Published in final edited form as:

Contemp Clin Trials. 2012 September ; 33(5): 1088–1093. doi:10.1016/j.cct.2012.06.007.

Sample Size Re-Estimation in an On-going NIH-Sponsored Clinical Trial: The Secondary Prevention of Small Subcortical Stroke Experience

Leslie A. McClure, PhD¹, Jeff M. Szychowski, PhD¹, Oscar Benavente, MD², and Christopher S. Coffey, PhD³

¹Department of Biostatistics, University of Alabama at Birmingham, Birmingham, AL, USA

²Department of Medicine, Division of Neurology, Brain Research Center, University of British Columbia, Canada

³Department of Biostatistics, University of Iowa, Iowa City, IA, USA

Abstract

Background and Purpose—When planning clinical trials, decisions regarding sample size are often based on educated guesses of parameters, that may in fact prove to be over- or underestimates. For example, after initiation of the SPS3 study, published data indicated that the recurrent stroke rates might be lower than initially planned for the study. Failure to account for this could result in an under-powered study. Thus, we performed a sample size re-estimation, and describe the experience herein.

Methods—We evaluated different scenarios based on a re-estimated overall event rate, including increasing the sample size and increasing the follow-up time, to determine their impact on both Type I error and the power to detect the initially planned treatment difference.

Results—We found that by increasing the sample size from 2500 to 3000 and by following the patients for one year after the end of recruitment, we would maintain our planned Type I error rate, and increase the power to detect the prespecified clinically meaningful difference to between 67% and 87%, depending on the rate of recruitment.

Conclusions—We successfully implemented this unplanned design modification in the SPS3 study, in order to allow for sufficient power to detect the planned treatment differences.

Clinical Trials Registration Information—Clinical Trials Registration - <http://clinicaltrials.gov/show/NCT00059306>. Unique identifier: NCT00059306

Keywords

sample size re-estimation; SPS3; randomized clinical trial

© 2012 Elsevier Inc. All rights reserved.

Corresponding Author: Leslie A. McClure, PhD, 1530 3rd Ave S, Birmingham, AL 35294-0022, lmccclure@uab.edu, P: 205-934-5924, F: 205-975-2540.

Disclosures

The authors disclose the following relationships: McClure – none, Szychowski – none, Benavente – grant support from Sanofi Aventis/BMS to conduct the study (drug donation), Coffey – none.

Publisher's Disclaimer: This is a PDF file of an unedited manuscript that has been accepted for publication. As a service to our customers we are providing this early version of the manuscript. The manuscript will undergo copyediting, typesetting, and review of the resulting proof before it is published in its final citable form. Please note that during the production process errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.

Introduction

During clinical trial planning, investigators make a number of assumptions that impact study design. Often, there may be a non-negligible amount of uncertainty associated with these decisions. For example, clinical trials are often designed to detect a clinically meaningful difference between two or more treatments at a specified error rate, with a sample size that achieves a particular power. Even after the ‘clinically meaningful’ effect of interest is specified, investigators must specify values for one or more nuisance parameters in order to determine the required sample size. Examples include the variance (for continuous outcomes) or the control group event rate (for binary or time-to-event outcomes). Usually the specified values of nuisance parameters are based on prior data; however, misspecification of these parameters leads to underpowered or overpowered studies.

Since knowledge accrues as the study progresses, there is increasing interest in using adaptive designs to modify characteristics of ongoing trials, providing greater study flexibility and more efficient utilization of resources^{1–8}. Examples of adaptive designs include a broad range of modifications including adaptive dose finding^{9,10}, enrichment designs¹¹, adaptive randomization¹², adaptive seamless designs^{13,14}, and sample size re-estimation (SSRE)^{15–17}. In 2005, the Pharmaceutical Research and Manufacturers of America (PhRMA) Adaptive Designs Working Group was convened to “...foster and facilitate wider usage and regulatory acceptance of adaptive designs to enhance clinical development...”.¹⁸ Early on, this working group published a white paper providing one of the first formal definitions of an adaptive design: “By adaptive design we refer to a clinical study design that uses accumulating data to modify aspects of the study as it continues”.¹⁸ The definition stresses three important points when using an adaptive design: 1) the changes should be made “without undermining the validity and integrity of the trial”, 2) changes should be made “...by design, and not on an ad hoc basis”, and 3) adaptive designs are “...not a remedy for inadequate planning”. A similar definition appeared in the recent FDA draft guidance document on adaptive designs.²

