The Effect of an Unplanned Sample Size Re-Estimation on the Type I Error Rate

Leslie A. McClure¹, Jeff M. Szychowski¹, Oscar Benavente², Robert G. Hart³ & Christopher S. Coffey⁴

¹Department of Biostatistics, University of Alabama at Birmingham; ²Department of Neurology, University of British Columbia; ³Department of Neurology, University of Texas Health Sciences Center; ⁴Department of Biostatistics, University of Iowa

Research supported in part by NINDS: U01 NS0385929
Outline

• Introduction
• Background: The Secondary Prevention of Small Subcortical Strokes (SPS3) Study
• Simulation Study
• Conclusions
Introduction

• Sample size estimation for time-to-event analyses requires information *a priori* regarding specific parameters of the study

• Using the approach of Shoenfeld*, computation of the sample size requires specification of the anticipated event rate in each treatment group, as well as an average length of follow-up
  – These parameters are often estimated based on published rates
  – Often, these rates could be quite out-dated by the beginning of the planned study
    • This may impact the true event rate, through advances in technology and changes in the standard of care

SPS3 Background

• The Secondary Prevention of Small Subcortical Strokes (SPS3) study is a 2-by-2 factorial study with a goal of preventing recurrent strokes

• Patients are randomized to one of four treatment combinations

<table>
<thead>
<tr>
<th>Antiplatelet Therapy (clopidogrel)</th>
<th>Usual (130-149 mmHg)</th>
<th>Intensive (&lt;130 mmHg)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Yes</td>
<td>1</td>
<td>2</td>
</tr>
<tr>
<td>No</td>
<td>3</td>
<td>4</td>
</tr>
</tbody>
</table>
The study was designed to address two hypotheses: Whether the time to second stroke is reduced with either:

- Treatment with combination antiplatelet therapy
  - Aspirin + Placebo vs. Aspirin + Clopidogrel
- Aggressive treatment of blood pressure
  - SBP between 130 and 149 vs. SBP < 130

Planned primary analyses for each hypothesis utilizes Cox’s proportional hazards models.

In addition, SPS3 has an interim monitoring plan that includes two interim analyses to assess both efficacy and futility.
SPS3 Background

• Initial sample size computations for SPS3 assumed the following:
  – 7% rate of recurrent stroke annually in the placebo group*
  – A 25% relative risk reduction in recurrent stroke attributable to combination therapy
    • A 5.25% annual risk of recurrent stroke in the combination therapy group
  – 3 years of follow-up
  – Lost-to-follow-up rate of 3% per year
  – Interim monitoring, utilizing Haybittle-Peto boundaries
    • Requires inflation of estimated number of events by 1.007

• Led to a required sample size of 2500

SPS3 Background

- Recent data have emerged, citing recurrent event rates much lower than those upon which the study was planned

<table>
<thead>
<tr>
<th>Study</th>
<th>Sample Size Treated</th>
<th>Antiplatelet Therapy</th>
<th>Annual Recurrent Stroke Rate</th>
</tr>
</thead>
<tbody>
<tr>
<td>SPARCL (2006)</td>
<td>2366</td>
<td>None</td>
<td>2.7%</td>
</tr>
<tr>
<td>ESPRIT (2006)</td>
<td>1376</td>
<td>Aspirin</td>
<td>2.6%</td>
</tr>
<tr>
<td>PROFESS (2008)</td>
<td>10,151</td>
<td>Clopidogrel</td>
<td>3.5%</td>
</tr>
</tbody>
</table>
SPS3 Background

• Recruitment into SPS3 began in February, 2003 and is currently on-going
  – Recruitment has proceeded at a slower than expected rate, requiring an extension of the initial recruitment period

• A lower than expected event rate, coupled with slower than expected recruitment, could result in an underpowered study

• Some solutions:
  – Extend the recruitment period to achieve the planned sample size
  – Re-estimate the sample size mid-study, utilizing newly acquired data regarding the study parameters
Simulation Study

