 An Alternative Variable: The analyses described above are complicated by the need to include among the "adjusting" variables term(s) defining an interaction between Age and some version of baseline NIHSS. In Section 5.4, it was noted that in some models, the inclusion of the interaction term literally made the terms corresponding to Age and BsNIHSS appear insignificant. In one final effort to examine this complex question of the impact of the baseline NIHSS imbalance within models that of necessity include this interaction among the adjusting variables, we defined an alternative variable as the simple product of age times BsNIHSS and classified the patients into quintiles on the basis of that variable. Table 8.1.4 illustrates, for each outcome measure, the rates of favorable outcomes for Placebo and t-PA treated patients within each of these quintiles.

The role of this variable defined by multiplying age by BsNIHSS is not easy to understand and a few examples may help. A 75 year old patient with an NIHSS of 4 would have a value of 300 for this Age*BsNIHSS product, placing him/her in the middle of the first (lowest) quintile. Similarly, a 65-year-old patient with an NIHSS of 10 would have a score of 650, placing him/her in the middle of the 2nd quintile. A 75 year old whose NIHSS is 12 or 13 would be in the 3rd quintile and one with an NIHSS of 17 would be in the 4th quintile. The 5th quintile will contain mostly very elderly patients with a very high NIHSS (e.g. an 85 year old with an NIHSS of 24).

The table illustrates that the odds of a favorable outcome decrease dramatically as we look from the 1st to the 5th of these quintiles. For example, for the Barthel index, the odds of a favorable outcome for patients whose combination of age and NIHSS at baseline place them in the first of these quintiles is 4.4, indicating that such a patient has a very good chance of a favorable outcome. While the odds are not as high for the other scales, the estimates of odds are all greater than one, indicating that by any of the 4 scales, patients in the first quintile are more likely to have a favorable outcome than not. In contrast, in the fifth quintile the prospects are grim with the odds of a favorable outcome ranging from 0.14 for the Barthel scale to 0.02 for the NIHS scale.

The above assessment of the likelihood of a favorable outcome notwithstanding, the comparison of t-PA therapy to Placebo is remarkably constant over the 5 quintiles. We will briefly discuss the consequences of using this classification of patients into the 5 groups according to the product of age by NIHSS in our formal statistical models as an alternative to the more complicated use of Age, some version of BsNIHSS and the interaction between the two as adjusting variables in the models.

The sequence of Tables 8.1.5, 8.1.6, 8.1.7, and 8.1.8, one for each outcome measure, displays the results of this analysis. For each outcome measure, the table actually consists of summaries of 7 different models. The first 4 of these models are the same as seen in earlier table in this section and involve the standard models first with BsNIHSS and then with BsNIHSS(5) in the models. The last three models use the 4 degrees-of-freedom variable necessary to account for the 5 quintiles of the age*NIHSS product variable. The final model is testing the "no-interaction" hypothesis regarding the interaction between t-PA and the 5 quintiles. As we had observed in Table 8.1.4, there is no evidence of an interaction. Indeed, the 4 p-values for the no-interaction hypotheses are: 0.99, 0.98, 0.94, and 0.91, for Barthel, Rankin, Glasgow and NIHSS, respectively. The other point to make is that the use of this variable has had minimal effect on the other coefficients in the model or on the t-PA to Placebo Odds ratio estimate.

8.1.8 Influence of the Age by Baseline NIHSS Interaction on the t-PA Treatment Effect: As discussed previously (Section 5.4.1), our analyses determined that both stroke severity and age were negatively and significantly related to a favorable outcome and that these two variables interacted in a highly significant way in predicting a favorable outcome, with the combination of advanced age and a severe stroke reducing the chances of a favorable outcome to an extremely low level. While many who care for acute stroke patients recognized this, it is a phenomenon that, heretofore, does not seem to have been quantified. Consequently, all models aimed at comparisons of the t-PA and control groups must include variables for age, NIHSS and their interaction to adjust for these effects.

The fundamental concern in this discussion is whether the data in the t-PA trials provided any evidence that this relationship between age, stroke severity, and the chances of a favorable outcome, had an impact on the effect of t-PA. The problem of estimating the interaction between t-PA and stroke severity at randomization (as estimated by the baseline NIHSS), within models in which NIHSS and age interact, is complex, requiring multiple degrees-of-freedom. We summarized three methods of analysis and the tests for interaction were not statistically significant in any of them.

However, as summarized by Brookes et al. in a recent publication[28], interaction tests are typically very underpowered in studies where the sample size was determined on the basis of main effect tests. To quantify the power of interaction tests and sub-group analyses in randomized trials these authors performed simulations. They reported that a clinical trial with 80% power to detect a specified main effect would have only a 29% chance (power) of detecting an interaction of the same magnitude as the main effect. They also reported that the overall sample size would have to be quadrupled to increase the power of the interaction test in the above situation to 80% and that, if the desired detectable interaction was only 20% of the main effect, the sample size would have to be increased “dramatically.”

In our study, the less complex, one degree-of-freedom, tests of no interaction have poor power (Section 5.5.2.2). Therefore, we concluded that the non-significant results of the complex tests were likely due to low power and included warnings in our conclusions that these results should not be taken as evidence of lack of an interaction. However, because of the importance of the baseline imbalance issue, we undertook further analyses to determine if more specific information regarding these interactions could be obtained. The results of this further examination in the three methods of analysis are summarized below.

1. In order to focus directly on the simultaneous impact of age and severity on outcome and t-PA effect, we created a new regressor by multiplying age by baseline NIHSS value (Section 8.1.7). We then subdivided the patients into quintiles of this predictor. As expected, those in the lowest quintile (relatively younger with less severe stroke) fared much better than those in the highest quintile (relatively older with more severe stroke) regardless of their randomization group. However, when we formally tested whether the effect of t-PA was the same for all five quintiles, the corresponding p-value, based on a 4-degree-of-freedom chi-square of 0.37, was $p = 0.99$ (within the global model while adjusting for all the covariates other than age and NIHSS which were included in this artificial variable). Thus, we conclude that there is no evidence of a difference in the t-PA to placebo comparison over these quintiles. The fact that the chi-square value is so small indicates that the estimate of any difference in effect based on an analysis of these data is so small as to be clinically irrelevant. The analyses for the 4

individual outcome measures (Barthel, Rankin, Glasgow, NIHSS) were just as dramatically null.

Barthel: Chi-square = 0.3415, p = .99
Rankin: Chi square = 0.4499, p = .98
Glasgow: Chi square = 0.7663, p = .94
NIHSS: Chi square = 1.0022, p = .91

2. With age and baseline NIHSS each being treated as continuous variables (Section 5.5.2.1), the presence of the highly significant interaction between them is manifested by the inclusion of a variable identical to the one discussed above (i.e., the product of age and NIHSS). In such models, the hypothesis of no interaction between NIHSS and t-PA requires the addition of three new terms in the model; hence the hypothesis is tested based on a chi-square value with 3 degrees-of-freedom. Again referring only to the global model, the chi-square value (df = 3) was 2.72, corresponding to p = 0.44. Here, again, the lack of statistical significance is not nearly as important as the fact that, among the 3 degrees of freedom there is no evidence of a meaningful indication of effect, since even a one degree-of-freedom test requires a chi-square of 3.84 or higher to provide such evidence. The analyses of the 4 individual outcome measures led to similar conclusions.

