

Adaptive Designs

Sample size re-estimation: A review and recommendations

Keaven M. Anderson

Clinical Biostatistics and Research Decision Sciences

Merck Research Laboratories

Outline

- Introduction/background
- Methods
 - Fully sequential and group sequential designs
 - Adaptive sample size re-estimation
 - ◆ Internal pilot studies
 - ◆ Conditional power methods
- Recommendations
- Case study
- References

Background

■ Origin

- PhRMA Adaptive Design Working Group
- Chuang-Stein C, Anderson K, Gallo P and Collins S, Sample size re-estimation: a review and recommendations. *Drug Information Journal*, 2006; 40(4):475-484

■ Focus

- Late-stage (Phase III, IV) sample size re-estimation
- Frequentist methods
- Control of Type I error
- Potential for bias is critical in these ‘confirmatory’ trials
 - ◆ Implications for logistical issues

Introduction

Adaptive designs

allow design specifications to be changed based on accumulating data (and/or information external to the trial)

Extensive literature exists on adapting through sample size re-estimation, the topic of this talk

Since sample size in group sequential and fully sequential trials are data-dependent, we consider these to be included in a broad definition of adaptive design/sample size re-estimation

Introduction

Why consider sample size re-estimation?

- Minimize number of patients exposed to inferior or highly toxic treatment
- Right-size the trial to demonstrate efficacy
 - ◆ Reduce or increase sample size
- Stop the trial for futility if insufficient benefit
- Incorporate new internal or external information into a trial design during the course of the trial

Introduction

Reasons NOT to use adaptive design

- Methods used to adapt must be carefully chosen
- Planning and logistics setup may prolong startup
- Regulatory scrutiny over methods may prolong startup or put the trial at risk
- May not improve efficiency of non-adaptive method (e.g., group sequential design)

Fully sequential design

- Not commonly used due to continuous monitoring
- May be useful to continuously monitor a rare serious adverse effect
 - Intracranial hemorrhage in a thrombolytic/anti-platelet trial
 - Intussusception in rotavirus vaccine trial
- Unblinded analysis suggests need for an independent monitor or monitoring committee
- References
 - Wald (1947), Sequential Analysis
 - ◆ Sequential probability ratio test (SPRT)
 - Siegmund (1985), Sequential Analysis: Tests and Confidence Intervals

Group sequential design

- Classic
 - Fixed sample sizes for interim and final analyses
 - Pre-defined cutoffs for superiority and futility/inferiority at each analysis
 - Trial stops (adapts) if sufficient evidence available to decide early
 - Independent data monitoring committee often used to review unblinded interim analyses
- Variations
 - Adjustment of interim analysis times (spending functions)
 - Adjustment of total sample size or follow-up based on, for example, number of events (information-based designs)
- Properties well understood and design is generally well-accepted by regulators
- See: Jennison and Turnbull (2000): Group Sequential Methods with Applications to Clinical Trials

SSR=Sample Size Re-estimation

SSR Strategies

- Update sample size to ensure power as desired based on interim results
- Internal pilot studies
 - Adjust for nuisance parameter estimates only
 - Testing strategy: no adjustment to usual tests at end of trial
 - Both blinded and unblinded nuisance parameter estimation methods are available
- Adjusting for interim test statistic/treatment effect
 - All methods adjust based on unblinded treatment difference
 - Adjust sample size to retain power based on interim test statistic
 - ◆ Assume observed treatment effect at interim
 - ◆ Assume original treatment effect
 - Testing strategy: adjust stage 2 critical value based on interim test statistic

Internal pilot studies: review articles

- Friede and Kieser, *Statistics in Medicine*, 2001; 20:3861-73
 - ◆ Also *Biometrical Journal*, 2006; 48:537-555
- Gould, *Statistics in Medicine*, 2001; 20:2625-43
- Jennison and Turnbull, 2000, Chapter 14
- Zucker, Wittes, Schabenberger, Brittain, 1999; *Statistics in Medicine*, 18:3493-3509