SSRE allows researchers to reexamine the power at an interim point in the study, given the planned sample size, and the currently observed parameters^{15–17}. If anticipated power is inadequate, the final sample size might be adjusted to allow for a study that would not be underpowered. Similarly, if the study is over-powered, the final sample size might be adjusted downward in order to save resources and complete the study in a more timely manner. There is some controversy surrounding SSRE, with a number of publications describing methods for SSRE based on a re-estimated treatment effect^{20–24}. It is not clear whether there are advantages to using these methods rather than relying on standard group sequential methodology to guide study decisions^{25–27}. However, SSRE procedures involving only nuisance parameters (i.e. those not involving the treatment effect), also known as internal pilot designs (IPs), are generally well accepted. There is vast literature describing their use in a variety of settings.^{4,15–16,28} Importantly, with continuous or binary outcomes, it is generally recognized that IPs can be used in moderate to large trials without concern about increasing type I error.^{16,28,29} This same assumption should apply for time-to-event data, but this needs to be verified.

Materials and Methods

In this paper, we describe use of an unplanned SSRE in the Secondary Prevention of Small Subcortical Strokes (SPS3) study (NCT00059306), an on-going clinical trial that uses a time-to-event outcome. Although there have been a number of publications describing methods for SSRE, there have been very few published examples of studies that have used these methods. Thus, descriptions of practical examples utilizing these methods are

important for advancing both the knowledge and comfort level of investigators considering these designs for possible future studies.

The SPS3 study is an on-going two-by-two factorial study, designed to examine the relationship between antiplatelet therapy (aspirin vs. aspirin plus clopidogrel) and recurrent stroke, and between systolic blood pressure (SBP) control (usual: 130–149 mmHg vs. intensive: <130 mmHg) and recurrent stroke, the details of which are provided elsewhere.³⁰ At the time the study was designed, reports in the literature, including from the WARSS study (comparing warfarin to aspirin),³¹ in which more than half of the participants enrolled had lacunar stroke, reported the recurrent stroke rate among patients on aspirin to be 7% annually. Thus, SPS3 was designed to detect a 25% relative risk reduction in recurrent stroke attributable to combination therapy (5.25% vs. 7%), with 90% power and assuming a loss-to-follow-up rate of 3% per year, and after adjustment due to interim monitoring utilizing the Haybittle-Peto boundaries³² (2 planned efficacy and futility interim analyses), a total of 417 events were required. It was anticipated that patients would be recruited from 35 sites, would be accrued at a rate of 2 patients per site per month over 3 years, and would be followed until one year after enrollment of the last subject. These assumptions led to an assumed average follow-up time of 3 years, which required enrolling a total sample size of 2500 subjects in order to provide the required number of events³³. A difference of the same magnitude and recurrent stroke rates for each of the usual and intensive blood pressure groups similar to those for the antiplatelet arm, were assumed for the blood pressure control arm of the trial, thus requiring the same sample size. The planned analysis was a time-to-event analysis, using a Cox proportional hazards model to determine whether there was a difference in the rate of recurrent stroke between those randomized to aspirin plus clopidogrel, versus those randomized to aspirin alone, as well as to determine whether there was a difference in the rate of recurrent stroke between those randomized to usual versus intensive blood pressure control. SPS3 was approved by the institutional review boards at each participating site, and each patient enrolled provided written informed consent.

