Challenges of Decision Making Regarding Futility in a Randomized Trial
The Interventional Management of Stroke III Experience

Sharon D. Yeatts, PhD; Renee H. Martin, PhD; Christopher S. Coffey, PhD; Patrick D. Lyden, MD; Lydia D. Foster, MS; Robert F. Woolson, PhD; Joseph P. Broderick, MD; Marco R. Di Tullio, MD; Charles A. Jungreis, MD; Yuko Y. Palesch, PhD; for the IMS III Investigators

Background and Purpose—Interventional Management of Stroke (IMS) III is a randomized, parallel arm trial comparing the approach of intravenous tissue-type plasminogen activator followed by endovascular treatment with intravenous tissue-type plasminogen activator alone in patients with acute ischemic stroke presenting <3 hours of symptom onset. The trial intended to enroll 900 subjects to ensure adequate statistical power to detect an absolute 10% difference in the percentage of subjects with good outcome, defined as modified Rankin Scale score of 0 to 2 at 3 months. In April 2012, after 656 subjects were randomized, further enrollment was terminated by the National Institute of Neurological Disorders and Stroke based on the prespecified criterion for futility using conditional power <20%.

Methods—Conditional power was defined as the likelihood of finding statistical significance at the end of the study, given the accumulated data to date and with the assumption that a minimum hypothesized difference of 10% truly exists between the 2 groups. The evolution of study data leading to futility determination is described, including the interaction between the unblinded study statisticians and the Data and Safety Monitoring Board in the complex deliberation of analysis results.

Results—The futility boundary was crossed at the trial’s fourth interim analysis. At this point, based on the conditional power criteria, the Data and Safety Monitoring Board recommended termination of the trial.

Conclusions—Even in spite of prespecified interim analysis boundaries, interim looks at data pose challenges in interpretation and decision making, underscoring the importance of objective stopping criteria.

Clinical Trial Registration—URL: http://www.clinicaltrials.gov. Unique identifier: NCT00359424.
(Stroke. 2014;45:1408-1414.)

Key Words: clinical trial ■ endovascular techniques

Intravenous tissue-type plasminogen activator (IV tPA; Alteplase) is the only US Food and Drug Administration–approved medical treatment for acute ischemic stroke. The use of endovascular approaches, including intra-arterial delivery of tPA as well as the use of thrombectomy and stent retriever devices, has become increasingly more common. Although such devices have been cleared by the Food and Drug Administration for thrombus removal, there is no evidence from a randomized controlled trial that demonstrates whether the endovascular approach is more effective than IV tPA alone. The Interventional Management of Stroke (IMS) III trial was designed as a randomized, parallel arm trial comparing the approach of IV tPA followed by endovascular treatment with IV tPA alone.

The design of IMS III incorporated observations from the IMS I trial, which, at the time of grant submission, was the only available trial data regarding clinical outcome for endovascular therapy after IV tPA. Subjects with the highest National Institutes of Health Stroke Scale (NIHSS) scores were most likely to exhibit major arterial occlusion after IV tPA and were more likely to benefit from endovascular therapy. The risk of symptomatic intracerebral hemorrhage was also greatest in subjects with NIHSS >20 in the National Institute of Neurological Disorders and Stroke (NINDS) tPA Stroke Trial. For these reasons, the IMS III trial randomized patients by severity strata defined according to baseline NIHSS score, and the primary efficacy analysis called for adjustment by NIHSS strata.
IMS III was designed to include interim efficacy and futility analyses as the trial progressed. In April 2012, after 656 subjects had been randomized, the fourth interim analysis led the Data and Safety Monitoring Board (DSMB) to recommend to the sponsor (NINDS) that the trial be halted because the futility boundary had been crossed.

The IMS III investigators have previously reported the primary results of the trial. Briefly, there was insufficient evidence to conclude that endovascular therapy increased the proportion of subjects with a favorable outcome as compared with IV tPA alone (absolute adjusted risk difference, 1.5%; 95% confidence interval [CI] −6.1 to 9.1). Predefined subgroup analyses also failed to demonstrate an effect of endovascular therapy among subjects with an NIHSS score of ≥20 (6.8%; 95% CI, 4.4 to 18.1) or ≤19 (−1.0; 95% CI, −10.8 to 8.8).