Internal Pilot Studies: Continuous Data

- Use some fraction of the planned observations to estimate error variance for continuous data, modify final sample size, allow observations used to estimate the variance in the final analysis.
- Plug the new estimate into the SS formula and obtain a new SS. If the SS re-estimation involves at least 40 patients per group, simulations have shown (Wittes et al, SIM 1999,18:3481-3491; Zucker et al, SIM 1999,18:3493-3509)
 - The type I error rate of the unadjusted (naïve) test is at about the desirable level If we do not allow SS to go down
 - The unadjusted test could lead to non-trivial bias in the type I error rate If we allow the SS to go down
 - Power OK
 - Unblinded methods may allow back-calculation of interim treatment difference
- Coffey and Muller (Biometrics, 2001, 57:625-631) investigated ways to control the type I error rate (including different ways to do SSR).
- Denne and Jennison, (Biometrika, 1999) provide a group sequential version

Internal pilot studies: blinded SSR

- When SSR is based on nuisance parameters
 - Overall variability (continuous data)
 - Overall rate (binary data)
- Advantage
 - No need to break the blind.
 - In-house personnel can do it.
 - Minimal implication for Type I error rate.
- Disadvantage
 - The estimate of the nuisance parameter could be wrong, leading to incorrect readjustment.

Internal pilot studies: blinded SSR

- Continuous (normal) data methods
- Gould and Shih (1992)
 - EM algorithm
 - Software: Wang, 1999
- Friede and Kieser (2001)
 - Assume treatment difference known (no EM algorithm required)
 - Adjust within group sum of squares using this constant
- Type I error and power appear good
 - Some controversy over appropriateness of EM (Friede and Kieser, 2002, 2005; Gould and Shih, 2005)
- Recent papers on blinded methods based on blocking
 - Van der Meulin, J Biopharm Stat, 2005; 15: 479-489
 - Xing and Ganju, Statistics in Medicine, 2005; 24:1807-1814
- Questions to ask:
 - How well will this work if treatment effect is different than you have assumed for the EM procedure?
 - Does this work with non-normal data?
 - Will it be under- or over-powered?
 - Group sequential version (Gould and Shih, 1998) may bail you out of this

Internal pilot studies: unblinded SSR

■ Advantage

- Could provide more accurate sample-size estimate.

■ Disadvantages

- Interim treatment difference might be calculated based on knowledge of formula and updated sample size
 - ◆ There could be concerns over bias resulting from knowledge of interim observed treatment effect.
- Typically require an external group to conduct SSR for registration trials

Internal pilot studies: binary data

- See, e.g., Herson and Wittes (1993), Jennison and Turnbull (2000)
- Estimate control group event rate at interim
 - Type I error OK if interim n large enough
- Options (see Jennison and Turnbull, 2000 for power study)
 - Assume $p_1 - p_2$ fixed
 - ◆ Power appears OK
 - Assume p_1/p_2 fixed
 - ◆ Can be underpowered

