1.0 Background

This is a plan for a Phase III trial to explore the efficacy of Argatroban and Eptifibatide in combination with rt-PA in treating stroke patients. One dose of each study drug will be compared to a control arm with respect to their ability to improve subjects’ 90 day scores on the modified Rankin scale (mRS). Improvement will be quantified using patient-centered utility scores for mRS values.

The trial is a Bayesian adaptive design that includes multiple key features:

1. adaptive sample size ranging from 500 to 1200 patients;
2. initial consideration of two active arms, with the ability to stop an arm for futility and continue as a 1:1 comparison between control and one of the active arms;
3. response-adaptive randomization is used to favor promising active arms;
4. frequent interim analyses can result in the trial stopping early for futility or for expected success;
5. a utility function on 90-day mRS scores to reflect patient and society valuation of outcome health states;
6. mRS scores are analyzed adjusting for baseline stroke severity as measured by the NIH Stroke Scale. The relationship between stroke severity and outcomes is modeled flexibly;
7. a longitudinal model relating 30-day mRS scores to 90-day mRS scores is utilized to bring more information from patients without 90-day data to stopping and adaptive allocation decisions.

The structure of this document is as follows: the next section describes the design in general terms, Section 3 elaborates on the subject population and the final analysis, and Section 4 fills in further details about decisions made during the trial. Section 5 presents tentative operating characteristics for the design obtained using simulation and based on a variety of assumed truths about the effectiveness of the drugs. Appendices elaborate on the statistical models used in the final analysis and in the interim analyses and on default assumptions about the subject population.
2.0 Design Overview

For the first 150 subjects, the randomization probabilities for the three arms remain fixed at 1:1:1 for the two active arms and the control arm. An interim analysis occurs at the 150th enrolled subject, and randomization probabilities for the active arms are adjusted, with the arm having the higher predictive probability of a successful final analysis assigned higher allocation probabilities. Interim analyses continue every 4 weeks.

Beginning with the first interim analysis after the 150th subject is enrolled, the design implements response adaptive randomization between two active arms (based on predictive probability) where control gets the same randomization probability as the maximum active arm.

After 500 subjects have been enrolled, the stopping rule for futility is based on 20% predictive probability and there is a shift to equal randomization between control and one or two active arms. From this point until the end of the trial, all arms remaining in the trial have equal allocation probabilities.

Interim analyses occur after 700 and 900 patients have been enrolled to allow for

- stopping one active arm for futility (based on 5% predictive probability based on three arms continuing)
- stopping the trial altogether for futility (based on 5% predictive probability based on two arms continuing)
- stopping the trial for expected success of one or both active arms.

The trial enrolls at most 1200 subjects. When final data for all enrolled subjects are available, the final analysis is conducted, and it may result in showing a significant benefit for one or two active arms.

Stopping decisions are based on Bayesian predictive probabilities, and in particular the predictive probability of a successful final analysis. Details about these predictive probabilities will be given in Section 4.

3.0 Study Population, Primary Endpoint, and Statistical Test

3.1 Entry criteria

The trial enrolls subjects with initial NIHSS scores of 6 or larger.

3.2 Treatment arms

Three treatment arms are under consideration:
1. Control arm with iv-tPA only;
2. Argatroban in addition to iv-tPA;
3. Eptifibatide in addition to iv-tPA;

Of the first 150 enrolled subjects, one-third will be assigned to the control arm and to each of the active arms.

3.3 Primary Endpoint

The primary endpoint for this trial is the 90-day mRS score. We choose to analyze this standard endpoint by converting the mRS scores into weights that directly reflect patient and society valuation of outcome health states. We then model a subject’s weighted mRS score as normally distributed with expected value depending on initial NIHSS and treatment assigned.

The weights assigned to the possible mRS scores are shown in Table 1 below. These weights were obtained through a synthesis of studies (O. Rivero-Arias, et al, “Mapping the Modified Rankin Scale (mRS) Measurement into the Generic EuroQol (EQ-5D) Health Outcome,” Medical Decision Making 2010 30:341, and K.-S. Hong and J.L. Saver, "Quantifying the Value of Stroke Disability Outcomes: WHO Global Burden of Disease Project Disability Weights for Each Level of the Modified Rankin Scale: Supplemental Mathematical Appendix,” Stroke 2009 40:3828-3833). Both these studies assigned utility values and confidence intervals to mRS scores; these are also shown in Table 1. We renormalized these utilities to a scale where an mRS of 6 implies a utility of 0 and and mRS of 0 implies a utility of 10. The two scales are quite similar, and we take the mean of the renormalized utilities to obtain our own weights. The second study reported more precise estimates, so in some cases the consensus value is closer to its value.

