An Evaluation of Inferential Procedures for Adaptive Clinical Trial Designs with Pre-specified Rules for Modifying the Sample Size

Gregory P. Levin, Sarah C. Emerson, & Scott S. Emerson

Cesar Torres

May 8, 2014

Motivating Example

Suppose researchers want to cure [insert type of cancer here], because said cancer is bad.

- Two-arm clinical trial
- Can observe X_{plac_i}'s and X_{treat_i}'s
  - $X_{plac_i} \overset{\text{iid}}{\sim} \mathcal{N}\left(\mu_{plac}, \sigma^2\right)$
  - $X_{treat_i} \overset{\text{iid}}{\sim} \mathcal{N}\left(\mu_{treat}, \sigma^2\right)$
  - $\sigma^2 > 0$ known
- Defining $\theta := \mu_{treat} - \mu_{plac}$, interested in testing $H_0 : \theta \leq 0$ vs. $H_1 : \theta > 0$
- Issue of concern: lots of treatments to evaluate

Possible designs for clinical trial

- "Well-understood" designs
  - Fixed design
  - Group sequential design
- "Less-well-understood" designs
  - Designs that adapt sample size based on interim-effect size estimates

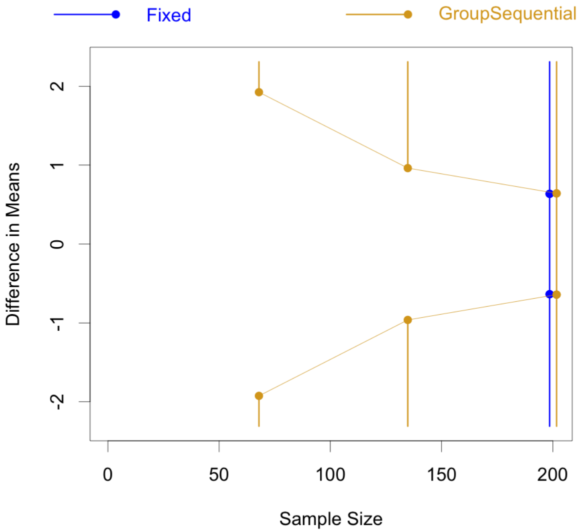
Fixed Design (old and boring)

- Prespecified sample size, decision rule
- Pros
  - Setup easy to understand
  - Easy calculations
  - Properties of $\hat{\theta} := \overline{X}_{treat} - \overline{X}_{plac}$ well understood
- Cons
  - Can be considered inefficient/unethical

Group Sequential Design (old and boring)

- J total analyses, with $J > 1$
- Decision rule at j^{th} analysis based on observed $\hat{\theta}_j$
- For some boundaries $a_j \leq d_j$,
  - If $j < J$, stop the trial and reject H_0 if $\hat{\theta}_j \geq d_j$, stop the trial and fail to reject H_0 if $\hat{\theta}_j \leq a_j$, and continue on with the trial otherwise
  - If $j = J$ (at the final analysis), stop the trial and reject H_0 if $\hat{\theta}_j \geq d_j$, stop the trial and fail to reject H_0 otherwise

Picture Similar to the one from Last Time



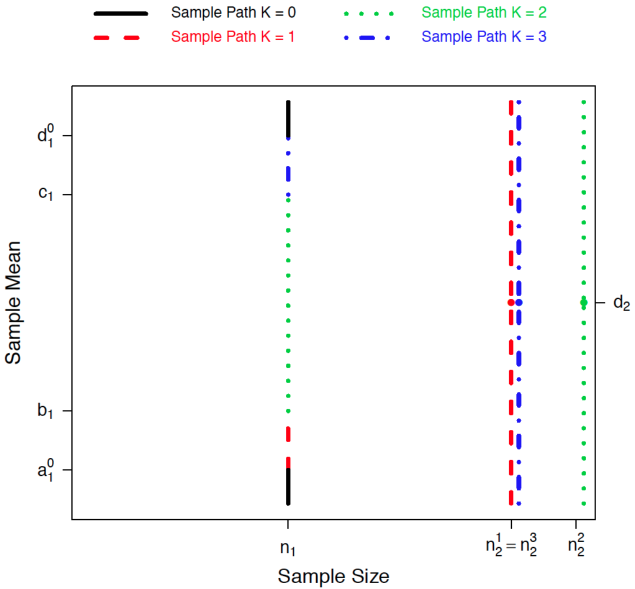
Group Sequential Design (old and boring)