Barthel: Chi-square = 3.5645, p = .31
Rankin: Chi square = 5.4728, p = .14
Glasgow: Chi square = 4.3866, p = .22
NIHSS: Chi square = 2.8117, p = .42

3. Because tables had been published with a subdivision of the baseline NIHSS scores into quintiles, which demonstrated an imbalance of assignment of patients to t-PA and placebo treatments, we also performed analyses with patients in these groups (Section 8.1.6.2). Such analyses require that the models contain 1 degree-of-freedom for age, 4 degrees-of-freedom to identify the 5 groups, and 4 degrees-of-freedom to describe the interaction of age and NIHSS. Thus, testing the interaction of t-PA with NIHSS (and, of necessity, age) requires that 9 degrees-of-freedom be added to the models. For the global analysis of this no interaction hypothesis, the chi-square value (df = 9) was 3.65, with an associated p-value of .93. Again, we conclude that these analyses provide no evidence of an interaction between t-PA and stroke severity. The analyses of the 4 individual outcome measures led to similar conclusions.

Barthel: Chi-square = 5.4077, p = .80
Rankin: Chi square = 5.5400, p = .78
Glasgow: Chi square = 6.0910, p = .73
NIHSS: Chi square = 2.9600, p = .97

Finally, in a further assessment of the impact of stroke severity on the t-PA effect for fixed ages, we used an argument described mathematically in the appendix below to estimate the t-PA effect associated with a 5 unit increase in NIHSS for individuals at 60, 70, and 80 years of age. The results of these investigations are shown in the table below. For each of these ages, as NIHSS increases by 5 units, the odds ratio estimates (in favor of t-PA) increases. In the Global analysis, for age = 60 years, the increase is 12% (95% CI: -15% to 49%); for age = 70 years, the increase is 32% (95%CI: -6% to 86%); for age = 80 years, the increase is 56% (95%

CI: -6% to 156%). These estimates, based on the variance estimates, suggest that a 5-unit higher level of stroke severity is associated with an increase in t-PA benefit with the magnitude of the increase rising with age. While none of these percent increases was significant at the 5% level (very generous in the context of the many tests being performed), some were close.

	AGE	theta	Std. Err of theta	ORR = OR(N+5) / OR(N)	95 % CI of ORR lower	upper	Minimally Detectable ORR
Global							
	60	0.114	0.1439	1.121	0.85	1.49	1.63
	70	0.279	0.1749	1.322	0.94	1.86	1.82
	80	0.444	0.2529	1.559	0.95	2.56	2.37
Barthel							
	60	0.004	0.1629	1.004	0.73	1.38	1.75
	70	0.242	0.1727	1.274	0.91	1.79	1.80
	80	0.480	0.2655	1.616	0.96	2.72	2.48
Rankin							
	60	0.205	0.2251	1.227	0.79	1.91	2.16
	70	0.277	0.2641	1.319	0.79	2.21	2.47
	80	0.350	0.3661	1.418	0.69	2.91	3.49
Glasgow							
	60	0.052	0.1675	1.053	0.76	1.46	1.77
	70	0.178	0.1949	1.195	0.82	1.75	1.95
	80	0.304	0.2977	1.355	0.76	2.43	2.77
NIHSS							
	60	0.216	0.1819	1.241	0.87	1.77	1.86
	70	0.394	0.2354	1.482	0.93	2.35	2.24
	80	0.571	0.3525	1.770	0.89	3.53	3.34

Using a more rigorous, two-sided 0.01 level of significance, we estimate that the minimally detectable (80% power) changes in the probability of a favorable outcome would be 63%, 82% and 137% respectively. These suggest that if a 5-unit increase in stroke severity did produce increases in t-PA effect of the magnitude indicated the analyses would have had an 80% chance of being significant. But, there was no statistically significant evidence of an interaction despite this unexpectedly robust power.

In summary, this study was not powered to detect subgroup interaction differences in the t-PA treatment effect. Nonetheless, our analyses provide no evidence that the effect of t-PA is clinically different for acute stroke patients with different levels of stroke severity. A post hoc power analysis allows us to conclude that there was no clinically important interaction between

baseline NIHSS and t-PA. Therefore, we conclude the baseline imbalance in NIHSS played a very minor role in the estimated benefit of t-PA.

Appendix

The equations below define some age specific assessments of the impact of increases in severity as measured by baseline NIHSS. In this assessment we treat NIHSS and age as continuous variables and ignore all covariates not involved in this assessment (of course they are not ignored in the actual analysis). Define an indicator variable (t-PA = 1 for those on t-PA and = 0 for those on Placebo) and specify an interaction model as follows:

$$\text{Log(OR)} = \beta_0 + \beta_1 \text{ Age} + \beta_2 \text{ NIHSS} + \beta_3 \text{ Age*NIHSS} + \beta_4 \text{ t-PA}$$

$$+ [\beta_5 \text{ Age} + \beta_6 \text{ NIHSS} + \beta_7 \text{ Age*NIHSS}]\text{*t-PA}.$$

The odds ratio (OR) is the ratio of the odds of a favorable outcome for those treated with t-PA to the odds of a favorable outcome for those treated with the placebo.

The last three terms in this model describe the interaction of t-PA with stroke severity because of the complex relationship of age and stroke severity with the likelihood of a favorable outcome. For a fixed age (A), and a fixed NIHSS (N), the log of the OR comparing t-PA to placebo is:

$$\text{Log(OR given A \& N)} = \beta_4 + \beta_5 A + \beta_6 N + \beta_7 A*N.$$

The no interaction hypothesis is that: $\beta_5 = \beta_6 = \beta_7 = 0$. If true, the t-PA to Placebo OR is $\exp(\beta_4)$. An interaction exists if any of these three betas are non-zero and for the non-null model we estimate their values from our data. Keeping age fixed at A and changing NIHSS from N to N + Δ,

$$\text{Log(OR given A \& N} + \Delta) = \beta_4 + \beta_5 A + \beta_6(N + \Delta) + \beta_7 A*(N + \Delta)$$

If these two equations give the same answers for age A, the t-PA to Placebo ORs are the same no matter what the NIHSS level is. The difference between these two equations, which we arbitrarily call Θ, is the Log of the ratio of OR at severity level N + Δ, call it OR(N + Δ), to the log of the OR at severity level N, call it OR(N). Define ORR(Δ) = OR(N + Δ)/OR(N), then:

$$\Theta = \beta_6\Delta + \beta_7 A*\Delta = \Delta\{\beta_6 + \beta_7 A\} = \text{Log(OR}(N + \Delta)) - \text{Log(OR}(N))$$

$$= \text{Log(ORR}(\Delta)).$$

If we test and reject the null hypothesis that the expected value of Θ is zero, we would have established that, at least for A-year olds, there is an interaction between t-PA and NIHSS.