<table>
<thead>
<tr>
<th>mRS</th>
<th>0</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
</tr>
</thead>
<tbody>
<tr>
<td>Rivero-Arias et al</td>
<td>10</td>
<td>8.7</td>
<td>7.3</td>
<td>6.0</td>
<td>2.8</td>
<td>-0.1</td>
<td>0</td>
</tr>
<tr>
<td>Hong &amp; Saver</td>
<td>10</td>
<td>9.5</td>
<td>7.9</td>
<td>6.7</td>
<td>3.5</td>
<td>0.1</td>
<td>0</td>
</tr>
<tr>
<td>This Trial</td>
<td>10</td>
<td>9.1</td>
<td>7.6</td>
<td>6.5</td>
<td>3.3</td>
<td>0</td>
<td>0</td>
</tr>
</tbody>
</table>

**Table 1:** weights used for 90 day mRS scores.

Relative to an approach that dichotomizes the 7 possible mRS scores into two possibilities, weighting the 7 Rankin levels by utilities improves the precision of the scale as a measure of disability. The weighted approach should also not be confused
with an approach based on the raw mRS scores, which would erroneously treat each single-point increase in mRS as equally valuable to the subject.

Figure 1 below gives an example of what treatment effects look like for this endpoint. Each vertical bar depicts a probability distribution: the height of the darkest blue represents the probability of an mRS of 0, the darkest red shows the probability of an mRS of 6, etc. The leftmost bar shows the results of the control arm for the NINDS tPA study (Tissue plasminogen activator for acute ischemic stroke. The National Institute of Neurological Disorders and Stroke rt-PA Stroke Study Group. The New England journal of medicine 1995;333:1581-7), which represents an expected utility of 5.01, and the second bar shows the tPA arm in the NINDS study, which represents an expected utility of 5.91. The remaining bars show distributions that represent improvements of 0.1 to 0.9 as compared to the tPA arm.

Figure 1: graphical depiction of mRS distributions. The first two bars depict the two arms in the NINDS study, and the remaining nine bars depict further improvements to expected utility. Dark blue represents the probability of an mRS of zero, dark red represents the probability of an mRS of 6, and so on.
### Table 2: translating treatment effects on the utility scale into changes in probabilities of the mRS outcomes. The point estimate of the effect of tPA in the NINDS study is 0.9 units of utility. The table shows what further improvements to expected utility above tPA might do.