Combination tests

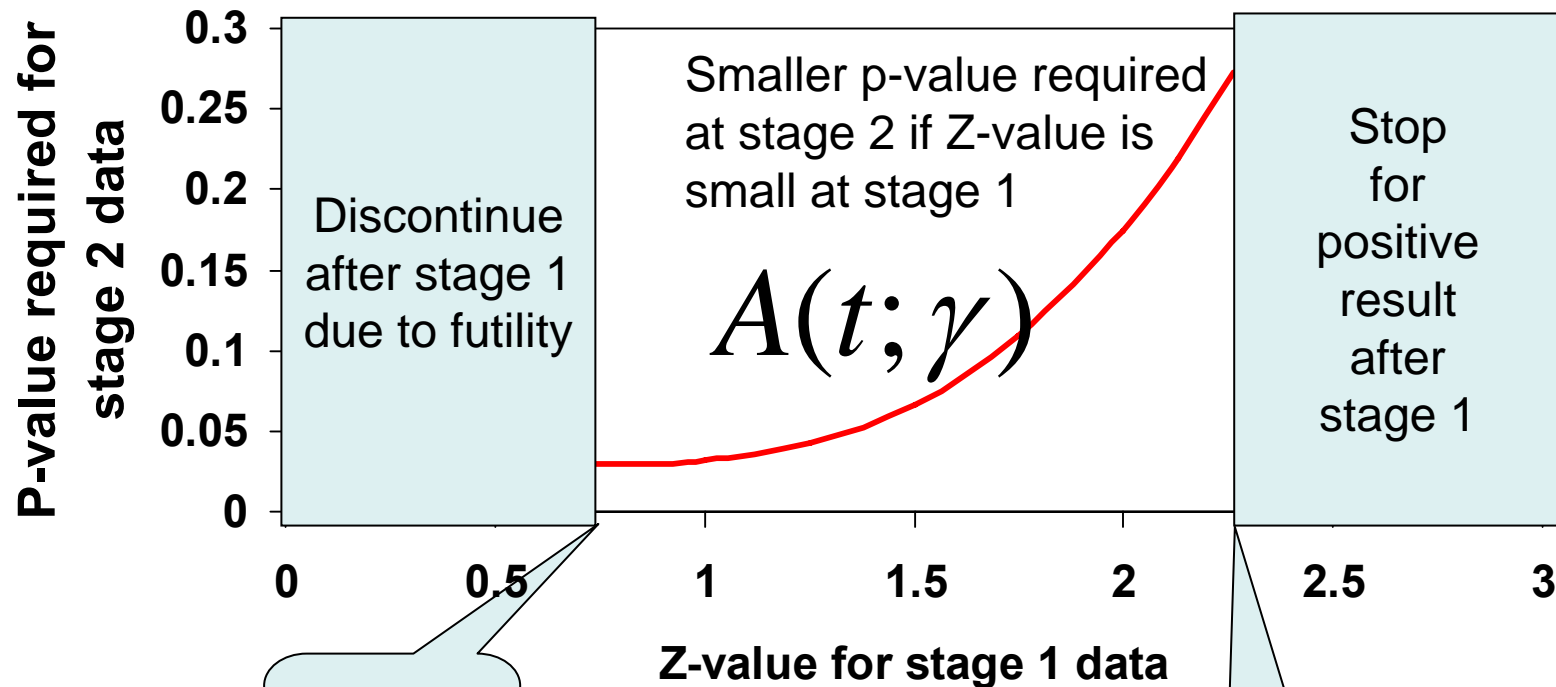
- Methods for controlling Type I error
- Method: calculate separate standardized test statistics from different stages and combine them in a predefined way to make decisions.
 - Weighting of a stage does not increase if sample size for that stage is increased
 - Violates 'one-patient one-vote principle'
 - ◆ Efficiency issue (Tsiatis and Mehta, 2003)
 - ◆ Can be taken to ridiculous extreme (Burman, 2006)
 - Many methods available, including
 - ◆ Fisher's combination test (Bauer, 1989)
 - ◆ Conditional error functions (Proschan and Hunsberger, 1995; Liu and Chi, 2001)
 - ◆ Inverse normal method (Lehmacher and Wassmer, 1999)
 - ◆ Variance spending (Fisher, 1998)

Combination tests

- Apply combination test method to determine the critical value for the second stage based on the observed data from the first stage.
- Make assumption on treatment effect; options include:
 - Observed effect (highly variable)
 - External estimate
 - Original treatment effect used for sample size planning
- Compute next stage sample size based on critical value, set conditional power to originally desired power given interim test statistic and assumed second stage treatment effect
 - Generally, will only raise sample size – not lower