- Prespecified decision rule for each analysis, maximum sample size
- Pros
  - Setup easy to understand
  - Doable calculations
  - Properties of $\hat{\theta}$ numerically derivable
    - Recursively done by noting that
      $f_n(\hat{\theta}|\theta) = \int_{a_{n-1}}^{d_{n-1}} f_{n-1}(\hat{\theta}|\theta) \frac{1}{\sqrt{2\pi}} e^{-\frac{\left(\hat{\theta}-\theta\right)^2}{2}} d\theta \times \mathbb{1}_{\{a_n \leq \hat{\theta} \leq d_n\}}$
- Cons
  - Underpowered for some values of theta

The Third Design (New and Exciting!!)

- Similar to group sequential design
- Adaption occurs at interim analysis time $j = h$
  - For $j \in \{1, 2, \ldots, h-1\}$, essentially the same as regular group sequential design
- Adaptation based on $\hat{\theta}_h$, determining future analysis times and boundaries
- Boundaries for this design are combination of boundaries of two group sequential designs (according to Scott Emerson)

Same Picture as Last Time



perhaps lifted from the paper...

Inference when using Group-Sequential-like Designs

- Neyman-Pearson lemma:
  If testing $H_0 : \theta = \theta_0$ vs. $H_1 : \theta = \theta_1 \neq \theta_0$, likelihood ratio test most powerful test of size α
- Karlin-Rubin theorem (extension of Neyman-Pearson lemma):
  If likelihood ratio is monotone non-decreasing in θ, then likelihood ratio test also most powerful for testing $H_0 : \theta \leq \theta_0$ vs. $H_1 : \theta > \theta_0$
- Issue: likelihood ratio not monotone non-decreasing when using group-sequential-like designs
  - Need some way (some ordering) to determine what are "extreme" observations under the null hypothesis

Considered Orderings

Three orderings focused on in paper:

- Sample mean
- Signed LR: If $\forall$ fixed θ^*,

$$sign\left(\hat{\theta}_{(1)} - \theta^*\right) \frac{f\left(\text{outcome } 1 | \theta = \hat{\theta}_{(1)}\right)}{f(\text{outcome } 1 | \theta = \theta^*)} > sign\left(\hat{\theta}_{(2)} - \theta^*\right) \frac{f\left(\text{outcome } 2 | \theta = \hat{\theta}_{(2)}\right)}{f(\text{outcome } 2 | \theta = \theta^*)},$$

  then outcome 1 ordered higher than outcome 2, with $\hat{\theta}_{(i)}$ the sample mean from outcome i
- Conditional Error Ordering: Outcomes ordered according to the stage-wise p-value of "backward image"

After selecting ordering, p-values and confidence interval can be derived.

Inference (Point Estimates)

Three point estimates considered

- Sample mean $\hat{\theta}$
- Bias adjusted mean $\hat{\eta}$
  - $\hat{\eta}$ satisfies $E(\hat{\theta}|\theta = \hat{\eta}) = \theta$
- Median-unbiased estimate $\hat{\zeta}$
  - Given the observed outcome, and an ordering, $\hat{\zeta}$ satisfies $P\Big(\text{observed} \succ \text{all outcomes}|\theta = \hat{\zeta}\Big) = \frac{1}{2}$

Aim of Paper

Evaluate by simulation the behavior of third design, under different scenarios

In particular, looking at

- Coverage probabilities and average length of confidence intervals
- Performance of point estimates and p-values

Varying parameters:

- Many

Challenges

- Unsure of how to combine stopping boundaries of two group sequential designs to define stopping boundaries of third design
- Unsure of how to code up orderings

Summary

- Want to cure cancer
- Have several choices for design of randomized clinical trial
  - Fixed design and group sequential design "well-understood"
  - Group sequential design with one sample size adaptation, not so much
    - Paper investigates performance of this last design via simulation study
- Conceptual/coding errors in the way of cure

What's next?

- Beg for help
- Run simulations for days

Questions?