<table>
<thead>
<tr>
<th>Treatment</th>
<th>Expected Utility</th>
<th>mRS=0</th>
<th>mRS=1</th>
<th>mRS=2</th>
<th>mRS=3</th>
<th>mRS=4</th>
<th>mRS=5</th>
<th>mRS=6</th>
<th>0-1</th>
</tr>
</thead>
<tbody>
<tr>
<td>NINDS Control</td>
<td>5.01</td>
<td>0.11</td>
<td>0.16</td>
<td>0.12</td>
<td>0.14</td>
<td>0.20</td>
<td>0.07</td>
<td>0.21</td>
<td>0.27</td>
</tr>
<tr>
<td>NINDS tPA arm</td>
<td>5.91</td>
<td>0.18</td>
<td>0.24</td>
<td>0.08</td>
<td>0.13</td>
<td>0.13</td>
<td>0.06</td>
<td>0.17</td>
<td>0.42</td>
</tr>
<tr>
<td>tPA + 0.1</td>
<td>6.01</td>
<td>0.19</td>
<td>0.25</td>
<td>0.08</td>
<td>0.13</td>
<td>0.13</td>
<td>0.06</td>
<td>0.17</td>
<td>0.44</td>
</tr>
<tr>
<td>tPA + 0.2</td>
<td>6.11</td>
<td>0.20</td>
<td>0.25</td>
<td>0.08</td>
<td>0.13</td>
<td>0.13</td>
<td>0.06</td>
<td>0.16</td>
<td>0.45</td>
</tr>
<tr>
<td>tPA + 0.3</td>
<td>6.21</td>
<td>0.21</td>
<td>0.25</td>
<td>0.08</td>
<td>0.13</td>
<td>0.13</td>
<td>0.06</td>
<td>0.15</td>
<td>0.46</td>
</tr>
<tr>
<td>tPA + 0.4</td>
<td>6.31</td>
<td>0.22</td>
<td>0.25</td>
<td>0.08</td>
<td>0.12</td>
<td>0.13</td>
<td>0.05</td>
<td>0.14</td>
<td>0.47</td>
</tr>
<tr>
<td>tPA + 0.5</td>
<td>6.41</td>
<td>0.22</td>
<td>0.26</td>
<td>0.08</td>
<td>0.12</td>
<td>0.13</td>
<td>0.05</td>
<td>0.14</td>
<td>0.48</td>
</tr>
<tr>
<td>tPA + 0.6</td>
<td>6.51</td>
<td>0.23</td>
<td>0.26</td>
<td>0.08</td>
<td>0.12</td>
<td>0.12</td>
<td>0.05</td>
<td>0.13</td>
<td>0.49</td>
</tr>
<tr>
<td>tPA + 0.7</td>
<td>6.61</td>
<td>0.24</td>
<td>0.27</td>
<td>0.08</td>
<td>0.12</td>
<td>0.12</td>
<td>0.05</td>
<td>0.13</td>
<td>0.51</td>
</tr>
<tr>
<td>tPA + 0.8</td>
<td>6.71</td>
<td>0.25</td>
<td>0.27</td>
<td>0.08</td>
<td>0.12</td>
<td>0.12</td>
<td>0.05</td>
<td>0.12</td>
<td>0.52</td>
</tr>
<tr>
<td>tPA + 0.9</td>
<td>6.81</td>
<td>0.26</td>
<td>0.27</td>
<td>0.08</td>
<td>0.12</td>
<td>0.11</td>
<td>0.05</td>
<td>0.11</td>
<td>0.53</td>
</tr>
</tbody>
</table>

**3.4 Primary Analysis**

The final analysis is Bayesian and includes a flexible normal dynamic linear model (NDLM) to account for different expected outcomes as a function of initial NIHSS. This is a flexible spline-like model that will capture that the average weighted mRS score in the control group is a (possibly non-linear) function of initial NIHSS. Meanwhile the average effect of a given treatment \(d\), \(\theta_d\), is the difference in the expected utility for the active treatment minus control. This average treatment effect, \(\theta_d\), is assumed to be equal over all values of initial NIHSS. Details of the statistical model are given in Appendix A. For each active arm, the hypothesis test is

\[
H_0: \theta_d \leq 0 \\
H_A: \theta_d > 0
\]
The treatment effect \( \theta_d \) is given a vague prior, \( \theta \sim N(0, 2.5^2) \): in particular, the prior probability that a drug is beneficial is the same as the prior probability that it is harmful. If there is a high posterior probability that the treatment effect \( \theta_d \) is positive, the treatment is declared to be efficacious. The posterior probability is conditional on the final results for all subjects enrolled in the trial. If two active arms remained in the trial all the way to the end, this posterior probability is computed for each, and potentially both arms can be declared to be successful.

### 3.5 Thresholds for a Successful Trial

The trial is successful if in the final analysis, the posterior probability of a positive benefit is at least 0.985. That is, success is claimed if

\[
\Pr(\theta_d > 0) \geq 0.985
\]

This threshold is chosen to control the experiment-wise type I error probability at 0.025.

### 3.6 Longitudinal Modeling of 30-day mRS Scores

Subjects with 30-day data but no 90-day data are also included in the interim analyses, and their data contribute to the adaptive allocation probabilities and the stopping and arm selection decisions. A subject with only 30-day data is less influential than a subject with complete data, because the statistical model in effect takes into account the fact that different 90-day outcomes are still possible. The relationship between 30-day and 90-day data is initially assumed to be unknown, and is updated as more data from the trial come in. Consequently, the trial will not make irreversible decisions as a result of incorrect beliefs about the relationship.
4.0 Prospectively Planned Interim Analyses

The first interim analysis takes place after 150 subjects have been enrolled. Subsequent interim analyses take place every four weeks (28 days) until 500 patients are enrolled, and less frequently thereafter.

4.1 Predictive Probabilities

Decisions made as a result of interim analyses are based on Bayesian predictive probabilities using the statistical model defined in Appendix B. The predictive probability of a successful final analysis is calculated based on different assumptions about the remaining subjects to be enrolled.