SPS3 began recruitment in 2003. The original assumptions regarding the rate of accrual proved to be incorrect, so the SPS3 investigators increased both the number of sites and the accrual period in an attempt to maintain the desired level of mean follow-up need to provide the required number of events. In addition, during the course of recruitment, new randomized trials of secondary stroke prevention enrolling a similar population to SPS3 reported overall stroke rates lower than those that SPS3 was planned upon. Table 1 provides estimates of the annual stroke recurrence rate across different types of secondary stroke prevention trials published since the initiation of SPS3. This prompted the study investigators to initiate a request to the National Institute of Neurological Disorders and Stroke (NINDS) and the study Data Safety and Monitoring Board (DSMB) to perform an unplanned SSRE. After approval from the DSMB and the NINDS, the SPS3 statistical team performed a simulation study in order to determine the impact of increasing the sample size and/or follow-up time after enrolling the final patient, and increasing the sample size on average follow-up per patient, study power and overall type I error.

Simulations

At the time of the simulation study (September, 2008), there were 2081 participants randomized into SPS3. In order to simulate a situation as close to the reality of the SPS3 study as possible, we examined both the rate of early termination from the study (either due to loss-to-follow-up, death or withdrawal), and the overall event rate for all study participants. It is important to note that the observed event rates in the treatment groups, and the difference between them, was not considered as a factor in these simulations. The early termination rate observed in the study at the time was assumed to stay fixed for the duration of the study. We allowed the overall event rate to vary but still considered the originally

planned 25% reduction of interest. For example, the estimated overall event rate of 3.4% annually led to assumed rates of 3.9% in the aspirin + placebo arm and 2.9% in the aspirin + clopidogrel arm. The overall observed event rate and its 95% confidence interval (2.9%, 4.0%) provided a range of event rates that might be expected for the duration of the trial, and we assumed the planned 25% relative risk reduction in recurrent stroke due to the combination therapy. We designed the simulation study to examine power and type I error under a variety of scenarios. We considered the following 3 scenarios:

1. The originally planned study design: recruitment continues until 2500 participants are accrued, and follow-up continues for one year following the end of recruitment;
2. Additional year of follow-up: recruitment continues until 2500 participants are accrued, and follow-up will continue for two years following the end of recruitment; and
3. Additional 500 participants recruited: recruitment continues until 3000 participants are accrued, and follow-up continues for one year following the end of recruitment.

We initially considered a fourth scenario, recruiting 3000 patients and continuing follow-up for two years following the end of recruitment. However, we determined prior to initiating the simulation study that this scenario would unreasonably extend the duration of SPS3 beyond what might be allowable to investigators, patients and the funding agency, and we thus abandoned this scenario. The scenarios are summarized in Table 2.

We used enrollment trends observed to date in the study to generate the first 2081 participants, then simulated a rate of entry into the study for the remaining participants to reflect the current trend in the study (recruit approximately 25 participants per month). However, because this did fluctuate some during the course of the study, we also examined the impact of variable numbers of participants recruited per month (20 and 30 per month, additionally).

We generated event times using exponential random variables, with the parameter specified as the observed overall event rate in each group, the lower bound of the 95% confidence interval around that rate for each group, or the upper bound of the 95% confidence interval around that rate for each group. We also generated loss-to-follow-up times using exponential random variables, assuming the rate to be 5% per year, approximating the rate observed in the study at the time of the simulation. We then determined the minimum of the maximum possible follow-up time (based on the end-of-study date), the event time, and the loss-to-follow-up time, to be the observed event or censoring time for each patient.

We carried out our simulation study using the SAS System[®], and performed 2000 replications of each simulation scenario. We determined the average number of events and the type I error and power from the log-rank test, across each scenario and monthly recruitment trend combination. Because we were interested in determining whether an unadjusted test would inflate the Type I error, we did not do any adjustments to our tests.

Results

Table 3 provides the simulated Type I error rates for the combinations of scenarios presented in Table 2, when the simulation was performed assuming no treatment difference. We found that with one exception, the overall Type I error rate did not exceed the 5% level. The exception occurred for the scenario in which the sample size was increased to 3000, follow-up continued for 1 year after the end of randomization, and the assumption was made that 30 patients were recruited per month. In this case, the simulated Type I error rate was 5.1%, exceeding the nominal 5% level only marginally. Type I error rates were slightly

higher when the sample size was increased to 3000 compared to the scenarios with a sample size of 2500; however none were above the sampling error of the simulations. Thus, while increasing the sample size above the initially planned sample size may slightly inflate the type I error compared to increasing only the follow-up time, it does not do so to an extent that would impact the results of the trial.