Conditional Error Function

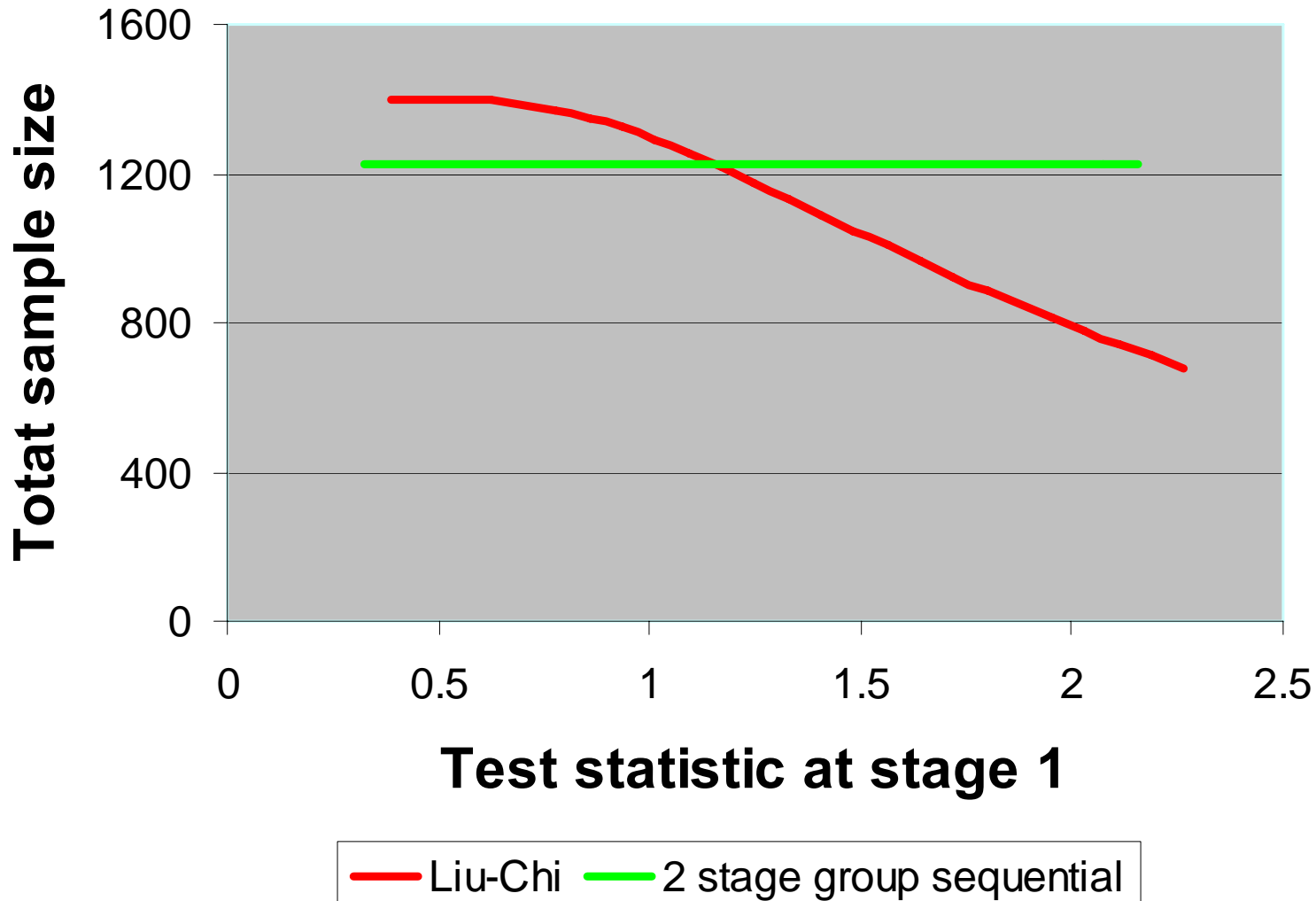
2-stage adaptive design



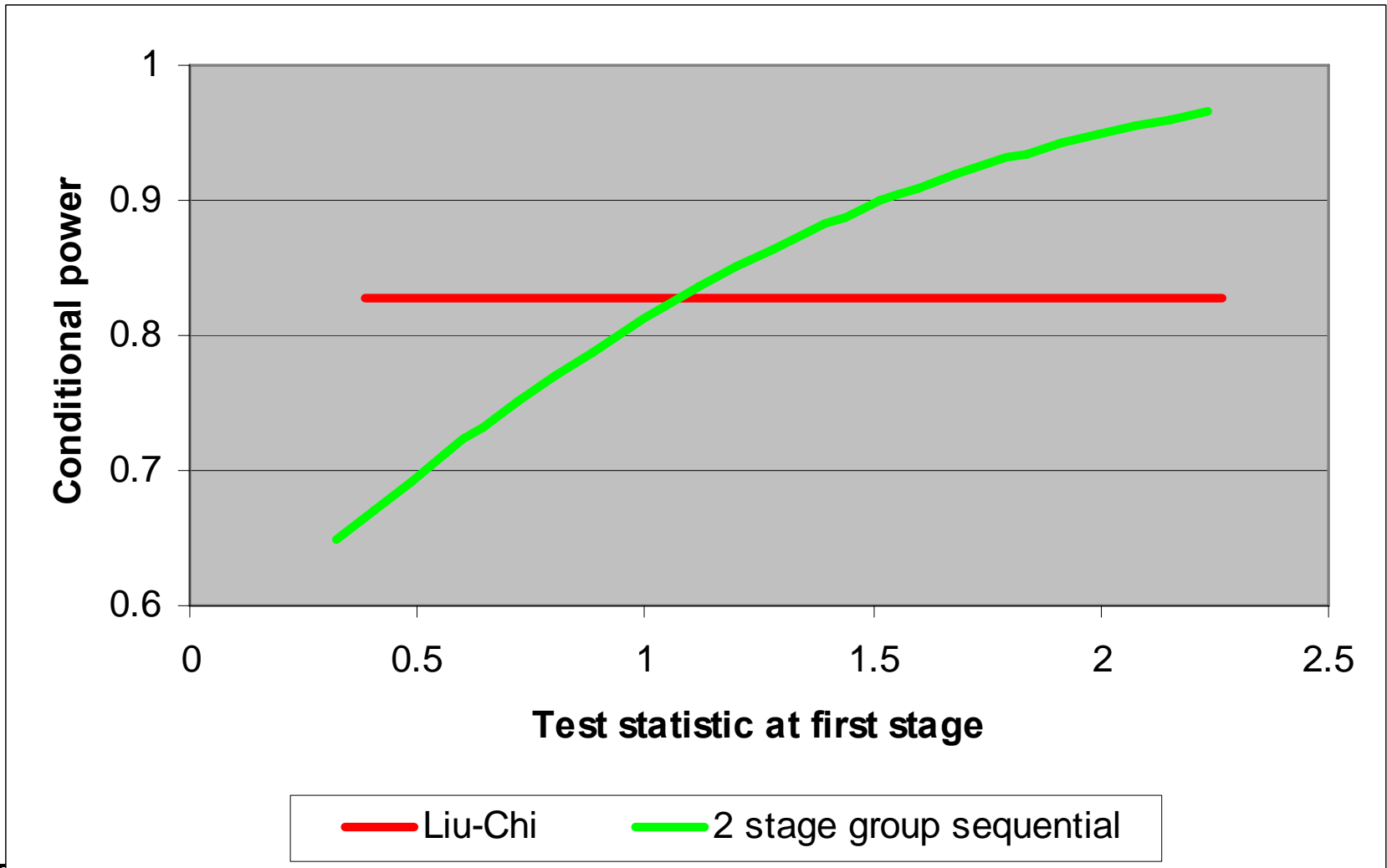
$$z_{1-\alpha_1^*}$$

$$z_{1-\alpha_1}$$

Total sample size based on stage 1 test statistic



Conditional Power Functions (conditioning on $\theta=\delta$)



Group Sequential vs SSR Debate

■ Efficiency

- The adaptive designs for SSR using combination tests with fixed weights are generally inefficient (Jennison and Turnbull, 2003; Tsiatis and Mehta, 2003).
- Efficient adaptive designs for SSR have little to offer over efficient group sequential designs in terms of sample size (Jennison and Turnbull, 2006). However, the latter might require more interim analyses and offer minimum gain. In addition, the comparisons were made as if we knew the truth.

■ Flexibility and upfront resource commitment

- SSR offers flexibility and reduces upfront resource commitment. The flip side is the need to renegotiate budget and request additional drug supply when an increase in SS is necessary.
- SSR addresses uncertainty at the design stage.

Recommendation #1

- Before considering adaptive sample-size re-estimation, evaluate whether or not group sequential design is adequate
 - Pros:
 - ◆ Regulatory acceptance
 - ◆ Well-understood methods allow substantial flexibility
 - ◆ Experienced monitoring committee members available
 - Cons:
 - ◆ May not work well in some situations when trial cannot be stopped promptly (long follow-up, slow data collection, cleaning or analysis)

Recommendation #2

- Anticipate as much as possible at the planning stage the need to do SSR to incorporate information that will accumulate during the trial
 - Treatment effect size
 - Nuisance parameters
 - The effect of environmental changes on the design assumptions
- Do not use SSR to
 - Avoid up-front decisions about planning
 - As a 'bait-and-switch' technique where a low initial budget can be presented with a later upward sample size adjustment.

Recommendation #3

- For SSR based on variance, consider using blinded SSR
 - However, when there is much uncertainty about the treatment effect, consider using unblinded SSR.

- For a binary outcome, one can either do blinded SSR based on the overall event rate or an unblinded SSR based on the event rate of the control group. There is no clear preference, choice dependant on several factors.
 - If there is much uncertainty about treatment effect, unblinded SSR using conditional power methods (see next slides).
 - If SSR is blinded, consider conducting interim analysis to capture higher than expected treatment effect early.