First, for each active arm, we assume that the remainder of the trial consists of 1:1 randomization between that arm and the control arm, and calculate the predictive probability that the remaining patients generate data that lead to a significant result (high posterior probability of a positive treatment effect). These predictive probabilities are used in futility decisions and response adaptive randomization probabilities. For convenience we will refer to these predictive probabilities as 1:1 predictive probabilities.

Second, we assume that the remaining subjects to be accrued are allocated 1:1:1 to the two active arms and to the control arm, and calculate the predictive probability that those subjects generate data resulting in a significant final analysis for the Argatroban arm, and respectively for the Eptifibatide arm. We will refer to these predictive probabilities as 1:1:1 predictive probabilities.

Third, we also calculate the predictive probability that an arm would have a successful final analysis if enrollment were stopped immediately, based on predictions of results for subjects enrolled in the trial but without final data. This predictive probability is used to determine whether enrollment should be stopped early for expected success. We will refer to these predictive probabilities as expected success predictive probabilities.

4.2 Interim Monitoring for Early Futility and Expected Success.

The trial cannot stop for futility or drop an arm before 500 subjects. Interim analyses before 500 subjects update response adaptive randomization probabilities.

After 500 subjects have been enrolled, one of the arms can be dropped if its 1:1:1 predictive probability is less than 20% or the trial can be stopped altogether if both active arms have 1:1 predictive probability is less than 20%.
After 700 and 900 subjects have been enrolled, one of the arms can be dropped if its 1:1:1 predictive probability is less than 5% or the trial can be stopped altogether if the 1:1 predictive probability is less than 5%.

When 700 patients and 900 patients have been enrolled the trial may to stop for expected success. Beginning with the first analysis after this time, the expected success predictive probabilities are calculated. If an active arm has an expected success predictive probability of at least 99%, that arm stops for expected success, and its final analysis is conducted when all subjects enrolled are followed up for 90-day data. If it is the last remaining active arm, the trial stops altogether, and all enrolled subjects are followed up to determine the success of the trial.

4.4 Response-Adaptive Randomization

During the response-adaptive randomization regime, which begins at 150 subjects and ends when the 500th subject has been enrolled, the control arm retains a 1/3 allocation probability throughout. The allocation probabilities for the two active arms are set to be proportional to their 1:1 predictive probabilities. For example, suppose that the two active arms have 1:1 predictive probabilities of 0.2 and 0.8 respectively. The 2/3 total allocation probability assigned to active arms is divided up proportionally to the predictive probabilities, and in this case the two arms have allocation probabilities of (2/15, 8/15) respectively.

5.0 Operating Characteristics

In this section we present tentative operating characteristics. These results are obtained through simulation as illustrated in Section 5, and are based on 1000 simulated trials per scenario, except for the null scenario whose results are based on 10000 simulated trials.

In all simulations, we assumed that 50% of patients receive endovascular therapy (ET) in addition to the study drug or standard tPA. The treatment effect for ET patients (projected half of participants in trial) is 75% of the effect in non-ET patients. The impact of ET on treatment effect is unknown but we assumed the worst scenario, i.e., attenuation of the treatment effect. Patients with ET are assumed to have the same initial NIHSS distribution as those without.

5.1 Operating characteristics when all active arms are equivalent

First we present preliminary estimates of power and sample size characteristics in scenarios in which both active arms have the same effect: specifically, an increase of
zero (a null scenario), 0.3, 0.4, and 0.5 units for the non-ET subjects. (The assumed overall treatment effects were 0.2625, 0.35, and 0.4375.) These effects benefit all injury severities equally with respect to expected utility. Examples of the scenarios are pictured in Figure 2 below, which shows the assumed distribution of mRS as a function of initial NIHSS (shown on the x-axis), for the control arm (left plot), for an arm with a 0.5 unit effect (center plot), and for an arm with a 1.0 unit (right plot; this effect is unrealistically large). For example, examining the heights of the dark blue bars on the far left of the plots show that the probability of an mRS of zero for subjects with an initial NIHSS of 6 treated with the control arm is assumed to be almost 60%. An expected utility effect of 0.5 points raises this to about 70%, while an effect of 1.0 points raises it to about 85%.

Figure 2: example distributions of mRS as a function of initial NIHSS.

