Adaptive Blinded Sample Size Adjustment for Comparing Two Normal Means – A Mostly-Bayesian Approach

Andrew Hartley, PhD
Associate Statistical Science Director

Summary

- The need for sample size adjustment (SSA)
- Review of standard approaches
- Intuition & logic underlying proposed approach
- Effects of blinded data on beliefs about mean treatment difference
- Comparing new approach vs. standard ones
  - Mean-Mean Absolute Deviation
  - Risk-adjusted Net Present Value (rNPV)
- Effects on estimation & power
- Sensitivity to Normal assumption
- Discussion - Q&A, comments

Context

- Clinical trial - Randomized, Blinded
- Comparing 2 normal means
- Sample size usually chosen to provide targeted level of statistical power
  - Power too high? $\Rightarrow$ Waste of resources, unethical to expose too many patients
  - Power too low? $\Rightarrow$ Might get equivocal results
- Power, for any given sample size, depends on

  Δ = population mean treatment difference

  Σ = population within-treatment variance (MSE)
- Problem: Usually,

  Much uncertainty attends Δ and Σ

  $\Rightarrow$ power cannot be determined precisely (!)

$\Rightarrow$ Hence, popularity of sample size adjustment

Standard Sample Size Adjustment Approaches

Common Steps

- At protocol planning stage, set provisional sample size target
- Collect some on-trial data
- Re-estimate Δ and/or Σ
- Re-estimate sample size requirement

Standard Sample Size Adjustment Approaches

Unblinded and **blinded** approaches exist

- Standard **Unblinded** approaches
  - Require
    - Data Monitoring Committee (DMC)
    - Adjusted test statistics (Cui-Hung-Wang 1999, Chen-DeMets-Lan 2004, etc.)—often are inefficient
    - Procedures for protecting blind
  - Complicate reporting / interpretation
  - Support re-estimating both Δ and Σ
- Standard **Blinded** approaches
  - Usually, based on S_b^2 = blinded (overall) sample variance
    - Approximately, $E(S_b^2) - \Delta_0^2/4 = \Sigma$
    - So, Σ estimated as $\underline{S_b^2 - \Delta_0^2/4}$
  - Obviate DMC, adjusted statistics
  - Support re-estimating Σ only
  - Advisable <u>only when Δ estimated with high precision</u> (this severely curtails usefulness)

Proposed Blinded Sample Size Adjustment Method

- We have *some* prior ideas of plausible values of Δ and Σ; if we didn't, would we be running the clinical trial?
- Often, we can summarize these beliefs well using

$$\Delta \sim N(\theta, \tau^2) \quad \perp\perp \quad \Sigma \sim Gamma(\alpha, \beta)$$

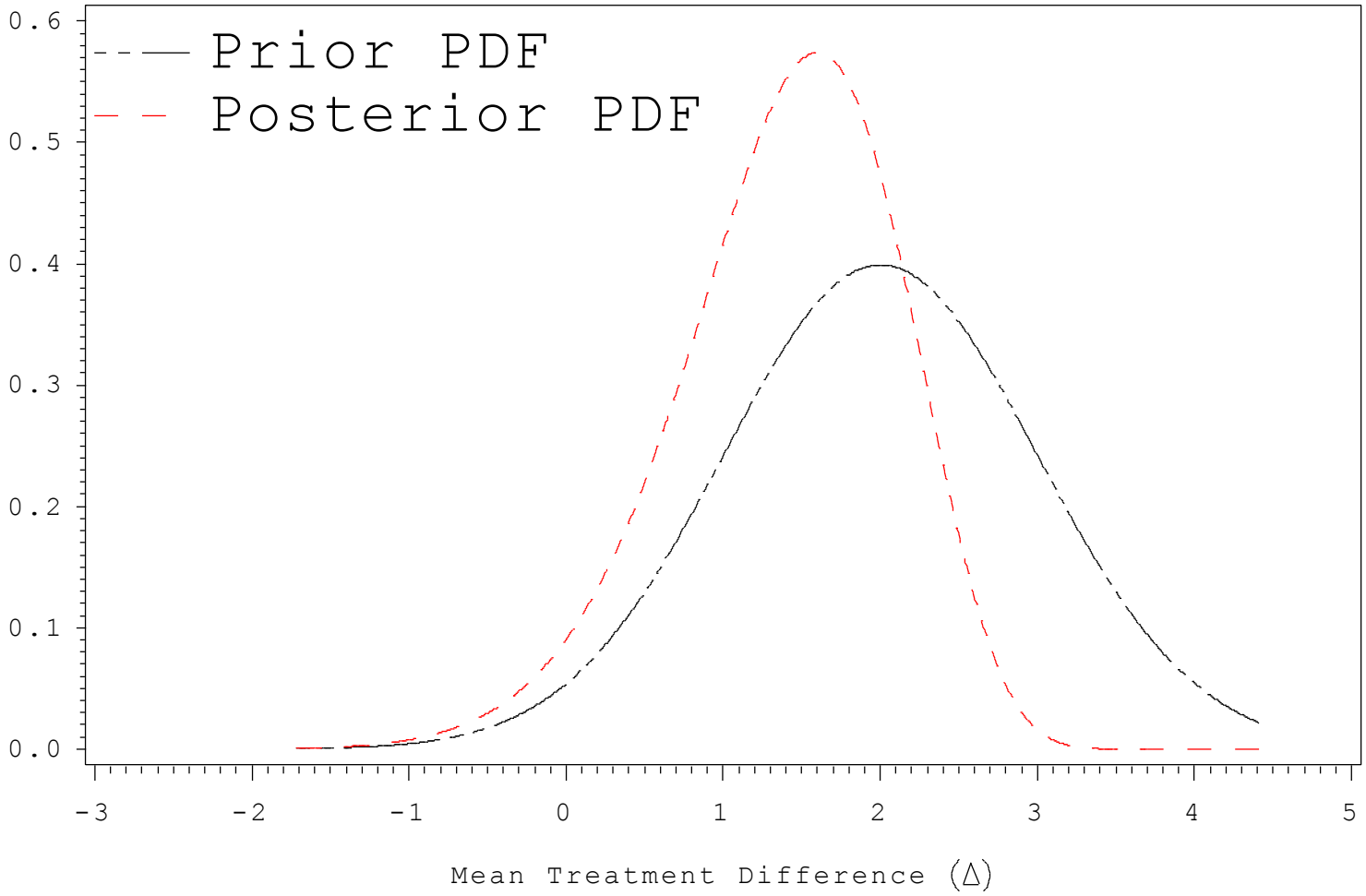
Proposed Blinded Sample Size Adjustment Method (con't)