• We assessed the impact of both extending the planned recruitment period in order to achieve the planned sample size AND re-estimating the sample size on the Type I error rate and the power for the primary hypotheses in SPS3
Simulation Study

• We designed the simulation study to mimic the recruitment scheme observed in SPS3, including losses to follow-up, deaths, and withdrawals

• We assessed the power and Type I error rate based on the overall event rate
  – Currently observed rate, and the upper and lower bounds of the 95% confidence interval around that rate

• We assumed no treatment differences in order to assess Type I error and the planned reduction in the rate of recurrent strokes in order to assess power

• We varied the total follow-up time in order to achieve the optimal estimates of power
Simulation Study

• We considered the following scenarios:
  – Two different recruitment schemes:
    • Recruitment to the planned sample size (n=2500), assuming 20, 25 or 30 patients enrolled per month
    • Recruitment of an additional 500 patients (n=3000), assuming 20, 25 or 30 patients enrolled per month
  – Two different follow-up schemes to try to balance recruitment vs. follow-up time
    • N=2500: One and two additional years of follow-up after the close of recruitment
    • N=3000: One additional year of follow-up after the close of recruitment
  – We generated 2000 simulated datasets, and assessed the power and Type I error
# Results from Simulation Study

Table 1: Type I error rates and power resulting from the simulation scenarios

<table>
<thead>
<tr>
<th>Sample Size</th>
<th>Patients Per Month</th>
<th>Additional Years of FU</th>
<th>Type I Error</th>
<th>Power (95% CI)</th>
</tr>
</thead>
<tbody>
<tr>
<td>2500</td>
<td>20</td>
<td>1</td>
<td>4.6%</td>
<td>68% (56%-71%)</td>
</tr>
<tr>
<td></td>
<td>25</td>
<td>1</td>
<td>4.4%</td>
<td>67% (56%-69%)</td>
</tr>
<tr>
<td></td>
<td>30</td>
<td>1</td>
<td>4.8%</td>
<td>66% (55%-68%)</td>
</tr>
<tr>
<td>2500</td>
<td>20</td>
<td>2</td>
<td>4.4%</td>
<td>74% (62%-79%)</td>
</tr>
<tr>
<td></td>
<td>25</td>
<td>2</td>
<td>4.6%</td>
<td>74% (61%-78%)</td>
</tr>
<tr>
<td></td>
<td>30</td>
<td>2</td>
<td>4.2%</td>
<td>71% (61%-77%)</td>
</tr>
<tr>
<td>3000</td>
<td>20</td>
<td>1</td>
<td>4.8%</td>
<td>81% (69%-87%)</td>
</tr>
<tr>
<td></td>
<td>25</td>
<td>1</td>
<td>4.9%</td>
<td>78% (67%-85%)</td>
</tr>
<tr>
<td></td>
<td>30</td>
<td>1</td>
<td>5.1%</td>
<td>75% (66%-82%)</td>
</tr>
</tbody>
</table>
Conclusions

• We found that in the time-to-event setting, adjusting the sample size mid-study based on the overall event rate did not inflate the Type I error rate
  – This was true regardless of the design options

• Power was highest when sample size was increased to 3000 with one additional year of follow-up
  – As recruitment occurs more quickly, power decreases due to the decrease in follow-up time
Conclusions

• The SPS3 Steering Committee recommended that the sample size be increased to 3000 patients, and that recruitment continue for 1 year after recruitment concludes
  – NINDS accepted this proposal, and the SPS3 study has been thusly modified

• Note that this ad-hoc modification to the SPS3 study design does not meet the PhRMA Working Group definition of an adaptive design*

• However, it does classify as a “generally well understood adaptive design with a valid approach to implementation”, as described in the recently released FDA draft guidance document on adaptive designs
  – Adaptation based on blinded interim analyses of aggregate data

• Ultimately, the choice to extend the sample size and/or recruitment relies on a balance between statistical power and resources

Additional References