Since the trial involves two different active drugs and potentially both of them can be successful, power has multiple aspects:
- the probability that at least one active arm is successful;
- the probability that the Argatroban arm is successful;
- the probability that the Eptifibatide arm is successful (this should be the same as the probability that an Argatroban arm is successful for the scenarios examined in this section);
- the probability that both active arms win.

Results based on 1000 simulated trials for each scenario, except with 10000 simulated trials for the null scenario, are shown in Table 3. Simulations and analysis have not yet been extensive enough to claim control of Type I error, but the numbers in the Null column are consistent with 0.025 Type I error. If the overall true effect is 0.4375 units for all active arms, the design has high power of 97% to be successful for some arm, and each drug has a 12% chance of being unsuccessful.
Table 3: power-related operating characteristics for scenarios in which both active arms are equally effective.
ET=endovascular therapy

Further operating characteristics are shown in Table 4 below. The table displays the probability of a futility stop at exactly 500 subjects, and the probability that the design reaches full enrollment of 1200. The average and standard deviation of total sample sizes are also shown, as well as the averages for trials that were ultimately unsuccessful.

Table 4: sample size-related operating characteristics.
5.2 Operating characteristics when arms differ

In this section we present operating characteristics for scenarios in which the Argatroban arm has a benefit of up to 0.5 units, while the Eptifibatide arm has no benefit or a smaller benefit, to explore how successful the design is at identifying the most promising arm. We have simulated scenarios in which Eptifibatide is less effective than Argatroban and not the reverse, but the two drugs are treated exchangeably, so those simulation results also apply to the analogous cases where Eptifibatide is the superior drug.

We see from Table 5 that if one of the active arms has an effect of 0.4 units of utility, the design has approximately 83% power. If one drug is ineffective, the probability that the more effective drug will be successful is increased relative to Table 3 because the ineffective drug will likely be dropped. The probability that an ineffective drug is successful is small.

<table>
<thead>
<tr>
<th>Assumed True Effect Overall (ET and non-ET patients)</th>
<th>0.2625</th>
<th>0.35</th>
<th>0.4375</th>
</tr>
</thead>
<tbody>
<tr>
<td>Assumed True Effect For non-ET patients</td>
<td>0.30</td>
<td>0.40</td>
<td>0.50</td>
</tr>
<tr>
<td>Assumed True Effect for ET patients (75% of effect w/o ET)</td>
<td>0.225</td>
<td>0.30</td>
<td>0.375</td>
</tr>
</tbody>
</table>

| Pr{Some Arm Wins}                                   | 0.562  | 0.834 | 0.945 |
| Pr{Arm A wins}                                      | 0.561  | 0.833 | 0.945 |
| Pr{Arm E wins}                                      | 0.015  | 0.012 | 0.009 |
| Pr{ A Arm and E arm both win}                       | 0.014  | 0.011 | 0.009 |

Table 5: power characteristics for scenarios where the active arms differ. ET=endovascular therapy

**Appendix A: Statistical Model for Final Analyses**

Denote the $j$'th subject’s 90-day mRS by $S_j$, and her resulting weight score by $Y_j$; $Y_j = W_k$ if $S_j = k$. Further denote the $j$'th subject’s initial NIHSS by $I_j$ and the treatment to which she was randomized by $d_j$ ($d_j$ is either 0 for control or 1, 2 for the active arms). Define, for $6 \leq i \leq 42$ and $d \in \{0,1,2\}$,

$$\Pr\{Y_j = k \mid I_j = i, d_j = d\} = p_k(i, d).$$
For the purposes of the final analysis, the \( p_k(i, d) \) are used to define the expected values of the weight scores. The expected values of the weight scores, \( Y_j \), are modeled as Gaussian, with expected values that depend on \( i \), with a common treatment effect \( \theta_d \) with \( \theta_0 = 0 \) by convention, and with variances \( \sigma_d^2 \) that depend on the treatment:

\[
E(Y_j | I_j = i, d_j = d) = \sum_{k=0}^{6} p_k(i, d) W_k = \phi_i + \theta_d,
\]

for all initial NIHSS scores \( i \). We model the \( \phi_i \) flexibly, and assume that they come from a second order normal dynamic linear model (NDLM). Specifically, the prior distribution for the \( \phi_i \) assumes that for \( 8 \leq i \leq 42 \), we have

\[
\phi_i \sim \text{Normal}(2\phi_{i-1} - \phi_{i-2}, \tau^2).
\]

This form of the normal dynamic linear model encourages the \( \phi_i \) to be linear.

