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

April 22, 2014

Why choose this paper?

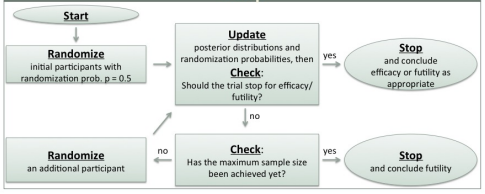
Estimation following a Bayesian adaptive design

Cesar Torres[1], Peter Müller[2] and Lurdes Inoue[1]
[1]Department of Biostatistics, University of Washington
[2]Department of Mathematics, University of Texas at Austin

Introduction

- Adaptive designs allow for some aspect of the design to change during the trial depending on interim outcomes.
  - There are both Bayesian and frequentist adaptive designs.
  - Examples: adaptive randomization; sequential stopping; arm additions and deletions; adaptive dose-finding
- Some concern that treatment effect estimates are biased using these adaptive methods

Setup



Methods

- Two Bayesian adaptive ideas implemented
  - Stopping early based on posterior distributions
  - Adaptive randomization based on posterior distributions
- Hypotheses: $H_0: \theta_0 + \delta \geq \theta_1$ vs. $H_1: \theta_0 + \delta < \theta_1$
- Outcomes: Two cases considered: Binary outcome, or normally distributed outcome (with variance 1)
- Relatively non-informative priors assumed for each of the two cases considered (at first)
  - Same prior for each arm in the clinical trial
    - Binary outcome case: Uniform(0, 1), or equivalently, Beta(1, 1)
    - Continuous outcome: N(0, 1)
- Evaluate $d = P(\theta_0 + \delta < \theta_1 \mid data)$
  - Stop for efficacy (reject null hypothesis) if d > some efficacy threshold
  - Stop for futility (fail to reject null hypothesis) if d < some futility threshold
  - Thresholds chosen to get acceptable type I and type II errors
    - Determined through simulations
- Randomization probabilities:
  - $p_0 = \frac{\left(1 - P(\theta_0 < \theta_1 \mid data)\right)^c}{\left(P(\theta_0 < \theta_1 \mid data)\right)^c + \left(1 - P(\theta_0 < \theta_1 \mid data)\right)^c}$
  - $c = \frac{n}{2N}$
  - $p_1 = 1 - p_0$
- 1000 simulations for each scenario
- Fixed θ_0 for binary scenarios and continuous scenarios separately
  - Binary: $\theta_0 = 0.2$
  - Continuous: $\theta_0 = 0$

Simulation Results

Table 1a. Results for binary outcome.