During the deliberations and data review leading up to the final interim analysis, the unblinded statisticians and the DSMB struggled with a variety of issues that might be seen in future trials of device therapies for complex disorders such as stroke. The purpose of this article is to describe the accumulation of data and the complexities associated with futility determination in the trial. The authors include members of both the statistical team and the DSMB.

Methods

IMS III Study Design
A full description of the trial methodology can be found in Khatri et al. Subjects were randomly assigned in a ratio of 2:1 to either endovascular therapy or IV tPA alone; as detailed in Khatri et al., the randomization was balanced with respect to clinical center and baseline stroke severity. For the purposes of randomization and analysis, stroke severity was stratified according to baseline NIHSS, with moderate strokes characterized by NIHSS ≤19 and severe strokes NIHSS ≥20. The primary efficacy outcome of functional independence was defined as a modified Rankin Scale score of 0 to 2 at 90 days. Thrombectomy treatment itself could not be blinded, so the protocol specified that blinded investigators conduct all primary outcome assessments.

The intended sample size of 900 subjects was calculated for an absolute treatment difference of 10% in the proportion of subjects with favorable outcome at 90 days, assuming a proportion of 40% in IV tPA alone (πIV=0.4; πendovascular=0.5), 2-sided type I error probability of 0.05, 80% power, 2:1 randomization, and an inflation factor of 1.03 to account for 2% noncompliance. These calculations took into account 3 planned interim efficacy analyses conducted according to O’Brien and Fleming type alpha spending function, intended to be conducted when 225, 450, and 675 subjects in moderate and severe strata, respectively, had completed the primary 90-day follow-up. Interim analyses for futility were intended to coincide with interim efficacy analyses.

The primary efficacy analysis was conducted via Cochran–Mantel–Haenszel test, adjusting for the baseline severity strata defined above. At each analysis, the Breslow–Day test for homogeneity of odds ratios was evaluated to assess whether treatment-by-severity interaction may exist; however, it was recognized that this and other tests have low power to detect interaction. An unfavorable outcome (modified Rankin Scale score >2) was imputed for all subjects with the primary outcome missing or obtained outside of the specified window.

The prespecified futility boundary was based on conditional power, defined as the probability of rejecting the null hypothesis (H0: πendovascular = πIV, where πIV is assumed to be 0.4) at the final analysis, given the data accumulated to date. Low conditional power meant that the trial was unlikely to demonstrate a difference between the treatment arms. The prespecified boundary for futility consideration was a conditional power <20% under the alternative hypothesis (H1: πendovascular ≠ πIV, where πIV is assumed to be 0.4 and πendovascular = 0.5).

The principal investigator (Y.Y.P.) of the Statistics and Data Management Center was the blinded statistician and a voting member of the trial Executive Committee. The unblinded statisticians (S.D.Y., R.H.M., L.D.F., R.F.W.) were members of the Statistics and Data Management Center, trial coinvestigators, nonvoting participants in trial management discussions, and liaisons to the DSMB. In order to maintain the blind, the blinded statistician took part in Executive Committee and DSMB teleconferences in a location separate from the unblinded team and was granted access only to aggregate trial data.

The unblinded statistical team was responsible for the generation of safety reports, which were provided to the DSMB on a semiannual basis, as well as interim analysis reports. Prespecified safety variables, including death <90 days as well as symptomatic and asymptomatic intracranial hemorrhage, were monitored via calculation of the relative risk and corresponding 95% confidence interval. Formal comparison of treatment arms with regard to safety was conducted after a prespecified number of events had been reported.

Results

Interim Findings
At the time of each interim analysis, the unblinded statistical team also provided a stratified contingency table demonstrating the relationship between treatment arm (endovascular versus IV tPA alone) and outcome (favorable versus not) separately for the 2 severity strata as well as overall. The interim findings also included analysis via Cochran–Mantel–Haenszel test and estimation of conditional power under the alternative hypothesis.

Figures 1 and 2 demonstrate trial progression with regard to the estimated treatment effect. Figure 1 presents the absolute risk difference and corresponding 95% confidence interval by severity strata. Figure 2 depicts the overall absolute risk difference, adjusted for strata via Cochran–Mantel–Haenszel weights and corresponding 95% confidence interval, as well as conditional power. Unless noted otherwise, point estimates and confidence intervals are calculated in such a way that positive values favor endovascular therapy. The published trial results correspond to the February 2013 analysis.