We use the following normal dynamic linear model distributions:

\[
\theta_d \sim \text{Normal}(0, 2.5^2) \quad (d = 1, 2)
\]

\[
\sigma_d^2 \sim \text{Inverse Gamma} (1, 10) \quad (d = 0, 1, 2), \text{ and}
\]

\[
\tau^2 \sim \text{Inverse Gamma} (10, 0.005).
\]

The final analysis is performed with either one or two active arms. We evaluate the posterior distribution of the parameters of this model using the Gibbs sampler. Conditionally on the other parameters, \( (\phi, \theta) \) have a multivariate normal distribution, and the remaining parameters have inverse Gamma conditional distributions.

The primary output of the final analysis is the posterior probability that \( \theta_d > 0 \), for any \( d \)'s that remain in the trial. If this probability is at least 0.985, the trial is considered to be a success. The threshold for defining significance is chosen so that the design has Type I error no larger than 0.025. Criteria for success (critical value for posterior probability of a positive benefit) was inflated from 0.975 to 0.985 to account for the two study drugs and the repeated interim looks (e.g. the trial can be stopped when data look favorable enough that a success is likely).

**Appendix B: Statistical Model for Interim Analyses**

The statistical model used during the trial to compute predictive probabilities to determine allocation probabilities and make the decision to stop for futility or expected success, is more detailed than the final analysis model. We use a longitudinal model to impute values of final endpoints for subjects for whom we have 30-day mRS scores but not 90-day scores; we estimate the probability distribution of final endpoint values given early endpoint values. Another major
change is that we also estimate the distribution of initial NIHSS scores for enrolled subjects; for predicting whether the trial will be successful it is critical to be able to forecast what kinds of subjects will appear in the future.

Whereas in the final analysis we use a noninformative prior with no information about the overall level of the \( \phi_i \) or the overall slopes of the \( \phi_i \) as a function of \( i \), we now use the following prior distributions:

\[
\phi_6 \sim N(5, 2.5^2), \ \text{and} \ \phi_7 \sim N(\phi_6, 0.25^2).
\]

Writing \( Y_j^{30} \) for the 30-day mRS value for the jth subject, we estimate the probabilities \( \lambda_{mk} = \Pr(Y_j = k \mid Y_j^{30} = m) \) using a multinomial model with prior distributions

\[
(\lambda_{m0}, \lambda_{m1}, \lambda_{m2}, \lambda_{m3}, \lambda_{m4}, \lambda_{m5}, \lambda_{m6}) \sim \text{Dirichlet}(\frac{1}{3}, \frac{1}{3}, \frac{1}{3}, \frac{1}{3}, \frac{1}{3}, \frac{1}{3}, \frac{1}{3})
\]

for \( m = 0,1,2,3,4,5 \). We have \( \lambda_{66} = 1 \) and \( \lambda_{6k} = 0 \) for \( k < 6 \). As mentioned earlier, the longitudinal model plays no role in the final analysis. The parameters of the longitudinal model are updated at each interim analysis and are based on all subjects with complete 30-day and 90-day data to that point. We note that we are not using data from other studies to inform the parameters of the longitudinal model. We use the same longitudinal model for all arms (i.e. we pool the data for all subjects to estimate the probability distribution of 90-day outcome given 30-day outcome). The final piece of the statistical model for interim analyses is the model for the initial NIHSS distribution \( \Pr(I_j = i) = u_i \), which is also a Dirichlet-multinomial model with prior distribution

\[
(u_6, u_7, ..., u_{42}) \sim \text{Dirichlet}(\frac{1}{3}, \frac{1}{3}, ..., \frac{1}{3}).
\]

The prior distributions for the \( \sigma_d^2 \) are as specified in the description of the final analysis.