- Recall

$$E(S_b^2 \mid \Sigma, \Delta) = \Sigma + k\Delta^2$$

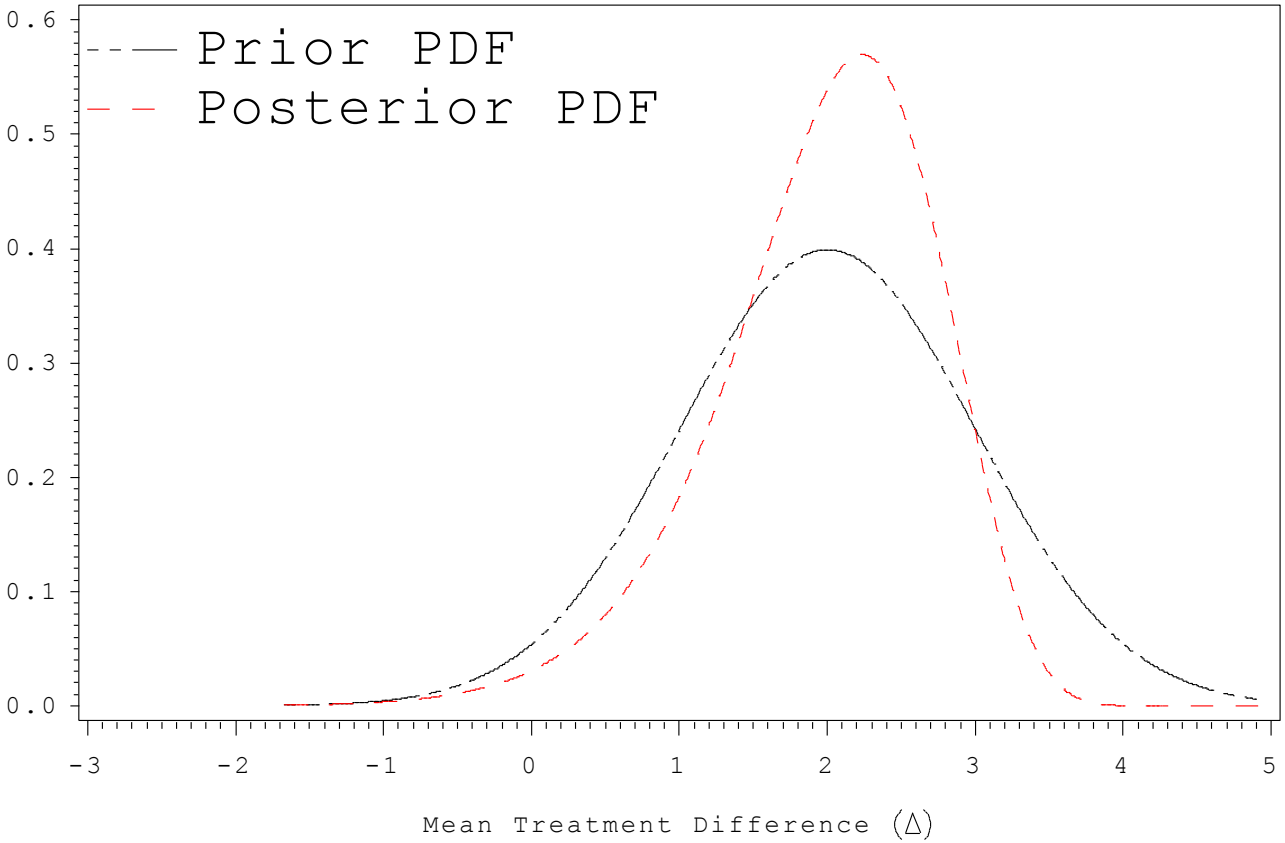
  where $k \rightarrow 1/4^+$.
- Observing S_b^2 refines the prior beliefs concerning both Δ and Σ (not just concerning Σ)
  - If Var(Δ)<<Var(Σ), then S_b^2 mainly shifts PDF of Σ
  - If Var(Σ)<<Var(Δ), then S_b^2 mainly shifts PDF of Δ
- Example in which Var(Σ)=Var(Δ)…