First Interim Analysis
The first interim analysis was conducted as scheduled in April 2009, when 225 subjects had completed the primary 90-day follow-up. Overall, the adjusted treatment effect was 3.6% in favor of IV tPA alone. However, with small sample sizes (158 and 67 subjects in moderate and severe strata, respectively), there was a great deal of variability around the point estimate (95% CI, −16.8% to 9.7%). Correspondingly, the estimated conditional power for the overall hypothesized 10% treatment effect was 41.8%, which did not come close to reaching the prespecified futility boundary of 20%.

From a safety perspective, an imbalance toward a greater frequency of asymptomatic intracerebral hemorrhage was noted among the moderate stroke group (NIHSS ≤19), which is the opposite of what was expected. The DSMB conjectured that the imbalance might be because of differences in blood glucose levels and blood pressure management, or because of patients who are less sick reperfusing a larger area, which could lead to more bleeding, and requested that the unblinded statisticians explore this data as a possible explanation. This review, based on 300 randomized subjects, determined that
Of more immediate concern, a potential qualitative (treatment effects in opposite directions) interaction with severity was observed. Among moderate strokes (NIHSS ≤19), the treatment effect was 8.5% (95% CI, −25.1% to 8.0%) in favor of IV tPA alone; among severe strokes (NIHSS ≥20), the treatment effect was 7.8% (95% CI, −13.8% to 29.4%) in favor of the endovascular approach. With small sample sizes, the confidence intervals were extremely wide, and the Breslow–Day test for homogeneity was not significant. If this qualitative interaction truly existed, implying that the original design assumptions were invalid, there was some concern that the trial would require a fundamental redesign. The study investigators were not informed of a potential interaction for 2 reasons. First, the enrollment numbers were small and confidence intervals wide; much more data were needed to assess the qualitative interaction. Second, the DSMB feared sending a message that might inadvertently impact negatively on recruitment. Therefore, the investigators remained unaware of the related discussions.

The DSMB requested that the unblinded statisticians develop a more formal rule for determining whether a qualitative interaction existed, as well as to evaluate the proposed guidelines via simulation. A follow-up call was held between the DSMB statistician and the unblinded statistical team to discuss what should be evaluated. Based on the results of that call, the statistical team developed a plan to conduct calculations to explore the likelihood of this qualitative interaction in the IMS III trial. For the purposes of this investigation, an apparent qualitative interaction was defined in 1 of 2 ways: (1) stratum-specific treatment effects that are in opposite directions, regardless of the magnitude of the effect, and (2) stratum-specific treatment effects that are in opposite directions and of absolute magnitude >0.07 (as observed at the time of the first interim analysis). Calculations were based on the original design parameters, rather than the observed study data.

Under the null hypothesis, the probability of observing an apparent qualitative interaction as defined by (1) was 0.5. This is somewhat intuitive; if there is no treatment effect, then the sign of the estimated stratum-specific treatment effect is just as likely to be positive as negative. The probability of observing an apparent qualitative interaction as defined by (2) was 0.12. Under the alternative hypothesis, the probability of observing an apparent qualitative interaction was 0.28 and 0.07 under definitions (1) and (2), respectively. Therefore, for the hypothesized treatment effect of 10% under which the trial was designed, the observed interaction at the first interim look would occur only 7% of the time. An assumed constant 10% treatment effect across the 2 strata did not explain the observed results very well. But, was there convincing evidence of an interaction?

Formal statistical tests for qualitative interaction, such as the Gail–Simon test and the pushback t procedure, are available. In consultation with the DSMB, the unblinded statistical

Figure 1. Stratum-specific treatment effects. The absolute risk difference and corresponding 95% confidence interval are depicted on the vertical axis, with the time of each interim analysis on the x axis. Effects pertaining to the severe stratum are represented with squares; effects pertaining to the moderate stratum are represented with circles. Negative values favor intravenous tissue-type plasminogen activator; positive values favor endovascular therapy. The dashed line indicates a 0 risk difference; inclusion of this line in the confidence interval implies a nonsignificant effect of treatment within the corresponding stratum.
team conducted simulation studies to assess the power of these tests at interim sample sizes comparable to those in IMS III, but results suggested that such tests would have limited power. These results, in addition to the likelihood calculations described above, were presented to the DSMB for consideration. The DSMB decided to maintain the interim analysis plan as designed, with the caveat that, if the futility boundary was crossed, the magnitude of the apparent qualitative interaction could be considered in the decision to stop the trial for futility. This decision was made before the review of the second interim analysis.