From the output of our models, we obtain the estimates of β_6 and β_7 as well as estimates of their variances and the covariance between them. This allows us to estimate the variance and hence, the standard error of the corresponding estimate of Θ. Then we obtain confidence intervals for Θ, and, using these empirical variance estimates, estimate the minimally

detectable value of Θ and hence, exp(Θ), the minimally detectable ratio of odds ratios, ORR(Δ), corresponding to a difference of Δ on the NIHSS scale for a fixed value, A, on the age scale.

Treating Θ as its own estimate, $Var(\Theta) = \Delta^2\{Var(\beta_6) + 2ACov(\beta_6, \beta_7) + A^2 Var(\beta_7)\}$

In the associated table we have summarized what we find in this situation for three choices of A (60, 70, & 80) and $\Delta = 5$.

8.1.9 Summary and Conclusions: After a thorough evaluation of this issue, we found no evidence that the imbalance in the distribution of baseline NIHSS between the treatment groups had either a statistically or clinically significant effect on the study results. We have determined that the original models using both Age and BsNIHSS as continuous variables properly adjust for the complex roles played by these two variables, both so strongly (negatively) related to the likelihood of a favorable outcome. There was a strong interaction between age and baseline NIHSS in the Global analysis and in the analyses of each of the four outcome measures. The likelihood of a favorable outcome was particularly low in patients older than 70 who had a baseline NIHSS more than 20. However, there was no evidence of any Age by BsNIHSS subgroup responding significantly differently to t-PA treatment than the study group at large.

8.2 Onset to Treatment Time

As detailed in Section 4.2, the NINDS study was stratified on onset to treatment time (OTT) with plans for an equal number of patients to be randomized with an OTT < 91 minutes and an OTT >90 minutes but not greater than 180 minutes. For the sake of the following discussion we will refer to these two strata as the first (#1) and second (#2) respectively.

8.2.1 Restricted Randomization: Following the onset of symptoms there were variable delays prior to patient arrivals in the emergency departments. Subsequently, further delay resulted before a patient could be consented, randomized and treated due to the requirements of the study protocol, including the performance of a CT scan to determine patient eligibility. Consequently, therapy was initiated on few patients in less than an hour after symptom onset. As a result, the NINDS investigators found it much easier to enter patients into the second stratum than the first. In order to assure satisfaction of the treatment protocol that equal numbers of patients be randomized within the two strata at each center, it was necessary for a restriction to be placed on the entry of patients into the study within stratum #2. Specifically, each center was instructed that whenever the number of patients in stratum #2 exceeded the number in stratum #1 by three (3), they could not randomize a patient into stratum #2. This design modification worked quite well with only a minor imbalance; 302 patients were randomized into stratum #1 and 320 into stratum #2. Of course, this quota rule resulted in 267 otherwise eligible, patients not being entered into the clinical trial[29]. The review committee has no reason to believe that this recruitment restriction in any way violated the randomization process or that it was anything more than an inconvenience in the conduct of the study.

8.2.2 Distribution of OTT: The NINDS investigators examined in some detail the role of the actual value of OTT (not the dichotomized version) on the effectiveness of t-PA and concluded that their study demonstrated that earlier treatment was better[19]. In that manuscript they displayed a histogram of OTT values demonstrating that a high proportion of the patients entered in stratum #1 were entered with values of OTT between 80 and 90 minutes (see Figure 1 below).

Indeed, 150 (50%) of the patients randomized into stratum #1 had values of 89 or 90 minutes. We present the distribution of all OTT values in the form of a cumulative distribution function (see Figure 2 below) showing the sharp rise as the OTT values approach 90 minutes, and the cumulative percent approaches 50%.

Considering the questionable precision with which many patients' "time of onset" must have been estimated and the intense setting of an emergency department the precision of these OTT values and their accumulation just before 90 minutes is questionable. Consequently, the Review Committee is somewhat skeptical of the analysis reported wherein the NINDS investigators used the OTT variable as a continuous variable[19] rather than as the protocol mandated dichotomized version.

Figure 1.