Effects of Blinded Data on Beliefs about Δ (1)



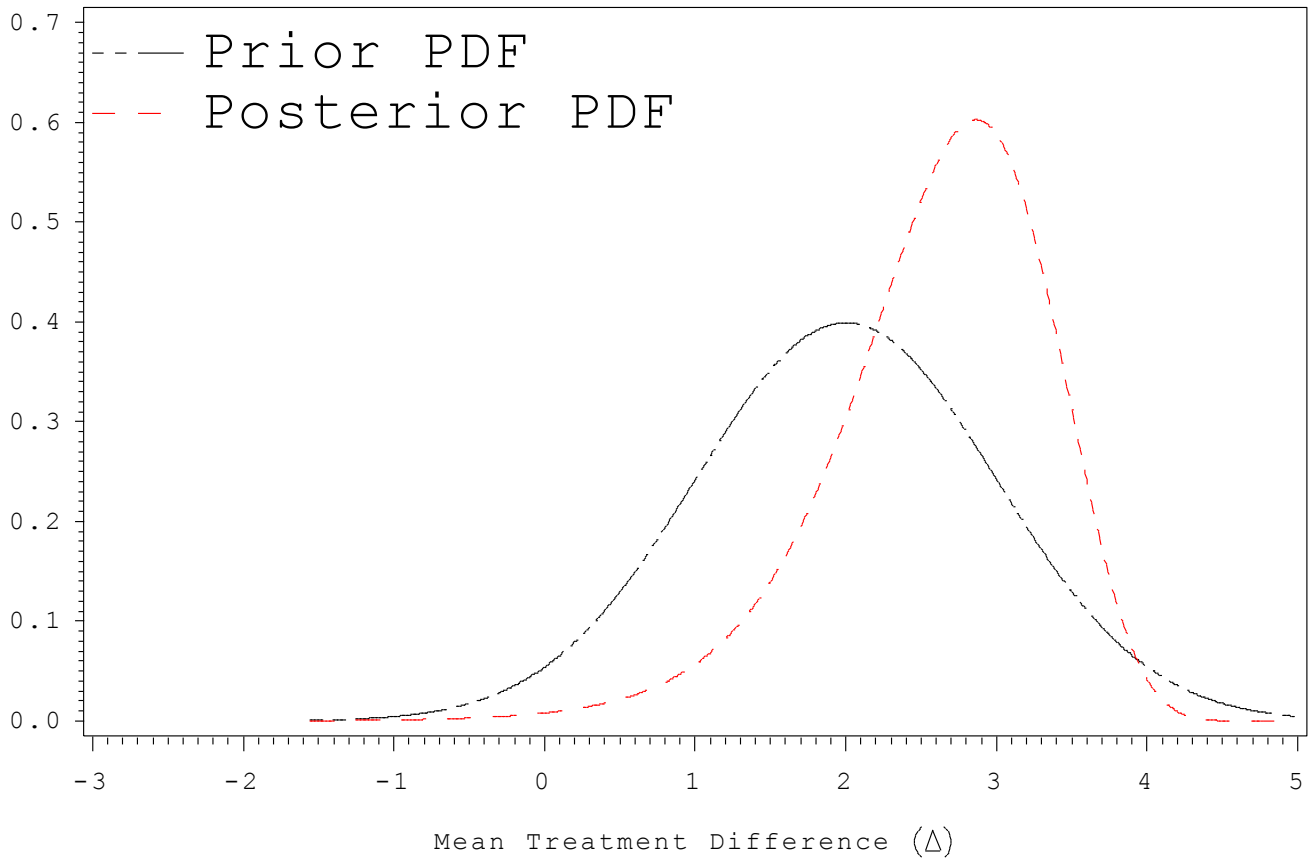
1.1. Expected and Observed Blinded Variance=(6.2931,5.03448).