Recommendation #4

- To help maintain confidentiality of the interim results, we recommend considering the following:
 - Do not reveal exact method for adjusting sample size.
 - Make the outcome of SSR discrete with only 2-3 options.
- Under the first approach, details on SSR methodology will not be described in the protocol, but documented in a stand-alone statistical analysis plan for SSR not available to study personnel.
- For SSR based on observed treat effect (continuous case), it will be beneficial to base SSR on both variability and effect to limit the ability to back-calculate the interim treatment difference.
- We recommend that the protocol include the maximum sample size allowed to minimize the need to go back to the IRB.

Recommendation #5

- For unblinded SSR
 - Invite a third party to do the calculations following a pre-specified rule.
 - If possible, combine SSR with a group sequential design where SSR will be conducted at the same time with an interim analysis.
 - ◆ Convene a DMC (or preferably an IDMC) to review the SSR recommendation from the third party. If an IDMC is used, the IDMC statistician can carry out the SSR.
 - Assuming Recommendation #4 is followed, the new sample size will be communicated to the sponsor. The investigators will be told to continue enrollment.

Recommendation #6

- Carefully consider the number of times to do SSR.
 - E.g., for variance estimation, is once enough?
- Timing of the SSR should be based on multiple considerations such as
 - available info at the design stage,
 - disease,
 - logistics
 - ◆ delay from enrollment until follow-up complete and data available
 - ◆ enrollment rate,
 - Method
 - ◆ whether the SSR will be based on variance or treatment effect
 - Gould and Shih (1992) recommend early update as soon as variance estimate stable due to administrative considerations, while Sandvik et al. (1996) recommend as late as possible to get accurate variance estimate

Recommendation #7

- Acceptance of SSR by regulators varies, depending on the reasons for SSR. In general, blinded SSR based on a nuisance parameter is acceptable.
- When proposing unblinded SSR, should include
 - The objective for SSR
 - Statistical methodology including the control of Type I error
 - When to do the SSR
 - How to implement (e.g., DMC, third party)
 - How to maintain confidentiality
 - How will the results be shared
 - Efficiency (power/sample size) considerations
- Discuss the plan with regulatory agencies in advance.

Case Study: Phase II/III Adaptive Stroke Trial

- Background
 - Only 'weak' proof of concept for benefit in early studies
- Design properties desired
 - Kill program early if no emerging benefit seen
 - ◆ Aggressive futility rule required
 - Start a second, confirmatory trial if strong evidence of emerging benefit observed
 - Apply modest sample size increase if early benefit is moderate, but promising (delay start of 2nd trial)
 - Maintain integrity of this trial as a potentially pivotal trial
 - ◆ Independent DMC to review unblinded interim analyses
 - ◆ Only reveal broad trend of results to sponsor at interim analysis 1

Stroke Trial: Interim Analysis Decision Rules

Early interim analysis performed when n=200/arm

<u>Z-value</u> (1-sided P-value)	<u>Interpretation</u>	<u>Trial action</u>	<u>Program action</u>
$Z \leq 0.679$ ($P \geq 0.25$)	No benefit or small benefit observed	Stop trial for futility	Stop development
$0.679 < Z < 1.763$ ($0.25 > P > 0.04$)	Small to moderate benefit observed	Enroll 750/arm	Delay confirmatory trial decision
$1.763 \leq Z < 3.023$ ($0.04 \geq P > 0.001$)	Moderately large benefit observed (not definitive)	Enroll 600/arm	Accelerate confirmatory trial
$Z \geq 3.023$ ($P \geq 0.001$)	Definitive benefit observed	Stop trial for efficacy	Initiate confirmatory trial

FDA draft guidance on sponsor unblinding at interim analysis (excerpts)

- In general, sponsors should avoid seeking information about unblinded interim data and should consider the significant possibility that they may wind up impairing trial management or even making the trial results uninterpretable by doing so.
- The sponsor should consider discussing such an action with FDA in advance.
- The sponsor should determine the minimum amount of information needed.
 - **The sponsor should formulate written questions, preferably with yes/no rather than numerical answers, that will elicit only that minimal required information**
 - Should identify individuals with a critical “need-to-know” and SOPs should ensure that no one else has access to such information.
 - Individuals with access to the information should avoid any further role in the management of the trial and should minimize interactions with others in that role.
 - Where possible, individuals who have access to such information should avoid taking actions that will assist others in inferring what the information is.