Tables 4–6 present the results of the simulation study for each of the scenarios presented in Table 2. Table 4 shows the average follow-up per patient as a function of the sample size, accrual rate and length of follow-up post randomization. As can be seen by the table, increasing both the sample size and the amount of follow-up time increases the average follow-up, while decreasing the accrual rate increases the average follow-up, although marginally so compared to the increase in power seen by increasing the sample size. The amount of variation around the estimates of the average follow-up time within each scenario is small: that is, the average follow-up time does not vary much whether the simulation was run using the observed rate, or the lower or upper bound of the 95% confidence interval around that rate (all standard deviations within 0.04 for all scenarios). Across scenarios, fluctuations in the average follow-up time are seen. By increasing the sample size from 2500 to 3000, for a single year of follow-up after the end of randomization, assuming 20 patients per month enter the study, the average follow-up time can increase by as much as 25%, while assuming 30 patients per month enter the study, the average follow-up time varies less, between 12–18%. This reflects the increase in the total length of time necessary to recruit the entire sample. Conversely, within an assumption of a single sample size (for example, 3000 total participants), the impact of recruiting 30 per month vs. 20 per month results in an approximately 15% decrease in the average follow-up time, reflecting the increase in the speed with which the total sample is obtained.

Table 5 shows the total number of events as a function of the sample size, accrual rate, and length of follow-up post randomization. As with average follow-up, increasing both the sample size and the amount of follow-up time increases the number of events, while decreasing the accrual rate increases the number of events. Due to the increased study time needed to recruit an additional 500 patients, increasing the total sample size from 2500 to 3000 results in large increases in the total number of events. Further, as would be expected, as the overall event rate increases across simulation scenarios, the total number of events increases as well. Within a sample size of 2500 participants, increasing the number of years of follow-up after the sample is recruited from one to two results in an increase in the total number of events; however, the increase is not as large as that resulting from increasing the sample size from 2500 to 3000.

Table 6 describes the simulated power for each of the scenarios described in Table 2. The results indicate that the power increases with sample size, increases as the simulated event rate increases, and slightly decreases as the rate of accrual increases. Thus, scenarios that lead to increased follow-up time and total events correspond to increased power, while scenarios that incur fewer events and less follow-up time correspond to decreased power. In fact, the power for scenarios that do not increase the total sample size from 2500 rarely achieve sufficient power to detect differences of the magnitude that the study was initially powered to detect. With 2500 participants and follow-up until one year after recruitment ends, the power only exceeds 70% for the situation in which 20 patients per month are recruited, and the simulated event rate reflects the upper bound of the 95% confidence interval around the observed rate. The power is better if the follow-up time after the end of randomization is extended to two years; the power exceeds 70% for each scenario for which the observed rate or the upper bound of the 95% confidence interval is used. Regardless of the accrual rate, for the situations in which the sample size is increased to 3000, the simulated power is above 70% for the situations in which the observed event rate or the

upper bound of the 95% confidence interval around the observed rate is used. When the lower bound of the 95% confidence interval around the observed rate is used in the simulation, the power falls to a minimum of 66%. Given the viable options, this seemed to be the best choice for providing the desired level of power. Correspondingly, these results were used by the SPS3 steering committee to request an increase in the planned sample size from 2500 to 3000.