Effects of Blinded Data on Beliefs about Δ (2)



1.2. Expected and Observed Blinded Variance=(6.2931,6.2931).

Effects of Blinded Data on Beliefs about Δ (3)



1.3. Expected and Observed Blinded Variance=(6.2931,7.55172).

Proposed Blinded SSA Procedure – How to Perform

- For each candidate N per treatment group, calculate "Predictive Power:"

$$\int_0^{\infty}\int_{\Theta}\left[\begin{array}{l} ConditionalPower(\delta,\Sigma,N)\times \\ \pi_{\Delta}\left(\delta \mid \sigma^2, s_b^{\ 2}\right)\partial\delta\pi_{\Sigma}\left(\sigma^2 \mid s_b^{\ 2}\right)\partial\Sigma \end{array}\right]$$

Integration subject to

$$\Sigma + k\Delta^2 = E[\Sigma + k\Delta^2 \mid S_b^2]$$

(Predictive Power = probability of statistical significance at the end of the trial, accounting for the uncertainties with respect to Δ and Σ)

- Choose N such that Predictive Power = target desired

Calculating $E[\Sigma + k\Delta^2 \mid S_b^2]$

- $E(S_b^2 \mid \Sigma, \Delta) = \Sigma + k\Delta^2$ where $k \cong 1/4$.
- So, we could simply estimate $\Sigma + k\Delta^2$ as S_b^2
- However, bayesianly using prior information enhances estimation

General bayesian inferential result is that, if

$$\theta \sim N(\mu, \tau^2) \qquad \& \qquad X \mid \theta \sim N(\theta, \sigma^2)$$

then

$$\theta \mid x \sim N(\mu(x), v^2)$$

where

$$\mu(x) = \frac{\sigma^2 \mu + \tau^2 x}{\sigma^2 + \tau^2}, \quad v^2 = \frac{\sigma^2 \tau^2}{\sigma^2 + \tau^2}.$$

⇒ Improved estimate of θ is a weighted average between the *prior* expectation & the empirical estimate.

Calculating $E[\Sigma + k\Delta^2 \,/\, S_b^2]$ (con't)

$\Delta \sim N(\theta, \tau^2) \quad \perp\!\!\perp \quad \Sigma \sim Gamma(\alpha, \beta)$

$\Rightarrow E(\Sigma + k\Delta^2) = \alpha\beta + k(\theta^2 + \tau^2)$

$\Rightarrow$ Improved estimate of $\Sigma + k\Delta^2$ is

$$E[\ \ + k\Delta^2 | S_b^2] = \frac{a[\alpha\beta + k(\theta^2 + \tau^2)] + bS_b^2}{a + b}$$

for some a & b, as weights of $E(\Sigma + k\Delta^2)$ & S_b^2.

a & b are the <u>sampling</u> variance & the <u>prior</u> variance.

New SSA Method - Summary of Steps

1. Identify
   a) Desired power
   b) Priors for Δ, Σ (indexed by θ, τ, α, ß), using elicitation and/or historical data
2. Calculate
   a) S_b^2 (blinded sample variance)
   b) $E(\Sigma + k\,\Delta^2 \mid S_b^2)$ (if time allows)
3. For each candidate *n* (end-of-study sample size per group), calculate PredictivePower(*n*), integrating ConditionalPower(*n*, Δ, Σ) over parameter space subject to

$$\Sigma + k\,\Delta^2 = E(\Sigma + k\,\Delta^2 \mid S_b^2) \quad \text{or } \Sigma + k\,\Delta^2 = S_b^2$$

4. Select *n* that achieves the predictive power closest to that desired.

Generalization: Comparing ≥3 Treatments

- Standard 1-way ANOVA
- Model: $y_{ij} = \mu + \alpha_i + \varepsilon_{ij}$
  - $\varepsilon_{ij} \sim N(0, \Sigma)$ (same Σ for all i)
  - r treatment groups, indexed by i
  - n independent measurements per group at End of Study, indexed by j
- $\alpha_1 + \alpha_2 + \ldots + \alpha_r = 0$ (to guarantee unique solution to normal equations)
- H_o: $\alpha_i = 0 \ \forall \ i$
- Standard frequentist test statistic is

$$F = \frac{MSTr}{MSE}$$

- $F \sim F_{df(Tr),df(Error),\lambda}$ where

$$\lambda = \text{ noncentrality} = \frac{n}{\Sigma} \sum_{i=1}^{r} \alpha_i^{\,2}$$

  - Standard test of H_o requires comparing F to upper quantile of $c \equiv F_{df(Tr),df(Error),0}$

Generalization: Comparing ≥3 Treatments (con't)

So, at the interim analysis, if $\theta \equiv [\Sigma, \sum_{i=1}^{r} \alpha_i{}^2]$ were known, power conditional on θ could be given as

$$\Pr(F \geq c \mid Data, \theta)$$

More realistically (unconditionally), <u>predictive</u> power is

$$\Pr(F \geq c \mid Data) = \int [ConditionalPower(\theta \mid Data)] \pi(\theta \mid Data) \partial\theta =$$

$$\int \Pr(F \geq c \mid \Sigma, \sum_{i=1}^{r} \alpha_i^2) \times$$

$$\pi\left(\Sigma, \sum_{i=1}^{r} \alpha_i{}^2 \mid Data\right) \partial\left(\Sigma, \sum_{i=1}^{r} \alpha_i{}^2\right)$$