Such is the nature of early evidence in an ongoing trial. The accumulating data may be challenging a fundamental design feature of the trial, yet the information accumulated is limited. Weighing all of these factors, the DSMB recommended trial continuation, but requested that an unplanned interim analysis be performed halfway between the time of originally scheduled first and second interim analyses. The benefit of examining the potential qualitative interaction more closely was thought to outweigh the presumed cost in terms of alpha spent for this additional interim analysis. Moreover, it was noted that, because of the use of the O’Brien–Fleming boundary, the amount of alpha spent at this unplanned interim look would be minimal. Accordingly, a second, unplanned interim analysis was scheduled for the following year. The investigators were informed of the plan for this interim analysis.

Second Interim Analysis
In May 2010, the unplanned second interim analysis was conducted when 347 subjects had completed the 90-day follow-up. At that time, there was a very small treatment effect of 0.8% (adjusted) in favor of IV tPA alone (95% CI, −11.4% to 9.7%). Estimated treatment effects were 3.9% (95% CI, −17.2% to 9.4%) in favor of IV tPA alone and 6.7% (95% CI, −9.7% to 23.0%) in favor of the endovascular approach, respectively, in the moderate and severe stroke strata. Although the stratum-specific treatment effect estimates remained in opposite directions, the treatment effect in the moderate severity stratum had moved substantially toward the null, somewhat mitigating the previous concern over the potential qualitative interaction. The conditional power estimate decreased from the previous 41.8% to 29.5%. This calculation was based on the originally hypothesized treatment effect of 10% in both strata; because the upper limit of the 95% confidence interval for the moderate stratum excluded 10%, there was some concern that the conditional power calculation was overly optimistic.

Although none of the prespecified boundaries had been crossed, extensive discussion among the DSMB members and unblinded statisticians considered possible design changes. These discussions considered the feasibility of restricting recruitment to the severe stratum; however, <30% of enrolled subjects were classified as severe, and the trial would likely not be able to reach its intended sample size. An alternative baseline NIHSS threshold to stratify subjects was also considered; however, a descriptive exploration of the data failed to reveal an alternative that would both maximize recruitment and optimize the observed treatment effect. The pitfalls in making such a design change, based on accruing data, were acknowledged, and the DSMB decided that, because no serious safety concern had been demonstrated and
the prespecified futility boundary had not been crossed, the trial should continue as designed.

Third Interim Analysis
A third interim analysis was conducted in April 2011 when the planned 450 subjects had completed the 90-day follow-up. There was a very small adjusted treatment effect of 0.9% in favor of the endovascular approach, and the overall 95% confidence interval included the hypothesized 10% effect of interest (95% CI, −8.4% to 10.1%). In the moderate and severe strata, respectively, treatment effect estimates were 2.6% (95% CI, −14.4% to 9.1%) in favor of IV tPA alone and 9.1% (95% CI, −5.0% to 23.2%) in favor of the endovascular approach.

Although the stratum-specific treatment effect estimates remained in opposite directions, the magnitude of effect in the lower severity stratum continued to trend toward a null effect. There was another decrease in the conditional power to 23%.

Because the upper limit of the 95% confidence interval for the moderate stratum excluded the hypothesized 10% effect, the DSMB requested that the conditional power be calculated under varying plausible effect sizes in the moderate stratum, whereas the severe stratum was fixed as initially hypothesized at 10%. These results are provided in the Table. Estimates were generated using accumulated data from both strata; stratum-specific estimates were not calculated. Assuming a treatment effect of 9% in the moderate stratum, consistent with the upper bound of the 95% confidence interval, the estimated conditional power is right at, but not below, the prespecified futility threshold. As anticipated, the conditional power declines with the assumed treatment effect. Although these results suggested that it was unlikely that the trial would go on to be positive if continued to the end, this sensitivity analysis did not imply that the futility boundary would be crossed for all plausible values. The trial was allowed to continue in the interest of fully exploring the role of endovascular therapy. Again, however, the DSMB requested an additional unplanned interim analysis, to be conducted in either 1 year or when n=560 subjects (the midway point between the prespecified interim analysis sample sizes of 450 and 675) had observed primary outcomes, whichever came first. The investigators were informed of the need for an additional interim analysis, but, as with the previous additional analysis, were not informed as to the reason to avoid interfering with recruitment momentum.

