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

Leslie A. McClure[1], Jeff M. Szychowski[1],
Oscar Benavente[2], Robert G. Hart[3] &
Christopher S. Coffey[4]

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

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

*Schoenfeld DA (1983). "Sample Size Formulation for the Proportional Hazards Regression Model.", **Biometrics**: 39, 499-503.

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

Systolic Blood Pressure Control

		Usual (130-149 mmHg)	Intensive (<130 mmHg)
Antiplatelet Therapy (clopidogrel)	Yes	1	2
	No	3	4

SPS3 Background

- 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

*Mohr JP, Thompson JLP, Lazar RM, *et al* (2001). "A Comparison of Warfarin and Aspirin for the Prevention of Recurrent Ischemic Stroke", **New England Journal of Medicine**: 345, 1444-1451.

SPS3 Background

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

Study	Sample Size Treated	Antiplatelet Therpay	Annual Recurrent Stroke Rate
SPARCL (2006)	2366	None	2.7%
ESPRIT (2006)	1376	Aspirin	2.6%
PROFESS (2008)	10,151	Clopidogrel	3.5%

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

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

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

*Gallo P, Chuang-Steijn C; Dragalin V; Gaydos B; Krams M; Pinheiro J (2006). "Adaptive Designs in Clinical Drug Development – An Executive Summary of the PhRMA Working Group" . **Journal of Biopharmaceutical Statistics** 16: 275-283.

Additional References

- Amarenco P, Bogousslavsky J, Callahan A, 3rd, Goldstein LB, Hennerici M, Rudolph AE, Sillesen H, Simunovic L, Szarek M, Welch KM, Zivin JA (2006). "High-dose atorvastatin after stroke or transient ischemic attack." **New England Journal of Medicine** 355: 549-559.

- Halkes PH, van Gijn J, Kappelle LJ, Koudstaal PJ, Algra A (2006). "Aspirin plus dipyridamole versus aspirin alone after cerebral ischaemia of arterial origin (esprit): randomised controlled trial." **Lancet** 367: 1665-1673.

- Sacco RL, Diener HC, Yusuf S, Cotton D, Ounpuu S, *et al* (2008). "Aspirin and extended-release dipyridamole versus clopidogrel for recurrent stroke." **New England Journal of Medicine** 359: 1238-1251.