⇒ We want to find joint <u>posterior</u> PDF of

$$[\Sigma, \sum_{i=1}^{r} \alpha_i{}^2 \mid Data]$$

Generalization: Comparing ≥3 Treatments (con't)

- At interim, with *m* observations per group, (Blinded) total mean sum of squares is

$$\text{MSTo} = \frac{\sum_{ij}(y_{ij}-\bar{y})^2}{rm-1}.$$

MSTo has conditional expectation

$$E(\text{MSTo} \mid \Sigma, \sum_{i=1}^{r} \alpha_i{}^2) = \Sigma + \frac{m\sum_{i=1}^{r} \alpha_i{}^2}{rm-1}$$

⇒ Use MSTo

1. along with $E(\Sigma + \frac{m\sum_{i=1}^{r} \alpha_i{}^2}{rm-1})$ to estimate $\Sigma + \frac{m\sum_{i=1}^{r} \alpha_i{}^2}{rm-1}$ (as a weighted average)
2. to update joint prior PDF of $[\Sigma, \sum_{i=1}^{r} \alpha_i{}^2]$

Then, for each candidate *n*, integrate $Pr(F \geq c \mid \Sigma, \sum_{i=1}^{r} \alpha_i{}^2)$ over joint PDF of $[\Sigma, \sum_{i=1}^{r} \alpha_i{}^2 \mid MSTo]$, finding *n* that provides desired predictive power

Comparing Blinded Sample Size Adjustment Methods

- Alice came to a fork in the road. "Which road do I take?" she asked.
- "Where do you want to go?" responded the Cheshire Cat.
- "I don't know," Alice answered.
- "Then," said the cat, "it doesn't matter."

  - Lewis Carroll, Alice in Wonderland

$\Rightarrow$ Choices between methods are arbitrary, without a **loss function**

Comparing Blinded SSA Methods

- Objective: Minimize expected loss ("risk"), taking expectations over parameters *(Δ, Σ)* & data *(S_b^2)*
- One measure of risk:

MMAD =

$Mean_{Data}$[$Mean_{\theta}$(Absolute Deviation from Targeted Power|Data)]

$$= \int_0^{\infty} \int_{\Theta} | Power(\delta, s_b^{\,2}, N) - 0.9 | \, \pi_{\Delta}\left(\delta \,|\, s_b^{\,2}\right) \partial\delta \, f_{S_b^{\,2}}\left(s_b^{\,2}\right) \partial s_b^{\,2}$$

for N= sample size per treatment group

Most **established blinded** methods lead to similar sample size adjustments & (hence) similar MMADs, so method of Gould-Shih (StatMed, 1998) used as representing those.

Over a range of situations, the proposed method reduces MMAD 15% to 27% compared to those established methods…

Comparing G&S Procedure (N′) vs. Proposed Procedure (N″) - MMAD

- 10,000 simulations for each combination of
  - ✓ E(Δ)
  - ✓ Var(Δ)
  - ✓ E(Σ)
  - ✓ Var(Σ)
- MMAD(N″)/MMAD(N′)
  - ➢ Small Ratio ⇒ N″ provides an advantage
  - ➢ Found to lie between 0.73 & 0.85 for all combinations studied*
  - ➢ Smallest when Var(Σ)/Var(Δ) small
- Evidence of small
  - α inflation
  - bias of sample mean treatment difference

- *N″ does <u>much</u> better than N′ in minimizing mean-mean <u>squared</u> deviation (MMAD=Mean-mean absolute deviation from 90% Conditional Power)

Comparing G&S Procedure (N′) vs. Proposed Procedure (N″) - rNPV

- Objectives of pharmaceutical industry
  - Provide medicines that allow patients to
    - Live longer
    - Function more fully
    - Feel better (Medical Care)
  - Advance knowledge of human body (Science)
  - Reward researchers & investors ($ Profit)
- A simple measure of expected profit for a clinical trial is

"risk-adjusted Net Present Value" (rNPV) =

[Payoff upon Trial Success][Probability of Success]

- Sampling Cost

where, in simple situations, [Probability of Success] can be approximated by predictive power,