Discussion

We found that modifying the design for the SPS3 study could have a large impact on the likelihood of detecting a treatment difference, should one exist, without having impact on the overall Type I error rate for the study. Because previous work has shown that inflated type I error rates are generally not a concern in larger sample sizes^{4,16,28–29}, we were not surprised to find that in our survival setting and with a relatively large sample, the Type I error rate for the study would be preserved. Armed with this knowledge, the SPS3 Steering Committee discussed the feasibility of the different scenarios, and how best to balance increased power with an increase in the study sample size and the total study length. After lengthy and insightful discussion, the Steering Committee recommended to the NINDS and the DSMB that the sample size be increased to 3000 participants and that follow-up continue for a year following the end of recruitment. After deliberations, the DSMB recommended that this adaptation to the study be made, and the NINDS accepted their recommendation. Note that the results of our SSRE suggested that we increase the sample size; thus, that the trial based on the original study design was underpowered. However, it is possible that a SSRE could indicate that the sample size be reduced, or that a planned study design has more power than anticipated. Wittes & Brittain (1990) discuss this case; however, recommend against it¹⁵, while Birkett & Day (1994) indicate that there may be benefits to allowing a reduction in sample size in the internal pilot setting³⁴. Thus, in cases such as these, a decision may be made to end the study after recruitment of a sample smaller than planned, or SSRE may need to be coupled with interim analyses, in order to facilitate appropriate decision making.^{35–37}

It is worth noting that this modification to the SPS3 study was unplanned. Thus, it does not meet the PhRMA Adaptive Designs Working Group definition of “adaptive by design”.¹⁸ However, the definition in the FDA draft guidance document uses a more relaxed definition for what is meant by prospectively planned: “This can include plans that are introduced or made after the study has started if the blinded state of the personnel involved is unequivocally maintained when the modification plan is proposed.” Interestingly, this example brings up a source of potential confusion with the FDA definition. Although the study statisticians and members of the DSMB had seen interim data labeled in a blinded fashion (e.g., “A” vs. “a”) at the time the request for SSRE was made, the main study investigators had not seen any data. In fact, they will not see any results of the study until the trial ends. Hence, based on the fact that the sample size modification request came from the study investigators who had not seen any interim data, based on published recurrent stroke rates at the time, this example does seem to meet the definition of a “generally well understood adaptive design with a valid approach to implementation” per the FDA Draft Guidance Document. Accordingly, this modification to the study design should not bias the final results. On the other hand, although the DSMB had not seen unblinded treatment effects, they had seen pseudo-blinded interim data (e.g., labeled “A” vs. “a”), which some may argue then does *not* meet the FDA Draft Guidance regarding a valid trial adaptation, indicating a gray area in the literature, which warrants further discussion. This not uncommon situation raises some interesting questions regarding the role that the DSMB should play during the recommendation of an SSRE. This decision can become quite complicated once interim analyses have been completed, since some information about

efficacy is known and it is then impossible to clearly determine that this knowledge in no way affects the decision. One attempt to avoid this issue in the current study was to focus the DSMB discussion purely on the scientific aspects of the SSRE, without consideration of the financial aspects. After the DSMB recommendation was received, NINDS was strictly responsible for addressing the financial issues involved with an increase in the planned sample size. Ideally, SSRE procedures, and adaptive designs more generally, will be planned in the initial design of the study, as advocated in the recent paper by Cheung & Kaufmann.³⁸ As the properties of these designs become better understood, funding agencies will become more willing to allow for planned study adaptations. Importantly, we note that many of these issues would not have been present if the re-estimation of sample size had been planned in advance – and based on clearly pre-specified rules. In this type of instance the DSMB would merely be overseeing the planned design, as opposed to making a decision that affects the design of the study.

Adaptive designs, both pre-specified and unplanned (as in SPS3), require extensive discussions between the study investigators and the trial sponsor. Making modifications to an on-going study has broad implications for operations, and further has broad financial implications. Even in “planned” adaptations, practical issues arise that cannot be anticipated *a priori*, including modifications to institutional review board (IRB) documentation and consent forms, providing adequate funding to continue, dealing with site and participants fatigue, and retention of staff. Further, financial commitments from sponsoring agencies will be necessary to ensure the modifications in study design can be implemented and sustained. In addition to assessing the ramification of the design modifications on the statistical properties of the study, care should be taken to consider the impact of design modifications on each aspect of the study operations prior to implementation. Finally, it is important that investigators publish examples of trials that have utilized an adaptive design so that other researchers can learn from the knowledge obtained (both good and bad).