θ_1	MSS[1]	Initial	Under Null Hypothesis: Treatment effect (mean)	Under Null Hypothesis: Treatment effect (median)	Under Null Hypothesis: Treatment effect (mode)	Under Alternative Hypothesis: Treatment effect (mean)	Under Alternative Hypothesis: Treatment effect (median)	Under Alternative Hypothesis: Treatment effect (mode)
0.6	100	10	-0.03 (-0.38, 0.38)	-0.03 (-0.41, 0.41)	-0.04 (-0.50, 0.50)	0.35 (-0.14, 0.58)	0.38 (-0.16, 0.64)	0.47 (-0.22, 0.89)
		20	0.00 (-0.29, 0.29)	0.00 (-0.30, 0.31)	0.00 (-0.35, 0.34)	0.34 (-0.01, 0.64)	0.36 (-0.01, 0.68)	0.41 (-0.01, 0.78)
		30	0.00 (-0.24, 0.24)	0.00 (-0.25, 0.25)	0.00 (-0.28, 0.29)	0.36 (0.06, 0.61)	0.37 (0.06, 0.64)	0.41 (0.06, 0.70)
		100	0.00 (-0.15, 0.14)	0.00 (-0.15, 0.14)	0.00 (-0.15, 0.15)	0.39 (0.21, 0.55)	0.39 (0.22, 0.56)	0.40 (0.22, 0.58)
	200	10	-0.03 (-0.38, 0.38)	-0.03 (-0.41, 0.41)	-0.04 (-0.50, 0.50)	0.34 (-0.14, 0.58)	0.37 (-0.16, 0.64)	0.47 (-0.22, 0.89)
		20	0.00 (-0.29, 0.29)	0.00 (-0.30, 0.31)	0.00 (-0.35, 0.34)	0.34 (-0.01, 0.64)	0.36 (-0.01, 0.68)	0.41 (-0.01, 0.78)
		30	0.00 (-0.24, 0.24)	0.00 (-0.25, 0.25)	0.00 (-0.28, 0.29)	0.36 (0.06, 0.61)	0.37 (0.06, 0.64)	0.41 (0.06, 0.70)
		200	0.00 (-0.10, 0.10)	0.00 (-0.11, 0.10)	0.00 (-0.11, 0.10)	0.39 (0.27, 0.51)	0.40 (0.27, 0.52)	0.40 (0.28, 0.53)
0.8	100	10	0.00 (-0.38, 0.38)	0.00 (-0.41, 0.41)	0.00 (-0.50, 0.50)	0.43 (0.00, 0.71)	0.47 (0.00, 0.78)	0.62 (0.00, 1.00)
		20	0.00 (-0.29, 0.29)	0.00 (-0.30, 0.30)	0.00 (-0.35, 0.34)	0.50 (0.17, 0.76)	0.53 (0.18, 0.8)	0.61 (0.20, 0.92)
		30	0.00 (-0.24, 0.24)	0.00 (-0.25, 0.25)	0.00 (-0.28, 0.29)	0.53 (0.25, 0.76)	0.56 (0.26, 0.79)	0.61 (0.29, 0.87)
		100	0.00 (-0.15, 0.14)	0.00 (-0.15, 0.14)	0.00 (-0.15, 0.15)	0.58 (0.42, 0.72)	0.59 (0.43, 0.73)	0.60 (0.44, 0.75)
	200	10	0.00 (-0.38, 0.38)	0.00 (-0.41, 0.41)	0.00 (-0.50, 0.50)	0.43 (0.00, 0.71)	0.47 (0.00, 0.78)	0.62 (0.00, 1.00)
		20	0.00 (-0.29, 0.29)	0.00 (-0.30, 0.30)	0.00 (-0.35, 0.34)	0.50 (0.17, 0.76)	0.53 (0.18, 0.80)	0.61 (0.20, 0.92)
		30	0.00 (-0.24, 0.24)	0.00 (-0.25, 0.25)	0.00 (-0.28, 0.29)	0.53 (0.25, 0.76)	0.56 (0.26, 0.79)	0.61 (0.29, 0.87)
		200	0.00 (-0.10, 0.10)	0.00 (-0.11, 0.10)	0.00 (-0.11, 0.10)	0.59 (0.48, 0.70)	0.59 (0.48, 0.70)	0.60 (0.49, 0.71)

Table 1b. Results for continuous outcome.

θ_1	MSS[1]	Initial	Under Null Hypothesis: Treatment effect (mean)	Under Alternative Hypothesis: Treatment effect (mean)
1	300	10	-0.12 (-1.01, 1.16)	1.08 (-0.23, 1.91)
		20	0.01 (-0.82, 0.89)	0.94 (0.09, 1.76)
		30	0.01 (-0.71, 0.71)	0.95 (0.23, 1.64)
		300	0.00 (-0.24, 0.24)	0.99 (0.76, 1.23)
1.5		10	0.01 (-1.01, 1.14)	1.28 (0.21, 2.31)
		20	0.02 (-0.82, 0.86)	1.38 (0.54, 2.22)
		30	0.01 (-0.71, 0.71)	1.42 (0.70, 2.11)
		300	0.00 (-0.24, 0.24)	1.49 (1.25, 1.73)

[1] Maximum Sample Size

Table 2. Results for different prior distributions, under the conditions of the first rows for tables 1a and 1b.