$$\int [ConditionalPower(\theta)]\pi(\theta|Data)\partial\theta$$

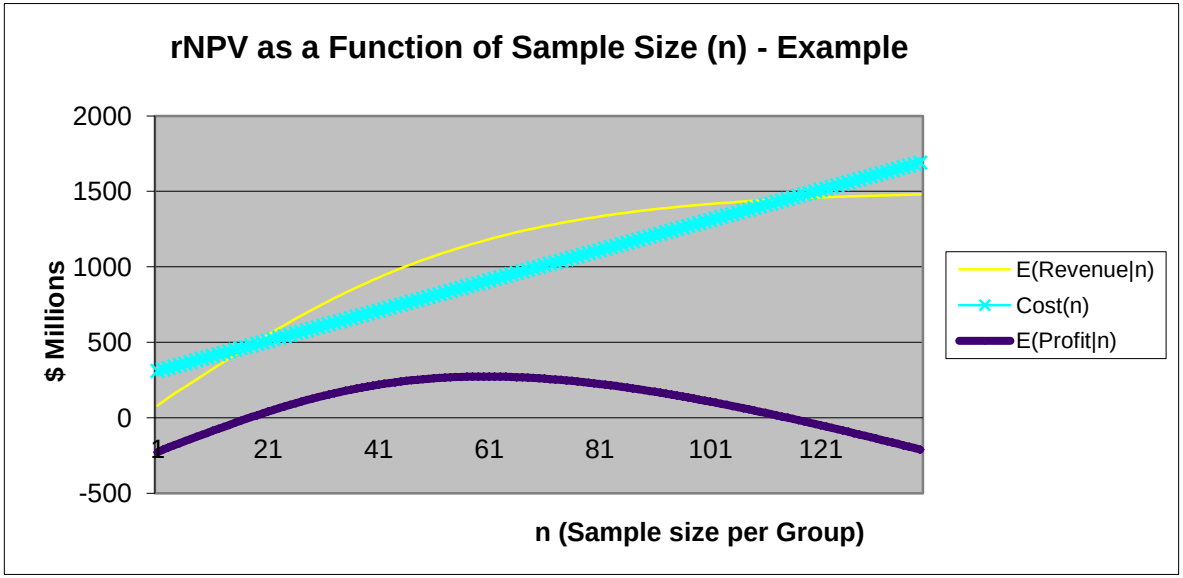
Comparing G&S Procedure (N′) vs. Proposed Procedure (N″) – rNPV (con't)

- That simple measure, rNPV, can be adapted to more complex situations by incorporating
  - Impacts of estimated treatment effect (mean treatment difference, hazard ratio, etc.) on sales forecast
  - Influences of additional trials on marketing approval
  - Delays in marketing approval due to larger sample sizes
  - Discounting cash flow ($1 spent now is worth more than $1 earned later)
  - Real Options…
- So, rNPV may be a satisfactory metric for comparing SSA procedures

Comparing G&S Procedure (N′) vs. Proposed Procedure (N″) – rNPV (con't)

rNPV for each method (N″ & N′): For each candidate *n*, calculate

- Unblinded probability of success (predictive power)
- Estimated predictive power (varies between blinded methods)
- Payoff given trial success
  - In simplest situation, constant WRT *n*
  - However, may decrease due to erosion of patent life & discounting
- Sampling cost – Increases linearly with *n*
- *n* maximizing rNPV



Comparing G&S Procedure (N′) vs. Proposed Procedure (N″) – rNPV (con't)

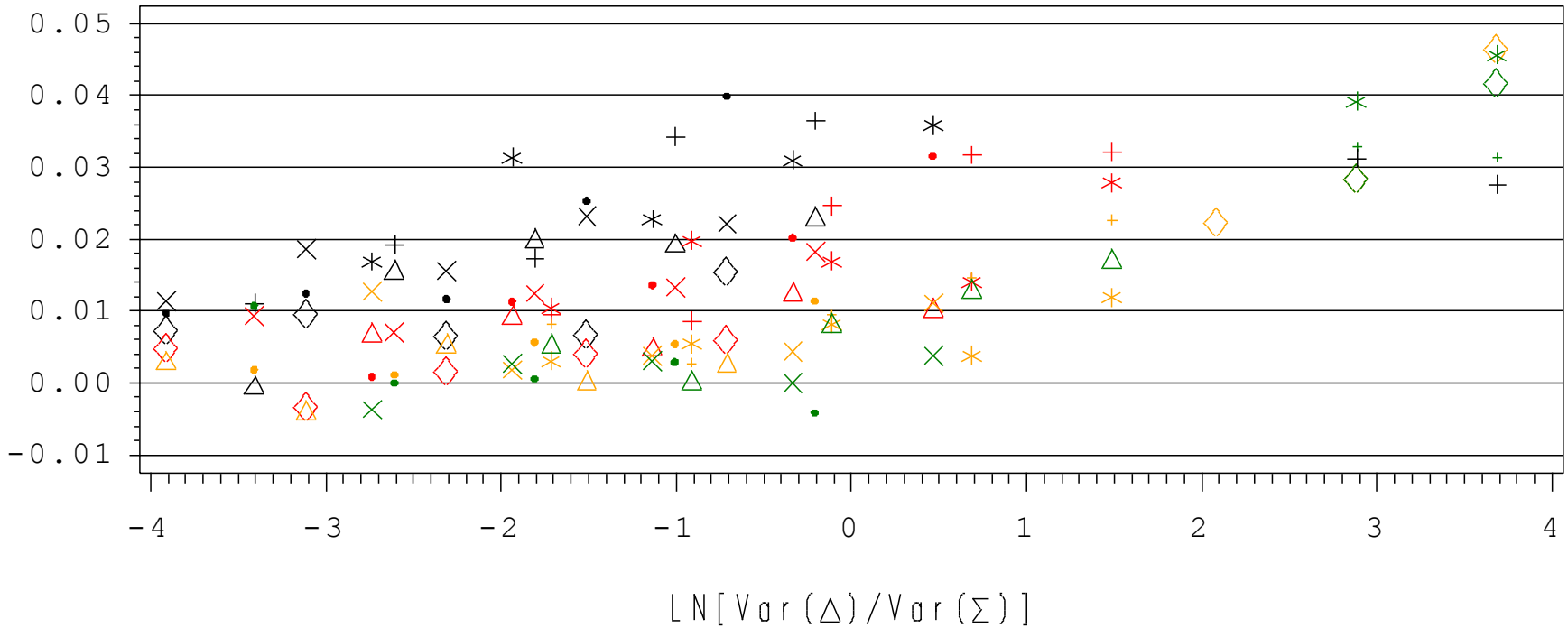
- N″ increases rNPV for *almost all* combinations (α,ß,θ,τ)
- Further investigation needed to ascertain conditions under which each is superior