Acknowledgments

The authors would like to acknowledge the members of the SPS3 Data and Safety Monitoring Board: K. Michael Welch, MD (Chairperson, Rosalind Franklin University); William Clarke, PhD (University of Iowa); Jeffrey Cutler, MD (NIH/NHLBI); Karen Furie, MD (Massachusetts General Hospital); Matthew Mayo, PhD (University of Kansas Medical Center) and the NINDS Liaison to DSMB: Scott Janis, PhD (NIH/NINDS). This research was funded by NIH/NINDS U01 NS0385929

Abbreviations

SSRE sample size re-estimation

References

1. Dragalin V. Adaptive designs: Terminology and classification. *Drug Inf J*. 2006; 40:425–435.
2. Chow, S.; Chang, M. Adaptive design methods in clinical trials. Boca Raton: Chapman & Hall/CRC; 2007.
3. Chow SC, Chang M. Adaptive design methods in clinical trials – A review. *Orphanet Journal of Rare Diseases*. 2008; 3:11. [PubMed: 18454853]
4. Coffey CS, Kairalla JA. Adaptive clinical trials: progress and challenges. *Drugs R D*. 2008; 9:229–42. [PubMed: 18588354]
5. Bretz F, Branson M, Burmann CF, Chuang-Stein C, Coffey CS. Adaptivity in drug discovery and development. *Drug Development Research*. 2009; 70:169–190.
6. Bretz F, Koenig F, Brannath W, Glimm E, Posch M. Adaptive designs for confirmatory clinical trials. *Statistics in Medicine*. 2009; 28:1181–1217. [PubMed: 19206095]

7. Gaydos B, Anderson KB, Berry D, et al. Good practices for adaptive clinical trials in pharmaceutical product development. *Drug Inf J*. 2009; 43:539–556.
8. Quinlan J, Gaydos B, Maca J, Krams M. Barriers and opportunities for implementation of adaptive designs in pharmaceutical product development. *Clinical Trials*. 2010; 7:167–173. [PubMed: 20338900]
9. Garrett-Mayer E. The continual reassessment method for dose-finding studies: A tutorial. *Clinical Trials*. 2006; 3:57–71. [PubMed: 16539090]
10. Bornkamp B, Bretz F, Dmitrienko A, et al. Innovative approaches for designing and analyzing adaptive dose-ranging trials. *J Biopharm Stat*. 2007; 17:965–995. [PubMed: 18027208]
11. Temple R. Enrichment of clinical study populations. *Clin Pharmacol Ther*. 2010; 88:774–778. [PubMed: 20944560]
12. Zhang, L.; Rosenburger, W. Adaptive randomization in clinical trials. In: Hinkelmann, Klaus, editor. *Design and Analysis of Experiments, Special Designs and Applications*. Vol. 3. Hoboken: John Wiley & Sons; 2012.
13. Maca J, Bhattacharya S, Dragalin V, et al. Adaptive seamless phase II/III designs: background operational aspects and examples. *Drug Inf J*. 2006; 40:463–473.
14. Stallard N, Todd S. Seamless phase II/III designs. *Stat Methods Med Res*. 2010; 20:623–634. [PubMed: 20724313]
15. Wittes J, Brittain E. The role of internal pilot studies in increasing the efficiency of clinical trials. *Stat Med*. 1990; 9:65–72. [PubMed: 2345839]
16. Proschan MA. Two-Stage Sample Size Re-Estimation Based on a Nuisance Parameter: A Review. *JBS*. 2005; 15:559–574. [PubMed: 16022163]
17. Proschan MA. Sample size re-estimation in clinical trials. *Biometrical J*. 2009; 51:348–357.
18. Gallo P, Chuang-Stein C, Dragalin V, Gaydos B, Krams M, Pinheiro J. Adaptive designs in clinical drug development – an executive summary of the PhRMA working group. *JBS*. 2009; 16:275–283. [PubMed: 16724485]
19. Food and Drug Administration. Guidance for Industry: Adaptive design clinical trials for drugs and biologics draft guidance. Accessed at: <http://www.fda.gov/Drugs/GuidanceComplianceRegulatoryInformation/Guidances/default.htm>
20. Bauer P, Kohne K. Evaluations of experiments with adaptive interim analyses. *Biometrics*. 1994; 50:1029–1041. [PubMed: 7786985]
21. Proschan MA, Hunsberger SA. Designed extension of studies based on conditional power. *Biometrics*. 1995; 51:1315–1324. [PubMed: 8589224]
22. Lehmacher W, Wassmer G. Adaptive sample size calculations in group sequential trials. *Biometrics*. 1999; 55:1286–1290. [PubMed: 11315085]
23. Chi L, Hung HMJ, Wang SJ. Modification of sample size in group sequential clinical trials. *Biometrics*. 1999; 55:853–857. [PubMed: 11315017]
24. Muller HH, Schafer H. Adaptive group sequential designs for clinical trials: combining the advantages of adaptive and classical group sequential approaches. *Biometrics*. 2001; 23:2497–2508.
25. Tsiatis AA, Mehta C. On the inefficiency of the adaptive design for monitoring clinical trials. *Biometrika*. 2003; 90:367–378.
26. Jennison C, Turnbull BW. Mid-course sample size modification in clinical trials based on the observed treatment effect. *Stat Med*. 2003; 22:971–993. [PubMed: 12627413]
27. Jennison C, Turnbull BW. Efficient group sequential designs when there are several effect sizes under consideration. *Stat Med*. 2006; 25:917–932. [PubMed: 16220524]
28. Friede T, Kieser M. Sample Size Recalculation in Internal Pilot Study Designs: A Review. *Biometrical Journal*. 2006; 48:537–555. [PubMed: 16972704]
29. Proschan MA. Sample size re-estimation in clinical trials. *Biometrics Journal*. 2009; 51:348–357.
30. Benavente OR, White CL, Pearce L, Pergola P, Roldan A, Benavente MF, et al. The Secondary Prevention of Small Subcortical Stroke (SPS3) study. *Int J Stroke*. 2011; 6:164–175. [PubMed: 21371282]