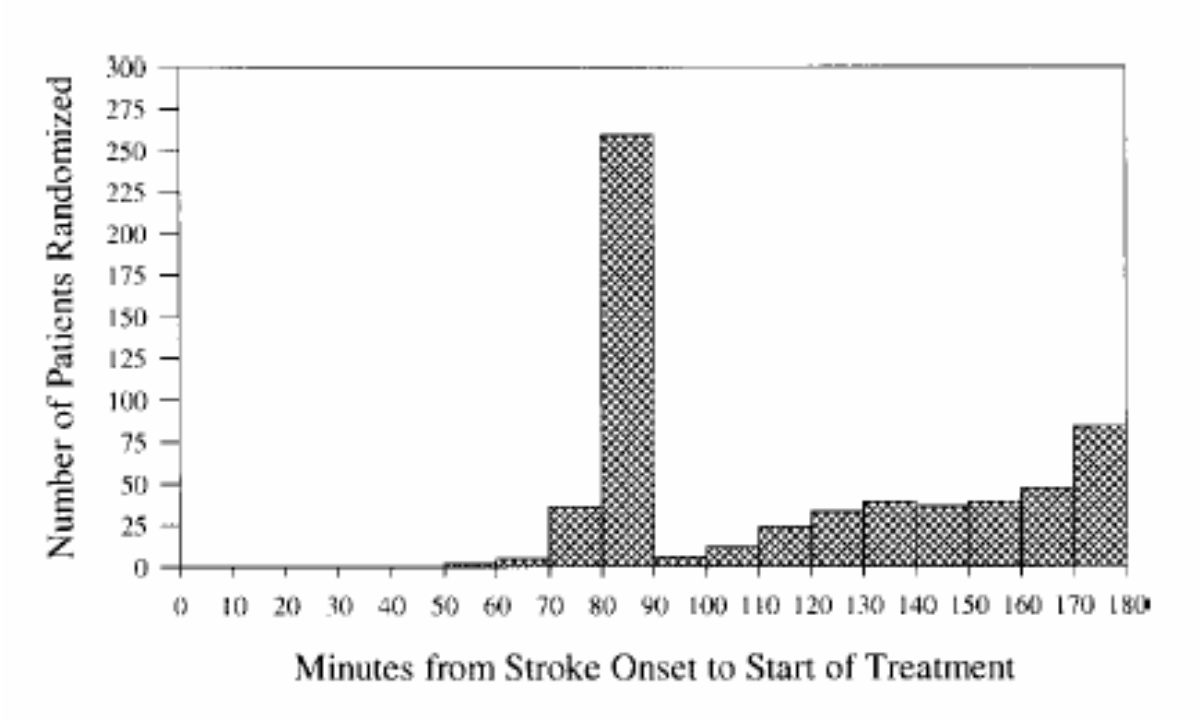
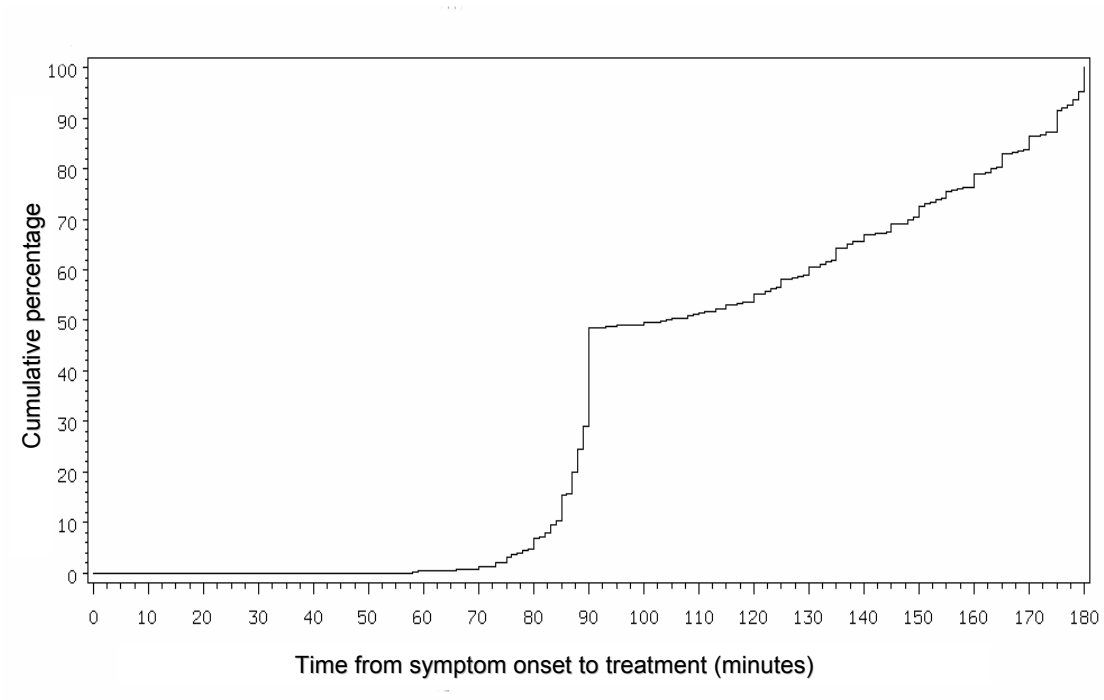


Figure 2.



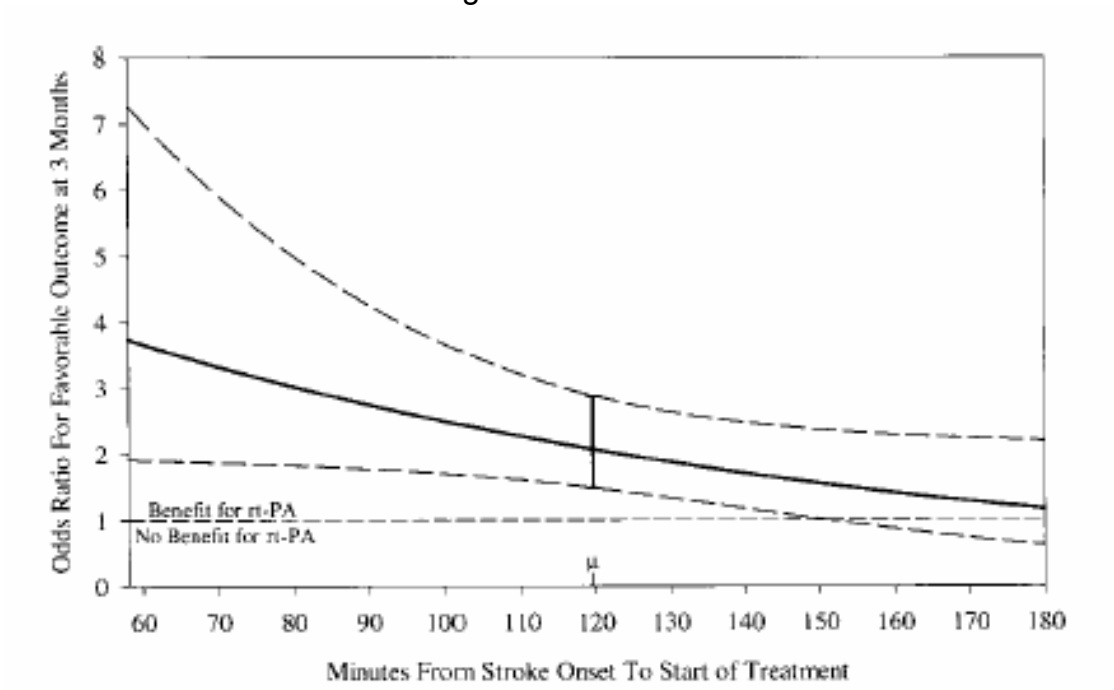
8.2.3 Does t-PA Effectiveness Decrease with Increasing OTT?: In order to investigate the issue of whether the NINDS study can lead to the conclusion that earlier t-PA therapy is better than later treatment, we performed a number of analyses.

8.2.3.1 OTT by t-PA Interactions (1): The first analysis considers whether the variable indicating OTT stratum interacted with the treatment group indicator in predicting outcomes.

We report these interaction tests in Tables 5.17 through 5.21 and discuss them in Section 5.5.2. For the Barthel, Rankin, Glasgow, NIHSS, and Global analyses, the tests of no OTT*t-PA interaction had p-values of 0.17, 0.87, 0.90, 0.22, and 0.19 respectively. These interaction tests, however, have 80 % power of detecting that two ORs differ only if their ratio is between 3.0 and 4.0, depending on outcome scale. Thus, if the true OR in stratum #2 is 1.0, the true OR in stratum #1 would have to be between 3.0 and 4.0 in order for the NINDS study to have 80% power. Put another way, these interaction tests had an 80% chance of being significant only if the two true ORs differ by a factor of 3 or more. We observed no such dramatic relative difference in ORs and the lack of statistical significance for the interaction tests is not surprising. Importantly, an interaction less than 3.0 may be clinically important, but the study has insufficient power to detect differences of such magnitude.

In the aforementioned article[19], the NINDS investigators presented a figure suggesting a range of ORs from 4.0 to 1.0 between OTT values of 60 and 180 minutes (see Figure 3). However, almost no patients had an OTT ~60 minutes. Indeed < 10% had OTT values as large as 82 mins, with a similar percent having OTT values between 176 and 180 minutes. According to the figure, the OR corresponding to 82 is <3 whereas the OR corresponding to 180 is > 1. Therefore, their own best estimate of OR differences suggests a less than 3-fold change over a reasonable OTT range, a change that the study has little power to detect.