- On following slide
  - Graph of

$$\frac{[rNPV(\mathrm{N}'')-rNPV(\mathrm{N}')]}{abs(rNPV(\mathrm{N}'))}$$

  - For 125 combinations of hyperparameters (α,ß,θ,τ)

Gain over GS2 Method, on LN[Var(Δ)/Var(Σ)] and CV(Δ)



CV(Δ)

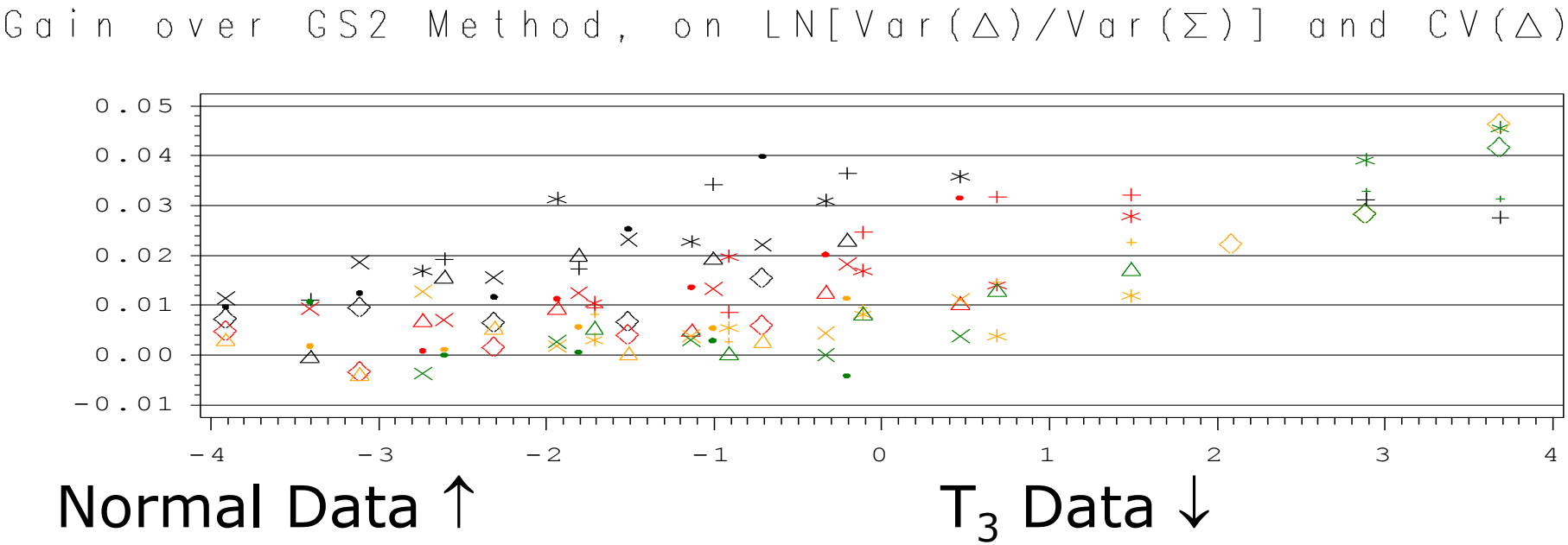
● ● ● 0.372	+ + + 0.429	× × × 0.479	* * * 0.524
△ △ △ 0.552	◇ ◇ ◇ 0.670	● ● ● 0.673	+ + + 0.733
× × × 0.773	* * * 0.942	△ △ △ 0.943	◇ ◇ ◇ 1.117
● ● ● 1.288	+ + + 1.319	× × × 1.571	* * * 2.198
△ △ △ 3.352	◇ ◇ ◇ 3.514	● ● ● 3.863	+ + + 4.518
× × × 4.714	* * * 6.325	△ △ △ 6.594	◇ ◇ ◇ 10.541
+ + + 31.623			