Data type	Prior dist.	Under Null Hypothesis: Treatment effect (mean)	Under Null Hypothesis: Treatment effect (median)	Under Null Hypothesis: Treatment effect (mode)	Under Alternative Hypothesis: Treatment effect (mean)	Under Alternative Hypothesis: Treatment effect (median)	Under Alternative Hypothesis: Treatment effect (mode)
binary	Beta(2.5, 2.5)	-0.02 (-0.27, 0.27)	-0.02 (-0.29, 0.29)	-0.03 (-0.33, 0.33)	0.25 (-0.10, 0.41)	0.27 (-0.11, 0.43)	0.31 (-0.12, 0.51)
	Beta(5.5, 5.5)	-0.01 (-0.18, 0.19)	-0.01 (-0.18, 0.20)	-0.02 (-0.20, 0.22)	0.17 (-0.06, 0.28)	0.18 (-0.07, 0.28)	0.19 (-0.07, 0.29)
	Beta(11.5, 11.5)	-0.01 (-0.1, 0.12)	-0.01 (-0.11, 0.13)	-0.01 (-0.11, 0.13)	0.11 (-0.04, 0.18)	0.11 (-0.04, 0.18)	0.12 (-0.04, 0.19)
continuous	N(0, 0.5)	-0.10 (-0.85, 0.99)	-	-	0.95 (-0.19, 1.62)	-	-
	N(0, 0.25)	-0.05 (-0.65, 0.79)	-	-	0.74 (-0.15, 1.26)	-	-
	N(0, 0.125)	-0.02 (-0.45, 0.59)	-	-	0.53 (-0.11, 0.89)	-	-

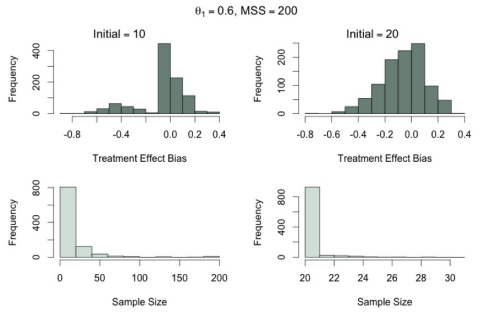


Figure 1. Histograms of treatment effect (mean) bias and sample sizes, under select scenarios.

Discussion

- Modest bias when assuming prior distributions that are not highly informative
- Bias minimally affected by maximum number of patients, when using prior distributions that are not highly informative
- Bias decreased when using larger number of initial patients, using prior distributions that are not highly informative
- Some bias introduced under more informative prior distributions
  - Bias for both arms tending toward the mean of prior distributions

Conclusions

- Bias minimized when using priors that are not highly informative, and increasing initial number of patients
- Potentially poor treatment effect estimates under informative ('biased') priors when trials stop early with a very small number of patients

Future Work

- Introduction of different prior distributions that allow for shrinkage of the bias
- Investigation of sensitivity of results to violations of assumptions
- Investigation of estimation following other popular adaptive designs (k-arm, CRM, TITE-CRM, etc)

Contact

Cesar Torres
Email: cdtorres@uw.edu

References

1. Berry, Scott M., Bradley P. Carlin, J. Jack Lee, and Peter Müller. *Bayesian Adaptive Methods for Clinical Trials.* Boca Raton: CRC, 2011. Print.

Summer project, completed under the guidance of Lurdes Inoue

Surprisingly, did not win any of the 'Best Poster' awards

What are adaptive clinical trials?

Adaptive clinical trials modify trial plans based on interim results.

Compared to clinical trials with fixed sample designs (and the same operating characteristics), adaptive trials

- Typically have higher maximum sample size
- But achieve lower average sample size
  - Save time
  - Save money
  - Save participants