During an interim analysis, we estimate the parameters \( (\phi, \theta, \sigma^2, \tau^2, \lambda, \iota) \) of this model using Gibbs sampling. We then use these samples to estimate several predictive quantities. First, for each active arm in the trial, we calculate the probability that the trial would end with a significant result if we assigned all remaining subjects 1:1 to control and that active arm, and enrolled subjects up to the maximum sample size. This calculation consists of the following steps: for a given Markov chain Monte Carlo (MCMC) sample,

1. Using the \( \lambda_s \), impute 90-day endpoint values for the subjects enrolled and with 30-day data.
2. Simulate random initial NIHSS scores and treatment assignments for the subjects yet to be enrolled, using the \( \iota \)'s and assuming that all remaining
subjects are enrolled 1:1 to control and that active arm. Augment this list of subjects with the subjects included in the trial who have not yet provided 30 day data.

3. Calculate the probability, given that list of subjects, that final 90 day data will result in a significant trial.

One may choose to repeat step 2 multiple times for a given MCMC sample. Compute the average of the resulting probabilities. These probabilities will be used to decide whether to to make arm selection decisions and futility stopping decisions.

Second, we calculate the probability of a successful final analysis if the trial assigned all future subjects to control and the two active arms equally 1:1:1. This probability is used to decide whether to drop one active arm.

Finally, we calculate the probability of a successful final analysis if the trial were to stop enrollment at once and then wait for final data for all enrolled patients. This probability is based on predicting the final data for enrolled patients with no data yet, as well as those with 30-day data only.

Appendix C: Default Population Assumptions

In this appendix we present the assumptions about the population that were used in the simulations. First we present the assumed distribution of NIHSS scores for enrolled subjects. Next we present the assumed distribution of mRS as a function of NIHSS for control subjects. Finally we present the assumed distribution of 90-day mRS given 30-day mRS scores.

Appendix C1: Initial NIHSS Distribution

In Figure C1 we display the assumed distribution of initial NIHSS scores. The distribution is based on an exponential distribution with mean 10 but with values less than 6 or more than 42 omitted. This assumption was chosen to be roughly consistent with Reeves et al (Distribution of National Institutes of Health Stroke Scale in the Cincinnati/Northern Kentucky Stroke Study; Mathew Reeves, et al; Stroke. 2013; 44:3211-3213).
Figure C1: the assumed distribution of initial NIHSS scores for enrolled subjects. Lower scores are considerably more likely.

Appendix C2: Distribution of 90-Day mRS Given Initial NIHSS

Figure C2 below displays the distributions of 90-day mRS given initial NIHSS assumed for the control arm. Good outcomes are expected to be very likely for NIHSS of 6 and very infrequent for initial NIHSS larger than 26.

Assumed Distribution of mRS as a function of NIHSS

Figure C2: distributions of 90-day mRS values conditional on initial NIHSS.
### Appendix C3: Distribution of 90-Day mRS Given 30-Day mRS

Table C1 below displays the assumed conditional distributions of 90-day mRS scores given 30-day mRS scores. These were taken from Ovbiagele, Lyden, and Saver (2010; Ovbiagele B, Lyden PD, Saver JL, Disability status at 1 month is a reliable proxy for final ischemic stroke outcome. Neurology 2010;75:688-92).

<table>
<thead>
<tr>
<th>30 \ 90</th>
<th>0</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>0.78</td>
<td>0.18</td>
<td>0.02</td>
<td>0.003</td>
<td>0.003</td>
<td>0</td>
<td>0.009</td>
</tr>
<tr>
<td>1</td>
<td>0.24</td>
<td>0.65</td>
<td>0.075</td>
<td>0.02</td>
<td>0.01</td>
<td>0.001</td>
<td>0.01</td>
</tr>
<tr>
<td>2</td>
<td>0.06</td>
<td>0.38</td>
<td>0.45</td>
<td>0.09</td>
<td>0.02</td>
<td>0</td>
<td>0.009</td>
</tr>
<tr>
<td>3</td>
<td>0.02</td>
<td>0.12</td>
<td>0.34</td>
<td>0.44</td>
<td>0.06</td>
<td>0.01</td>
<td>0.01</td>
</tr>
<tr>
<td>4</td>
<td>0.003</td>
<td>0.02</td>
<td>0.06</td>
<td>0.29</td>
<td>0.53</td>
<td>0.05</td>
<td>0.04</td>
</tr>
<tr>
<td>5</td>
<td>0</td>
<td>0.001</td>
<td>0.001</td>
<td>0.007</td>
<td>0.06</td>
<td>0.87</td>
<td>0.06</td>
</tr>
<tr>
<td>6</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
</tr>
</tbody>
</table>

Table C1: conditional distribution of 90-day mRS given 30-day mRS.