Fourth Interim Analysis
A fourth (and final) interim analysis was conducted in April 2012. At this point, of the 656 subjects randomized in the trial, 587 had completed the 90-day follow-up. Again, there continued to be a very small adjusted treatment effect of 0.2% in favor of the endovascular approach. Furthermore, the confidence interval now excluded the overall 10% treatment effect of interest (95% CI, −8.0% to 8.3%). The conditional power was estimated at 4%, thus triggering the DSMB to recommend discontinuation of the trial for futility. At the time of this recommendation, the treatment effects had decreased to 2.4% (95% CI, −12.8% to 8.0%) in favor of IV tPA alone and 5.8% (95% CI, −6.6% to 18.3%) in favor of the endovascular approach, respectively, in the moderate and severe stroke strata.

<table>
<thead>
<tr>
<th>Hypothesized Treatment Effect</th>
<th>NIHSS ≤19</th>
<th>NIHSS ≥20</th>
<th>Conditional Power</th>
</tr>
</thead>
<tbody>
<tr>
<td>$P_{\text{endovascular}} - P_{\text{IV}} = 0.10$</td>
<td>$P_{\text{endovascular}} - P_{\text{IV}} = 0.10$</td>
<td>0.230</td>
<td></td>
</tr>
<tr>
<td>$P_{\text{endovascular}} - P_{\text{IV}} = 0.09$</td>
<td>$P_{\text{endovascular}} - P_{\text{IV}} = 0.10$</td>
<td>0.20</td>
<td></td>
</tr>
<tr>
<td>$P_{\text{endovascular}} - P_{\text{IV}} = 0.08$</td>
<td>$P_{\text{endovascular}} - P_{\text{IV}} = 0.10$</td>
<td>0.170</td>
<td></td>
</tr>
<tr>
<td>$P_{\text{endovascular}} - P_{\text{IV}} = 0.07$</td>
<td>$P_{\text{endovascular}} - P_{\text{IV}} = 0.10$</td>
<td>0.141</td>
<td></td>
</tr>
<tr>
<td>$P_{\text{endovascular}} - P_{\text{IV}} = 0.06$</td>
<td>$P_{\text{endovascular}} - P_{\text{IV}} = 0.10$</td>
<td>0.115</td>
<td></td>
</tr>
<tr>
<td>$P_{\text{endovascular}} - P_{\text{IV}} = 0.05$</td>
<td>$P_{\text{endovascular}} - P_{\text{IV}} = 0.10$</td>
<td>0.087</td>
<td></td>
</tr>
<tr>
<td>$P_{\text{endovascular}} - P_{\text{IV}} = 0.04$</td>
<td>$P_{\text{endovascular}} - P_{\text{IV}} = 0.10$</td>
<td>0.07</td>
<td></td>
</tr>
<tr>
<td>$P_{\text{endovascular}} - P_{\text{IV}} = 0.03$</td>
<td>$P_{\text{endovascular}} - P_{\text{IV}} = 0.10$</td>
<td>0.053</td>
<td></td>
</tr>
<tr>
<td>$P_{\text{endovascular}} - P_{\text{IV}} = 0.02$</td>
<td>$P_{\text{endovascular}} - P_{\text{IV}} = 0.10$</td>
<td>0.042</td>
<td></td>
</tr>
<tr>
<td>$P_{\text{endovascular}} - P_{\text{IV}} = 0.01$</td>
<td>$P_{\text{endovascular}} - P_{\text{IV}} = 0.10$</td>
<td>0.029</td>
<td></td>
</tr>
<tr>
<td>$P_{\text{endovascular}} - P_{\text{IV}} = 0.00$</td>
<td>$P_{\text{endovascular}} - P_{\text{IV}} = 0.10$</td>
<td>0.022</td>
<td></td>
</tr>
</tbody>
</table>

NIHSS indicates National Institutes of Health Stroke Scale.