31. Mohr JF, Thompson JLP, Lazar RM, Levin B, Sacco RL, Furie KL, et al. A comparison of warfarin and aspirin for the prevention of recurrent ischemic stroke. *N Engl J Med*. 2001; 345:1444–1451. [PubMed: 11794192]
32. Jennison, C.; Turnbull, BW. *Group Sequential Methods with Applications to Clinical Trials*. Boca Raton: Chapman & Hall/CRC; 2000.
33. Schoenfeld DA. Sample Size Formula for the Proportional Hazards Regression Model. *Biometrics*. 1983; 39:499–503. [PubMed: 6354290]
34. Birkett MA, Day SJ. Internal pilot studies for estimating sample size. *Stat in Med* 1994. 1994; 13:15–30.
35. Morgan CC. Sample size re-estimation in group-sequential response-adaptive clinical trials. *Stat in Med*. 2003; 22:3843–2857. [PubMed: 14673942]
36. Mehta CR, Tsiatis AA. Flexible sample size considerations using information-based monitoring. *Drug Inf J*. 2001; 35:1095–1112.
37. Kairalla JA, Muller KE, Coffey CS. Combining internal pilots with an interim analysis for single degree of freedom tests. *Commun Stat Theory Methods*. 2010; 39(20):3717–3738. [PubMed: 21037942]
38. Cheung K, Kaufmann P. Efficiency perspectives on adaptive designs in stroke clinical trials. *Stroke*. 2011; 42:2990–2994. [PubMed: 21885845]
39. Amarenco P, Bogousslavsky J, Callahan A 3rd, Goldstein LB, Hennerici M, Rudolph AE, et al. High-dose atorvastatin after stroke or transient ischemic attack. *N Engl J Med*. 2006; 355:549–559. [PubMed: 16899775]
40. Halkes PH, van Gijn J, Kappelle LJ, Koudstaal PJ, Algra A. Aspirin plus dipyridamole versus aspirin alone after cerebral ischaemia of arterial origin (esprit): randomised controlled trial. *Lancet*. 2006; 367:1665–1673. [PubMed: 16714187]
41. Sacco RL, Diener HC, Yusuf S, Cotton D, Ounpuu S, Lawton WA, et al. Aspirin and extended-release dipyridamole versus clopidogrel for recurrent stroke. *N Engl J Med*. 2008; 359:1238–1251. [PubMed: 18753638]