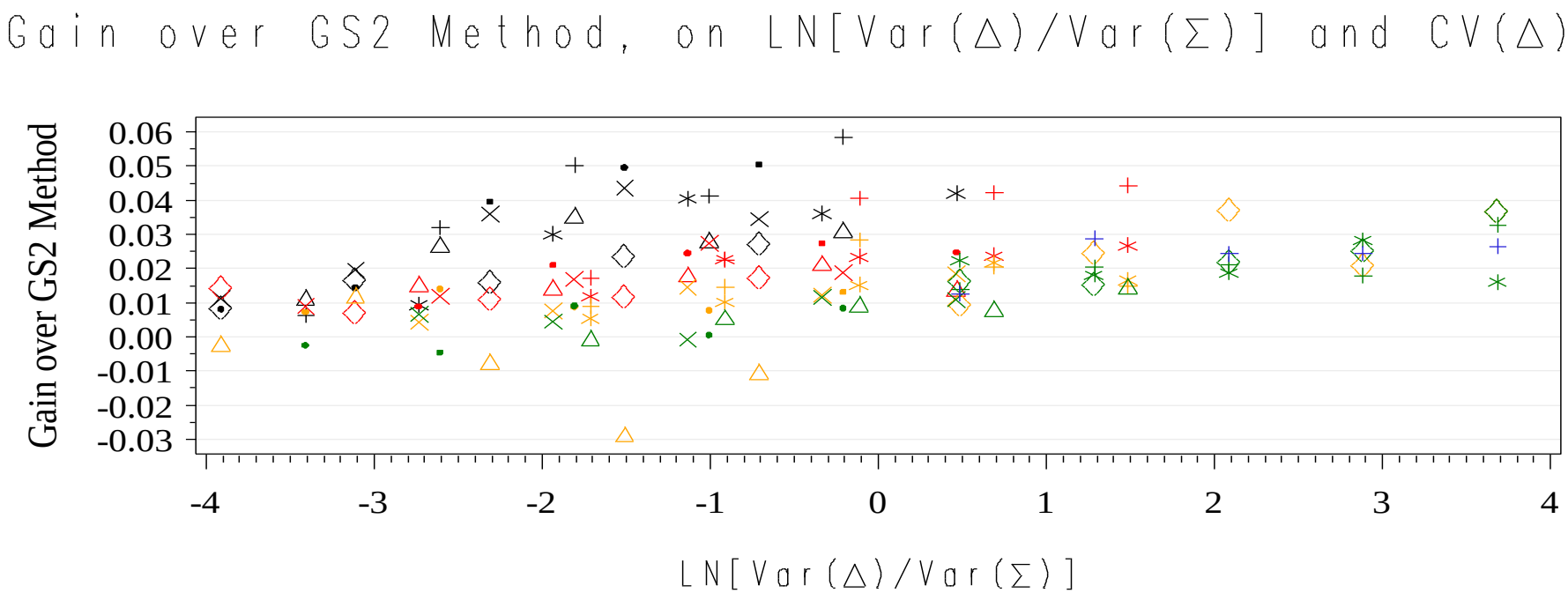
Sensitivity Analysis: t-distribution of Data

- Q: Does proposed PP method improve rNPV vis-à-vis GS method, if data are T_3 (i.e., 3 dof)?
  - Note if dof=1 or 2 then variance not well-defined
  - rNPV simulations of previous slides repeated
  - Extreme outliers occurred in raw data, causing computation problems for both GS & new SSA methods. Therefore, S_b^2 limited to 10 times its approx. expectation:

$$S_b^2 \leq 10E(\Sigma + k\Delta^2) = 10[\alpha\beta + k(\theta^2 + \tau^2)]$$

Next: Comparison of gains over GS method, for Normal data vs. T_3 data.


Gain over GS2 Method, on LN[Var(△)/Var(Σ)] and CV(△)
0.05
0.04
0.03
0.02
0.01
0.00
-0.01
-4
-3
-2
-1
0
1
2
3
4
Normal Data ↑
T3 Data ↓


Gain over GS2 Method, on LN[Var(△)/Var(Σ)] and CV(△)
Gain over GS2 Method
0.06
0.05
0.04
0.03
0.02
0.01
0.00
-0.01
-0.02
-0.03
-4
-3
-2
-1
0
1
2
3
4
LN[Var(△)/Var(Σ)]

Summary

- **Unblinded** SSA methods - logistically challenging, requires statistical adjustments
- Established **blinded** SSA methods - Useful only when Δ already estimated precisely
- Proposed **blinded** SSA method
  - Useful when both Δ and Σ are highly uncertain
  - Formally incorporates prior information
  - Appropriately adjusts PDFs of Δ and Σ
  - Generalizable to comparisons of ≥3 treatments
  - Compared to established blinded methods
    - Reduces mean-mean-absolute-deviation from targeted power
    - Almost always increases rNPV, for normal & t-distributed data
- For further information:
  - Andrew.Hartley@PPDI.com, 910-558-7147