Final Results
Primary outcome data for the final subjects randomized was collected in July 2012. The trend observed during the previous interim analysis was maintained in the final analysis, published in February 2013. There was insufficient evidence to conclude that the proportion of subjects with favorable outcome was different according to treatment (40.8% and 38.7% for endovascular therapy and IV tPA, respectively). Among severe strokes (NIHSS ≥20), the treatment effect was estimated to be 6.8% (95% CI, −4.4% to 18.1%) in favor of the endovascular approach, whereas among moderate strokes (NIHSS ≤19), the treatment effect was estimated to be −1.0% (95% CI, −10.8% to 8.8%). As in the preceding interim analyses, the Breslow–Day test for homogeneity failed to demonstrate a significant interaction between strata and treatment ($P=0.27$).

Discussion
The IMS III trial demonstrates the difficulties that may arise during interim monitoring of ongoing clinical trials because of competing desires to maintain rigidly defined stopping rules at the beginning of the study versus allowing sufficient flexibility to respond to unexpected trial findings. Although it is often the expectation that the magnitude of the treatment effect might vary according to some subgroup, it is rarely anticipated that the direction of effect might differ across primary subgroups within the same trial. In this example, the design was based on an assumption of equal treatment effect across both strata, but the IV tPA–alone moderate stroke group was expected to perform better than the IV tPA severe stroke group. If it became clear that the assumptions were invalid during the study (ie, if a statistically significant interaction were observed), the solution would have been more straightforward—the investigators would have been asked to consider how the study might be redesigned to account for this unexpected interaction. The
redesign might have separated the severity strata into independent trials or called for a trial large enough to detect a clinically relevant qualitative interaction and given due consideration to whether and how the accumulated data could be incorporated into the new design, without sacrificing trial integrity. Although an adaptive design can be useful in allowing design changes in response to accumulating data, a trial maintains its integrity only when it adapts to anticipated occurrences in the prespecified manner. Given that a qualitative interaction was not anticipated in the design stage, there was no prospectively defined plan for adaptation, and potential post hoc adaptations were ultimately rejected.

Similarly, if no concerns about an interaction were ever raised during the review of the study, the predefined stopping boundaries would have the desired properties. However, the IMS III trial presented the most difficult scenario for a statistical team and a DSMB to consider. Although no statistically significant interaction was ever observed, there were ongoing concerns for a possible interaction present across all interim analyses. Correspondingly, the DSMB thought that the possibility of an interaction needed to be considered when interpreting the interim data. This was particularly true given the treatment effects estimated in opposite directions and the relatively low conditional power observed at the interim analyses.

Counterbalancing the statistical issues, the DSMB was acutely aware that clinical stroke trials face significant recruitment challenges. A halt to trial recruitment during some of the data reviews could have powerfully and negatively impacted recruitment momentum. While continuously observing safety and repeatedly checking for evidence of harm, the DSMB recommended trial continuation without pause or notification of the investigators. Fortunately, these decisions did not—after the recruitment of a larger sample size—prove to have exposed additional patients to excessive risk. Whereas, had the trial been stopped sooner, the field could have reacted sooner to the data and initiated a redesigned trial sooner, perhaps with newer devices that might prove to be more effective. In real time, the DSMB wavered on multiple occasions and considered stopping, although no formal boundary had been crossed.

Despite a relatively low conditional power at the time of the interim analyses, the trial was allowed to continue until the prespecified futility boundary had been crossed. This decision can obviously be second-guessed. However, because there were no safety concerns and the DSMB thought the need for the study to reach as definitive a conclusion as possible, the DSMB allowed the trial to continue until the futility concern was overwhelmingly supported by the data, as evidenced by the crossing of the boundary. The continued accumulation of data provided important information regarding the treatment effect in IMS III. Over time, the favorable outcome proportions stabilized with increasing sample size. The final analysis suggested a neutral result in the moderate (NIHSS ≤19) stratum, in contrast with early interim findings that suggested a less favorable effect of endovascular therapy in the same subgroup. Thus, observations of the treatment effect over time made it clear that the finding was not the result of a random occurrence, but instead a trend that was unlikely to reverse itself over time.