Table 1

Annual recurrent stroke rates in secondary stroke prevention trials

Study	Sample Size	Interventions		Number of Lacunar Strokes at Entry	Annual Recurrent Stroke Rate
SPARCL ³⁸ (2006)	4,738	Atorvastatin	Placebo	1,409	2.7% on placebo
ESPRIT ⁴⁰ (2006)	2,739	ASA-ERDP	ASA	1,377	2.6% on ASA
PROFESS ⁴¹ (2008)	20,332	ASA-ERDP	Clopidogrel	10,572	3.5% on Clopidogrel

ASA-ERDP = aspirin plus extended-release dipyridamole

ASA = aspirin

Table 2

Scenarios simulated in order to assess the impact of adapting the study design

Total Sample Size	# Recruited per Month	Date Recruitment Ends	Date Study Ends
2500	20	04/15/2010	04/15/2011
	25	02/13/2010	02/13/2011
	30	12/14/2009	12/14/2010
2500	20	04/15/2010	04/15/2012
	25	02/13/2010	02/13/2012
	30	12/14/2009	12/14/2011
3000	20	05/15/2012	05/15/2013
	25	10/15/2011	10/15/2012
	30	05/15/2011	05/15/2012

Table 3

Simulated Type I error rates for the different scenarios

Total Sample Size	# Recruited per Month	Years of Follow-up Post Recruitment	Type I Error
2500	20	1	4.6%
	25	1	4.4%
	30	1	4.8%
2500	20	2	4.4%
	25	2	4.6%
	30	2	4.2%
3000	20	1	4.8%
	25	1	4.9%
	30	1	5.1%

Table 4

Simulated mean (standard deviation) follow-up (years) across the different scenarios

Total Sample Size	# Recruited per Month	Years of Follow-up Post Recruitment	Mean Follow-up in Years (SD)		
			Lower Bound 95% CI of Observed Event Rate	Observed Event Rate	Upper Bound 95% CI of Observed Event Rate
2500	20	1	3.7	3.6	3.6
	25	1	3.6	3.5	3.5
	30	1	3.4	3.4	3.4
2500	20	2	4.4	4.3	4.2
	25	2	4.3	4.2	4.1
	30	2	4.1	4.1	4.0
3000	20	1	4.5	4.5	4.4
	25	1	4.2	4.1	4.1
	30	1	4.0	3.9	3.8

Table 5

Simulated number of events across the different scenarios

Total Sample Size	# Recruited per Month	Years of Follow-up Post Recruitment	Simulated Number of Events		
			Lower Bound 95% CI of Observed Event Rate	Observed Event Rate	Upper Bound 95% CI of Observed Event Rate
2500	20	1	235	285	335
	25	1	229	276	325
	30	1	222	267	314
2500	20	2	277	337	398
	25	2	271	329	389
	30	2	265	321	379
3000	20	1	342	418	496
	25	1	318	390	462
	30	1	301	368	436

Table 6

Simulated power across the different scenarios

Total Sample Size	# Recruited per Month	Years of Follow-up Recruitment Post	Simulated Power		
			Lower Bound 95% CI of Observed Event Rate	Observed Event Rate	Upper Bound 95% CI of Observed Event Rate
2500	20	1	56%	68%	71%
	25	1	56%	67%	69%
	30	1	55%	66%	68%
2500	20	2	62%	74%	79%
	25	2	61%	74%	78%
	30	2	61%	71%	77%
3000	20	1	69%	81%	87%
	25	1	67%	78%	85%
	30	1	66%	75%	82%