Figure 3.



8.2.3.2 OTT by t-PA Interactions (2): Because of all the attention the NINDS publication on this topic has received, we pursued the issue further, working with indicator variables and substratifications rather than a continuous OTT. Since nearly all the OTT values in the first stratum were greater than 60 minutes, suggesting that the stratum was actually only about 30 minutes wide, we elected not to partition that stratum further. However, we did partition stratum 2 several ways to see if the “trend” reported by the investigators can be supported by an alternative analysis of the data. To facilitate the following discussion consider Table 8.2.1.

In that table, for each of the 4 outcome variables, there are 8 columns in three groups. The first column (labeled OTT: 0 – 90) summarizes data from the first stratum and the last column (labeled OTT: 91 – 180) presents the same summaries for the second stratum. Columns 2 through 5 correspond to a partition of the second stratum into 4 substrata, each containing nearly the same total number of randomized patients (i.e. quartiles of the distribution of OTT in stratum #2). Columns 6 & 7 contain the sums of columns 2 & 3 and 4 & 5 respectively. Each column contains 4 tables providing, for each outcome measure, the number of favorable and unfavorable responses among the placebo and t-PA patients randomized within the stratum (substratum) defined by that column. Two summary measures are then provided for each table. They are the odds of a favorable response among the placebo patients (PL odds F) and the t-PA vs. Placebo odds ratio (OR).

The most interesting aspect of this table is found in the substratum labeled OTT: 91 – 133 which summarizes information from the 81 patients randomized within stratum #2 whose OTT values were assessed to be between 91 minutes and 133 minutes inclusive. Of these patients, 50 (62%) were randomized to placebo. This is different from the expected 50% ($p = 0.035$). There is nearly perfect t-PA to placebo balance in the other three substrata and at the conclusion of the study 52% of the patients in stratum #2 were randomized into the placebo group. This still represents an excess of 7 patients from the expected 50%, almost all of which is attributable to the unexplained imbalance among those randomized with OTT values between 91 & 133. Of further interest in this regard is that on all outcome scales those 50 placebo patients randomized with OTT values between 91 and 133 had by far the lowest odds of a favorable outcome of any of the OTT substrata. As a result, on the OR scale this substratum stands out with exceptionally high values favoring t-PA. However, if the placebo patients in that substratum had odds for a favorable outcome more in line with the rest of the study, the corresponding ORs would be between one third and one half the quoted values.

The foregoing observations are relevant to the decreasing trend in OR as reported by the NIHSS investigators[19] seen here in Figure 3 . The inexplicable and likely artificial elevation of the OR during the 91 to 133 minutes interval could tilt the OR scale up at the earlier part of stratum #2 resulting in an estimate of a negative slope with increasing OTT. Furthermore, for the Barthel & Glasgow outcome measures, the period OTT: 174 – 180 demonstrates what appear to be equally inexplicable elevations in the odds of a favorable outcome among the placebo patients. These clearly contributed to the lower estimates of the OR in that period and would have also contributed to the negative slope estimate.

8.2.3.3 OTT by t-PA Interactions (3): These observations about the nature of the relationship between the distributions of favorable response among the placebo patients and OTT notwithstanding, we carried out two sets of additional formal logistic and GEE regression analyses, including all final covariates, which we summarize briefly in this paragraph. Both analyses involved partitioning the second stratum into substrata as illustrated in Table 8.2.1. In the first analysis we divided it into two substrata (OTT: 91-154 & OTT: 155-180) each containing 160 patients. In the second analysis we used the 4 substrata defined by the quartiles discussed earlier. In both analyses, stratum #1 was used as the comparison group. In the first analysis, 2 degrees-of-freedom were required to separate the resulting three OTT classes and in the second analysis 4 degrees-of-freedom were required to separate the 5 OTT classes. In the first step of each analysis we estimated the t-PA to Placebo odds ratio to assess the impact of this change in the OTT variable on the OR estimates. The results (first analysis, second analysis) for each outcome variable are: Barthel (2.19, 2.18), Rankin (2.43,

2.39), Glasgow (2.13,2.11), NIHSS (2.21, 2.22) and Global (2.14, 2.14). These results do not differ from each other nor do they differ from the adjusted odds ratios reported in Section 5.5.1.

The second step of the analyses was to determine whether t-PA and OTT interacted with each other in separate models each containing one of these two versions of OTT. For the first analysis, with the 2 degree-of-freedom OTT variable, the results of the no interaction test may be summarized as: Barthel ($p = 0.14$), Rankin ($p = 0.08$), Glasgow ($p = 0.31$), NIHSS ($p = 0.39$) and Global ($p = 0.13$). For the second analysis (4 dfs) the results are: Barthel ($p = 0.17$), Rankin ($p = 0.21$), Glasgow ($p = 0.15$), NIHSS p = 0.17) and Global ($p = 0.06$). Clearly these analyses failed to identify any significant interaction between the treatment and OTT variables (as did the fundamental analysis reported earlier in this section) that is, no collection of patients randomized at any of the five OTT levels discussed can be said to have a significantly different response to t-PA therapy than any other group.

8.2.4 Summary and Conclusions: In light of these results, the substantially nonlinear nature of the distribution of OTT when considered as a continuous variable, and the idiosyncratic distribution of favorable response rates among the placebo patients, we conclude that the data provided by this study failed to support a conclusion that the effect of t-PA therapy diminishes with increasing values of OTT within the protocol specified 3 hour time limit. However, this does not mean such a relationship does not exist, and further studies are needed to address the question of a differential t-PA treatment effect related to time from symptom onset to treatment. It is also important to recognize that the results from this study provide no data on the effectiveness of thrombolytic therapy administered to acute ischemic stroke patients more than 180 minutes after symptom onset.

8.3 Clinical Centers:

Randomization took place within each of 9 centers; however, one center randomized and treated only one patient who was followed by another center. The NINDS investigators considered those two centers as a single center. We therefore consider the study as involving 8 centers, or strata, and the "Center" variable carries 7 degrees-of-freedom in all of the models.

8.3.1 Center Comparisons of Favorable Outcome Rates: Center differences regarding such issues as recruitment and outcome are illustrated in Tables 8.3.1 and 8.3.2. These two tables have identical structure and support all of the statements regarding specific numerical values unless otherwise specified.

The capacity of the centers and their access to appropriate patients differed appreciably. Two centers (#'s 4 and 5) randomized 146 and 150 patients respectively (nearly 50% of the entire study population). The remaining centers randomized 103, 71, 62, 39, 37, and 14 patients. As was observed in Section 5.3 (Table 5.1), the chances of a favorable outcome, regardless of therapy, varies by outcome variable. The overall favorability percents are 45%, 35%, 38%, and 27% for the Barthel, Rankin Glasgow and NIHS outcome measures respectively, corresponding to odds of favorability of 0.82, 0.53, 0.62 and 0.38. Across the centers the percent favorable ranges over about 25 percentage points with Center 7 always low, but not always the lowest and Center 3 always the highest.