It is possible that alternative futility boundaries, including boundaries that become more strict with accumulation of data, might have led to a different outcome, resulting in either early or late trial termination. Varying the boundaries even slightly may dramatically impact the overall operating characteristics of the trial, and it is important to understand the impact that the choice of boundary might have on the probability of early stopping under varying conditions. However, regardless of the choice of boundary, the DSMB and unblinded statisticians would still have struggled with the complex balance between specifying stopping rules at the outset of a study and rigidly sticking with them during the course of the study versus allowing the stopping criteria to be flexible if unexpected findings emerge during the conduct of the study. There are obviously no right answers regarding how to best proceed in these instances, nor is this the first article to highlight these complex issues. Yet, we think that the publication of articles such as this can help to highlight some of these issues and help DSMB members who may encounter similar struggles during the review of some future trials.

Appendix

Cochran–Mantel–Haenszel (CMH) test: This tests the hypothesis that treatment is associated with outcome, when adjusted for the specified stratification variable (in this case, severity strata defined according to National Institutes of Health Stroke Scale score).

Breslow–Day test: This tests the hypothesis that the association of interest in the CMH test is consistent across levels of the stratification variable (in this case, whether the association between treatment and outcome is the same for both the moderate and severe strata).

Power: the probability of correctly rejecting the null hypothesis, given that the specified alternative hypothesis is true.

Conditional power: the probability of rejecting the null hypothesis at the final analysis, given the accumulated data, under the specified alternative hypothesis.

Interaction: the situation in which the effect of treatment differs according to the level of a prognostic variable.

Quantitative interaction: the magnitude of the treatment effect differs according to the level of the prognostic variable, but the direction of the effect remains the same.

Qualitative interaction: both the magnitude and the direction of the effect differ according to the level of the prognostic variable.

Acknowledgments

We would like to acknowledge the entire Data and Safety Monitoring Board (Pat Lyden, Chris Coffey, Marco Di Tullio, Chip Jungreis, and Christine Wijman) for their contributions to the trial and the article.

Sources of Funding

Supported by grants from the National Institutes of Health/National Institute of Neurological Disorders and Stroke (UC U01NS052220, and MUSC U01NS054630 and U01NS077304) and by Genentech, EKOS, Concentric Medical, Cordis Neurovascular, and Boehringer Ingelheim.

Disclosures

Dr Broderick: other research support, significant; Genentech Inc—study medication for National Institute of Neurological Disorders and Stroke–funded Interventional Management of Stroke (IMS) III and CLEARER trials; Genentech Inc—PRISMS trial consultant

Yeatts et al Futility Determination in IMS III 1413
on steering committee; Penumbra—THERAPY trial coinvestigator (paid to department). Dr Coffey: consultant—ZZ Biotech, LLC. Dr Foster: research grant, significant; member of the Statistics and Data Management Center (SDMC) and unblinded statistician for IMS III trial. Dr Lyden: research grant, significant; NN104 3K3A-APC in Ischemic Stroke Study; consultant—ZZ Biotech, LLC. Dr Martin: research grant, significant; member of the SDMC and unblinded statistician for IMS III trial. Dr Palesch: research grant, significant; PI of the SDMC for IMS III trial; consultant—BrainsGate and Edge Therapeutics. Dr Yeatts: research grant, significant; member of the SDMC and unblinded statistician for IMS III trial; consultant—GeneTech; PRISMS Steering Committee member. The other authors have no conflicts to report.

References


Challenges of Decision Making Regarding Futility in a Randomized Trial: The Interventional Management of Stroke III Experience
for the IMS III Investigators

Stroke. 2014;45:1408-1414; originally published online April 3, 2014;
doi: 10.1161/STROKEAHA.113.003925

Stroke is published by the American Heart Association, 7272 Greenville Avenue, Dallas, TX 75231
Copyright © 2014 American Heart Association, Inc. All rights reserved.
Print ISSN: 0039-2499. Online ISSN: 1524-4628

The online version of this article, along with updated information and services, is located on the World Wide Web at:
http://stroke.ahajournals.org/content/45/5/1408

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

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

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