Stroke trial results

- Interim primary efficacy results nearly identical in active and control groups
 - With aggressive boundary, trial was stopped for futility
 - Without aggressive boundary:
 - ◆ Considerably more patients would have been exposed to experimental agent with little chance of demonstrating benefit
 - ◆ Company resources could not be redirected towards more promising therapies

References

- Anderson et al (2004) JSM Proceedings, ASA.
- Anderson (2006) Biometrical Journal, to appear
- Bauer, P. (1989) Biometrie und Informatik in Medizin und Biologie 20:130–148.
- Birkett, Day (1994) Stat in Medicine, 13:2455-2463
- Brannath, Konig, Bauer (2003) Biometrical J, 45:311-324
- Bristol (1993) J of Biopharm Stat, 3:159-166
- Cheng, Shen (2004) Biometrics, 60:910-918
- Coburger, Wassmer (2003) Biometrical J, 45:812-825
- Coffee, Muller (2001) Biometrics, 57:625-631
- Cui, Hung, Wang (1999) Biometrics, 55:853-857
- Denne (2000) J of Biopharm Stat, 10:131-144
- Denne, Jennison (1999) Stat in Medicine, 18:1575-1585
- Denne, Jennison (2000), Biometrika, 87:125-134
- Fisher (1998) Stat in Medicine, 17:1551-1562

References

- Friede, Kieser (2002) Stat in Medicine, 21:165-176.
- Friede, Kieser (2003) Stat in Medicine, 22:995-1007.
- Friede, Kieser (2005): Stat in Medicine 24:154-156.
- Friede, Kieser (2006) Biometrical Journal, 2006; 48:537-555
- Friede, Kieser (2004) Pharm Stat, 3:269-279
- Gould (1992) Stat in Medicine, 11:55-66.
- Gould, Shih (1992) Comm in Stat (A), 21:2833-2853.
- Gould, Shih (1998) Stat in Medicine, 17:89-100.
- Gould, Shih (2005) Stat in Medicine 24:147–156
- Herson, Wittes (1993) Drug Info Journal, 27:753-760
- Jennison, Turnbull (2003) Stat in Medicine, 22:971-993
- Jennison, Turnbull (2004) Technical Reports, U of Bath, UK
- Jennison, Turnbull (2006) Biometrika, 93:1-21
- Jennison, Turnbull (2006) Stat in Medicine: 25:917-932

References

- Kieser, Friede (2000) Drug Info Journal, 34:455-460.
- Kieser, Friede (2000) Stat in Medicine, 19:901-911
- Muller, Schafer (2001) Biometrics, 57:886-891
- Mehta, Tsiatis (2001) Drug Info Journal, 35:1095-1112
- Lehmacher, Wassmer (1999) Biometrics, 55:1286-1290
- Liu, Chi (2001) Biometrics, 57:172–177
- Lokhnygina (2004) NC State stat dissertation (with Anastasio Tsiatis as the advisor).
- Posch, Bauer, Brannath (2003) Stat in Med, 22:953-969
- Proschan, Hunsberger (1995) Biometrics, 51:1315–1324
- Sandvik, Erikssen, Mowinckel, Rodland (1995), Statistics in Medicine, 15:1587-90
- Schwartz, Denne (2003) Pharm Stat, 2:263-27

References

- Shih (1993) Drug Information Journal, 27:761-764.
- Shih, Gould (1995) Stat in Medicine, 14:2239-2248.
- Shih, Long (1998) Comm in Stat (A), 27:395-408.
- Shih, Zhao (1997) Stat in Medicine, 16:1913-1923.
- Shun, Yuan, Brady, Hsu (2001) Stat in Medicine, 20:497-513
- Tsiatis, Mehta (2003) Biometrika, 90:367–378
- Wust, Kieser (2003) Biometrical J, 45:915-930.
- Wust, Kieser (2005) Comp Stat & Data Analysis, 49:835-855
- Wittes, Brittain (1990) Stat in Medicine, 9:65-72
- Wittes et al (1999) Stat in Med, 18:3481-3491
- Zucker, Wittes et al (1999) Stat in Medicine, 18:3493-3509