Chi-square tests (not presented) of the hypothesis that the rates of a favorable outcome are the same over all centers were not significant. Furthermore, in the final models described in Sections 5.4 and 5.5, the 7 degrees-of-freedom Center variables, always included because it is part of the study design, was never significant even after adjusting for all the other variables in the model. Thus, observed differences in the likelihood of a favorable outcome result from statistical variation and should not be taken as evidence of important, underlying center-to-center variation.

8.3.2 Center Comparisons of t-PA Effect: In Tables 8.3.1 and 8.3.2, we have listed for the 8 centers, within each of the 4 outcome measures, their associated numbers of favorable and unfavorable outcomes by treatment group with two measures of treatment effect. These are, the difference (delta) between the percents of t-PA and placebo patients experiencing a favorable outcome and the ratio (t-PA to Placebo) of the odds of a favorable outcome. In both tables the centers are ordered by their rank according to the t-PA by placebo odds ratio for the Barthel scale. Thus, center #4 with a Barthel OR of 2.77 and a delta favorable outcome percent of 24.8% is ranked first even though it does not rank first for all 4 scales. Center #7, with only 14 patients randomized, which had consistently among the lowest overall rates of favorable outcomes did rank last on all scales when comparing t-PA to placebo. Indeed, all of the center #7 estimated odds ratios were less than one by a considerable, although not statistically significant, margin, consistently indicating more favorable outcomes among the placebo treated patients than among the t-PA treated patients at that center.

The data in the tables suggest what appears to be considerable variability among the centers as regards the odds ratios comparing t-PA therapy to placebo. For the Barthel scale, for example, the maximum odds ratio of 2.77 (Center # 4) is nearly 9-fold higher than the

minimum of 0.33 (center #7). For the Rankin and Glasgow scales this ratio of maximum and minimum odds ratios is even greater. However, the 95% confidence intervals, most of which overlap the null value 1, indicate that very few of the within center odds ratio estimates are "significantly" greater than 1 and none is significantly less than 1. Most of these confidence intervals, especially those based on the centers with smaller numbers of patients randomized, are very wide, reflecting the substantial random error present in estimates obtained from such small numbers of observations.

8.3.3 Center by t-PA Interaction: These observations raise the question of whether there is evidence that the response of patients to t-PA therapy as estimated through the odds ratios comparing the response of t-PA treated patients to the response of placebo treated patients is different among the 8 centers. This is a verbal description of an interaction between the variable defining therapy and the variables defining Center and the question is formally addressed by introducing into the models upon which our comparisons of t-PA to Placebo are based an appropriate interaction.

We have already described and briefly discussed formal testing of those interactions in Section 5.5.2. Specifically, in Tables 5.17 through 5.21 we reported the p-values for these 7 degrees-of-freedom tests for each outcome variable as well as for the global analysis. These tests were all conducted within the models containing the covariates deemed appropriate for "adjusting" the treatment comparisons. The p-values reported were 0.16, 0.24, 0.17, 0.87 & 0.47 for the Barthel, Rankin, Glasgow, NIHSS, and Global analyses respectively and are based on Chi-squared statistics of 10.49, 9.12, 10.28, 3.20, and 6.61 respectively. Based on this lack of statistical significance, we conclude that there is little evidence of an important interaction. However, the study was not powered to detect interactions, so the lack of significance does not guarantee the absence of an interaction.

8.3.4 Estimates of Differences in Favorable Outcome Percentages: Motivated by the foregoing, we pursued the question of the influence of the "interactions" a step further. Using equation 2 from Section 4.3.7, which permits the estimation of the difference in favorable outcome percentages while weighting the individual contribution of groups of patients (in this case from the different centers) we estimated these differences based on three different scenarios. The first scenario is based on the crude data seen in Tables 8.3.1 and 8.3.2. Thus, for Barthel, Rankin, Glasgow and NIHSS scales, respectively, the direct estimates of the differences in percent favorable response (with 95% confidence intervals) are: 14.1% (6.4%, 21.9%), 16.3% (8.9%, 23.7%), 14.4% (6.8%, 22.0%), and 13.7% (6.8%, 20.6%). If we model these estimates without including an interaction term, forcing the OR estimates to be the same for all centers, but otherwise adjusting for all covariates (Section 5.4.3) the estimates become: 17.3%, 16.5%, 16.7% and 14.2%, respectively. Finally, if we estimate these differences using models that contain a t-PA by Center interaction, permitting the OR estimates to be different for the 8 centers, and thus taking into account the fact that some centers have an estimated negative difference, the estimates are: 19.1%, 15.3%, 17.4% and 14.6%, respectively.

A comparison of the several estimates of the differences (t-PA minus Placebo) in rates of favorable outcomes suggests that these estimated differences do not change notably as we go from direct estimates to estimates based on complex models. Specifically, if we allow the models to estimate different ORs for each center and use those ORs in the estimates of the difference we obtain difference estimates that are essentially the same as those obtained in

the absence of an interaction. Thus, using the most flexible and completely adjusted within-center estimates of ORs that are available in the estimation of the differences between t-PA and Placebo rates of a favorable outcome, the difference estimates are essentially the same as those obtained from the raw data. Thus, our position that there is a statistically and clinically significant net positive effect of t-PA remains.

8.3.5 Summary and Conclusions: We found no significant difference between the centers in the baseline characteristics of the patients. The likelihood of having a favorable outcome differed considerably between the centers, those with fewer patients often having the worst outcome. However, the between-center variation in t-PA treatment effect for either the global outcome, or the individual outcome measures, was not statistically significant and did not invalidate the trial results. Nevertheless, it will be important in future studies to identify the factors that lead to good outcomes at institutions administering t-PA to treat acute ischemic stroke patients. This information will be very helpful to other institutions that are looking to develop the resources needed to administer t-PA safely to acute ischemic stroke patients.

8.4 Stroke Subtype

8.4.1 Introduction: The NINDS investigators examined all pre-randomization records in an attempt to determine the ischemic stroke subtype into which each patient could be classified[1]. The result was a post-randomization classification of the patients into one of four subtypes: small vessel, cardioembolic, large vessel and other. In Table 4.2 and Section 4.1.3, we noted that the randomization of the patients into t-PA and Placebo treatment groups had resulted in a marginal imbalance regarding the subtype groups (p=0.064). In Table 4.2 and in Table 8.4.1 associated with this section, it can be seen that only the 273 cardioembolic patients were divided as nearly as possible equally between the two treatment groups. In contrast, 63% of the 81 small vessel patients were randomized into the t-PA treatment arm while 46% of the remaining 268 patients, mostly classified as having a large vessel stroke, were randomized to t-PA.

8.4.2 Analyses: The four subtype groups were examined analytically in three ways. First, the three variables necessary to uniquely indicate each patient's membership in one of the groups were, collectively, examined with all other potential covariates as part of the process of arriving at a "final" collection of covariates to be included in the treatment comparison models. In Table 5.10, it can be seen that these variables were quite significant in the first stage of this process when all covariate candidates were examined within models containing only the stratification variables. The corresponding 3-degrees-of-freedom Chi-squares and associated p-values for the 4 outcome measures were: Barthel (17.78, p = 0.0005), Rankin (15.16, p = 0.0017), Glasgow (10.74, p = 0.0132) and NIHSS (9.60, p = 0.0222). These analyses clearly indicate that stroke subtype is associated with the likelihood of a favorable outcome on all four scales. The nature of this association can be seen in the Table 8.4.1. For the Barthel, Rankin and Glasgow scales, the small vessel stroke patients had better than an even chance of a favorable outcome regardless of treatment, with odds of 1.9, 1.2, and 1.25 while for the other subtypes combined the odds were 0.72, 0.47 and 0.55 respectively. For the NIHSS scale the direction of the difference was the same but less dramatic.

However, as is seen in Table 5.10, there were other variables that were much more strongly related to the likelihood of a favorable outcome and, as the stepwise process of identifying the critical covariates continued, these variables entered the models, modifying the level of significance of the stroke subtype variables such that they never entered a single model. The logical conclusion to draw here is that the combined information contained in BsNIHSS, Age, PrDisability and PrDM, the variables that did enter the models, was highly correlated with the information that separated the small vessel stroke patients from the others and that the differences among the subtypes was no longer necessary in the models.

Nevertheless, we continued with the second and third stages of our examination of the subtype variables. We first included the three indicator variables, regardless of their p-values, in the outcome models that have been discussed so extensively in Sections 5.4 and 5.5. In these models, the 3-degrees-of-freedom Chi squares and associated p-values for the subtype variables are; Barthel (1.76. p = 0.62), Rankin (5.00, p = 0.17), Glasgow (5.06, p = 0.17). NIHSS (2.77, p = 0.43) and Global (3.09, p = 0.38). Since these variables are not statistically significant in these models, they will not be included. However lack of statistical significance does not prove that there are no differences among the stroke subtypes regarding a patient's likelihood of experiencing a favorable outcome. Even in the face of this lack of significance, it is interesting to note that, in these models with all of the influential covariates, the adjusted

estimates associated with the subtype variables now suggests that the cardioembolic group has the best chance of experiencing a favorable outcome, regardless of therapy.

Finally, we examined the question of whether the influence of t-PA as measured by the t-PA to Placebo odds ratio, is different across the subtype groups. In Table 8.4.1 we see, ignoring the "other" group because it is small (18 patients) and ill defined, the only consistent pattern is that the OR is smallest in the cardioembolic group for all outcome measures. In the 5 formal analyses, four outcome measures and the Global analysis, none of the interaction tests were significant (p – values between 0.40 and 0.58), indicating that the data do not provide statistically significant evidence suggesting a differential t-PA effect by stroke subtype. All previous caveats about insignificant tests of no interaction apply in this case as well.

8.4.3 Summary and Conclusions: We conclude that it was appropriate that the NINDS Investigators did not include stroke subtype as a covariate in the analytic models. Further, we conclude that the data of this trial do not support any claim regarding either the presence, or absence, of a differential t-PA treatment effect within stroke subtype.

8.5 Preexisting Disability

As illustrated in Table 8.5.1, 46 of the 622 patients randomized into this study were disabled prior to their stroke. The severity of the disability was assessed using the modified Rankin scale[1], and of the 46 patients with some preexisting disability 23 had slight disability (Rankin = 2), 17 had moderate disability (Rankin = 3) and 6 had severe disability (Rankin = 4). Very few of these 46 patients were classified as having a favorable outcome at 90 days. Indeed, for the Barthel, Rankin, Glasgow and NIHSS outcome measures, the number (%) of patients experiencing a favorable outcome was 6 (13%), 5 (11%), 5 (11%) and 3 (6.5%) respectively. None of the favorable outcomes occurred among patients with a severe disability and only one among those with a moderate disability (Table 8.5.1). In contrast, for the remaining 576 patients, the corresponding percents were; 48%, 37%, 40% and 29% respectively. This 3 to 4-fold difference in favorable response rate resulted in the inclusion of a variable indicating the 46 patients with a preexisting disability in the models assessing treatment response (Section 5.4.3).

In Section 5.5.2.1, we reported that this variable, while a highly significant predictor of an unfavorable outcome in all models, does not interact significantly with the treatment variable in any of the models. This lack of an interaction is seen in Table 8.5.1 where the ORs contrasting t-PA and Placebo are found to be very similar for those with and without a preexisting disability.

8.5.1 Summary and Conclusions: Thus, despite the fact that patients with a preexisting disability had a significantly reduced chance of experiencing a favorable outcome, there was no evidence that they responded any differently to t-PA therapy than those without a preexisting disability.

8.6 Diabetes Mellitus

8.6.1 Analyses: In our preliminary investigations we identified 8 patients with baseline blood glucose in excess of 400, in violation of the study protocol. Since elimination of this small number of patients would not substantially alter our conclusions, we continued to include them.

As seen in Table 8.6.1, of the 622 patients randomized into the study, 131 (21%) had a history of diabetes mellitus (DM). Of these 131, 34%, 30%, 31%, and 18% experienced a favorable outcome at 90 days according to the Barthel, Rankin, Glasgow, and NIHSS outcome measures respectively. The corresponding figures for those without DM are 48%, 36%, 40%, and 30% (see table entitled Diabetes Workbook). The differences between the corresponding favorability percents give a measure of the impact of DM on the likelihood of a favorable outcome. In Section 5.4.3 we reported that DM was the last of the "adjusting" covariates to enter the models but it had remained a significant predictor of an unfavorable outcome even after adjusting for the stratification variables as well as for the highly significant BsNIHSS, AGE and PrDisability.

In the Table 8.6.1, the ORs comparing t-PA therapy to Placebo are quoted for the DM and nonDM groups separately. It appears that there is little evidence of a t-PA advantage over Placebo among diabetics. However, in the Section 5.5.2.1, Tables 5.17 through 5.21, we reported the results of the tests of whether DM interacted with the treatment variable for the Barthel, Rankin, Glasgow, NIHSS and Global analyses, none of which was significant (p = 0.08, 0.25, 0.23, 0.96, and 0.27 respectively). In that section we also reported that these tests of "no interaction" would have 80% chance (power) of being significant only if the ORs among the nonDM patients were from 4 to 6-fold higher than the ORs among the DM patients. Since we observed (see Table 8.6.1) ratios of odds ratios between 1.3 (NIHSS) & 2.6 (Barthel), the fact that these tests of no interaction were all insignificant is no surprise. The caveat that we have stated before that the lack of evidence of difference does not constitute proof of the lack of a difference needs to be kept in mind here.

8.6.2 Summary and Conclusions: Although the observed data (Table 8.6.1) and the adjusted estimated t-PA effects, indicated a strong benefit for patients without DM, but no benefit among patients with DM, this comparison must be treated cautiously because there was no statistical evidence of a t-PA*DM interaction. The trial found no statistically significant evidence that diabetic and non-diabetic acute ischemic stroke patients responded differently to t-PA therapy.

9. CONCLUSION

The committee concluded that, despite an increased incidence of symptomatic intracerebral hemorrhage in t-PA treated patients and subgroup imbalances in baseline stroke severity, when t-PA was administered to acute ischemic stroke patients according to the study protocol, there was a statistically significant, and clinically important, benefit of t-PA treatment resulting in a higher likelihood of having a favorable clinical outcome at three months.

References

1. The National Institute of Neurological Disorders and Stroke rt-PA Stroke Study Group. Tissue plasminogen activator for acute ischemic stroke. *N Engl J Med* 1995;333:1581-1587.

2. Haley EC, Jr., Brott TG, Sheppard GL et al. Pilot randomized trial of tissue plasminogen activator in acute ischemic stroke. The TPA Bridging Study Group. *Stroke* 1993;24:1000-1004.

3. Haley EC, Jr., Levy DE, Brott TG et al. Urgent therapy for stroke. Part II. Pilot study of tissue plasminogen activator administered 91-180 minutes from onset. *Stroke* 1992;23:641-645.

4. Brott TG, Haley EC, Jr., Levy DE et al. Urgent therapy for stroke. Part I. Pilot study of tissue plasminogen activator administered within 90 minutes. *Stroke* 1992;23:632-640.

5. Lyden P, Brott T, Tilley B et al. Improved reliability of the NIH Stroke Scale using video training. NINDS TPA Stroke Study Group. *Stroke* 1994;25:2220-2226.

6. Lyden P, Broderick J, Mascha E. Reliability of the Barthel Index Outcome Measure selected for the NINDS t-PA Stroke Trial. In: Yamaguchi T, Mori E, Minematsuk K, del Zoppo G, editors. *Thrombolytic Therapy in Acute Ischemic Stroke III*. Tokyo, Japan: Springer-Verlag; 1995. p. 327-333.

7. Levy DE, Brott TG, Haley EC, Jr. et al. Factors related to intracranial hematoma formation in patients receiving tissue-type plasminogen activator for acute ischemic stroke. *Stroke* 1994;25:291-297.

8. The National Institute of Neurological Disorders and Stroke (NINDS) rt-PA Stroke Study Group. A systems approach to immediate evaluation and managment of hyperacute stroke. Experience at eight centers and implications for community practice and patient care. *Stroke* 1997;28:1530-1540.

9. The NINDS t-PA Stroke Study Group. Generalized efficacy of t-PA for acute stroke. Subgroup analysis of the NINDS t-PA stroke trial. *Stroke* 1997;28:2119-2125.

10. Kwiatkowski TG, Libman RB, Frankel M et al. Effects of tissue plasminogen activator for acute ischemic stroke at one year. *N Engl J Med* 1999;340:1781-1787.

11. The NINDS t-PA Stroke Study Group. Intracerebral hemorrhage after intravenous t-PA therapy for ischemic stroke. *Stroke* 1997;28:2109-2118.

12. Brott T, Lu M, Kothari R et al. Hypertension and its treatment in the NINDS rt-PA stroke trial. *Stroke* 1998;29:1504-1509.

13. Lenzer J. Alteplase for stroke: money and optimistic claims buttress the "brain attack" campaign. *BMJ* 2002;324:723-729.

14. Mann H, Li J, Nathanson LA et al. Alteplase for stroke. *BMJ* 2002;324:1581.

15. Mann J. Truths about the NINDS study: setting the record straight. *West J Med* 2002;176:192-194.

16. Trotter G. Why were the benefits of tPA exaggerated? *West J Med* 2002;176:192-194.

17. Wardlaw JM, Lindely RI, Lewis S. Thrombolysis for acute ischemic stroke: still a treatment for the few by the few. *West J Med* 2002;176:198-199.

18. Warlow C. Commentary: Who pays the guideline writers? *BMJ* 2002;324:726-727.

19. Marler JR, Tilley BC, Lu M et al. Early stroke treatment associated with better outcomes. The NINDS rt-PA Stroke Study. *Neurology* 2000;55:1649-1655.

20. Tilley BC, Marler J, Geller NG et al. Use of a global test for multiple outcomes in stroke trials with application to the National Institute of Neurological Disorders and stroke t-PA stroke trial. *Stroke* 1996;27:2136-2142.

21. Clinical Review for PLA 96-0350 - Endpoints and Planned Analyses. *Food and Drug Administration* 1996;16. Available at: http://www.fda.gov/cder/biologics/review/altegen061896r2.pdf. Accessed December 12, 2003.

22. Efron B, Gong G. A leisurely look at the bootstrap, the jackknife and cross-validation. *The American Statistician* 1982;37:36-48.

23. Tilley BC, Lyden PD, Brott TG, Lu M, Levine SR, Welch KM. Total quality improvement method for reduction of delays between emergency department admission and treatment of acute ischemic stroke. The National Institute of Neurological Disorders and Stroke rt-PA Stroke Study Group. *Arch Neurol* 1997;54:1466-1474.

24. Report of the Quality Standards Subcommittee of the American Academy of Neurology. Practice Advisory: Thrombolytic therapy for acute ischemic stroke - Summary Statement. *Neurology* 1996;47:835-839.

25. Gore JM, Granger CB, Simoons ML et al. Stroke After Thrombolysis : Mortality and Functional Outcomes in the GUSTO-I Trial. *Circulation* 1995 November 15;92:2811-2818.

26. Selker HP, Beshansky JR, Schmid CH et al. Presenting pulse pressure predicts thrombolytic therapy-related intracranial hemorrhage. Thrombolytic Predictive Instrument (TPI) Project results. *Circulation* 1994;90:1657-1661.

27. Simoons ML, Maggioni AP, Knatterud G et al. Individual risk assessment for intracranial haemorrhage during thrombolytic therapy. *Lancet* 1993;342:1523-1528.

28. Brookes ST, Whitely E, Egger M, Smith GD, Mulheran PA, Peters TJ. Subgroup analyses in randomized trials: risks of subgroup-specific analyses; power and sample size for the interaction test. *J Clin Epidemiol* 2004;57:229-236.

29. Clinical Review for PLA 96-0350 - Screened Patient Characteristics. *Food and Drug Administration* 1996;23-24. Available at:

http://www.fda.gov/cder/biologics/review/altegen061896r2.pdf. Accessed December 12, 2003.