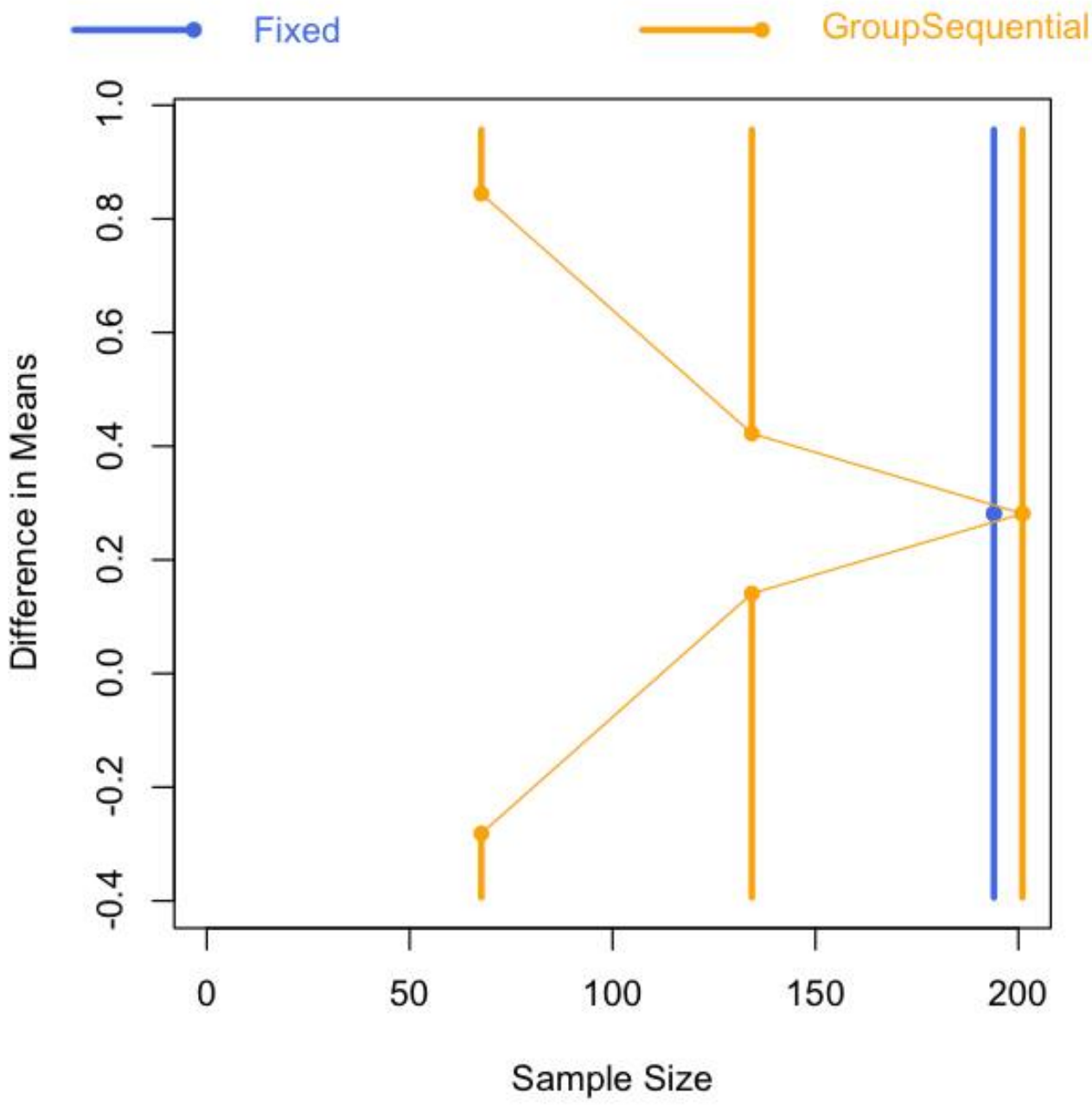
Ok, fine. But what are the specifics of the setup?

- Three designs mentioned
  - Fixed sample
  - Group sequential
  - Group sequential with one "adaptation" (will elaborate on this in a bit)
- Parameter of interest: $\theta = \mu_{\text{treatment}} - \mu_{\text{placebo}}$
- Interest in testing, for example, $H_0 : \theta \leq 0$ vs $H_1 : \theta > 0$

...group...sequential?

- J interim analyses
- Statistic T_j based on the data observed up until j^{th} analysis (interim or final)
- For some boundaries $a_j \leq d_j$,
  - If $j \leq J$, stop the trial and reject H_0 if $T_j \geq d_j$, stop the trial and fail to reject H_0 if $T_j \leq a_j$, and continue on with the trial otherwise
  - If $j = J + 1$ (at the final analysis), stop the trial and reject H_0 if $T_j \geq d_j$, stop the trial and fail to reject H_0 otherwise

Whoah, buddy! Where's the picture?

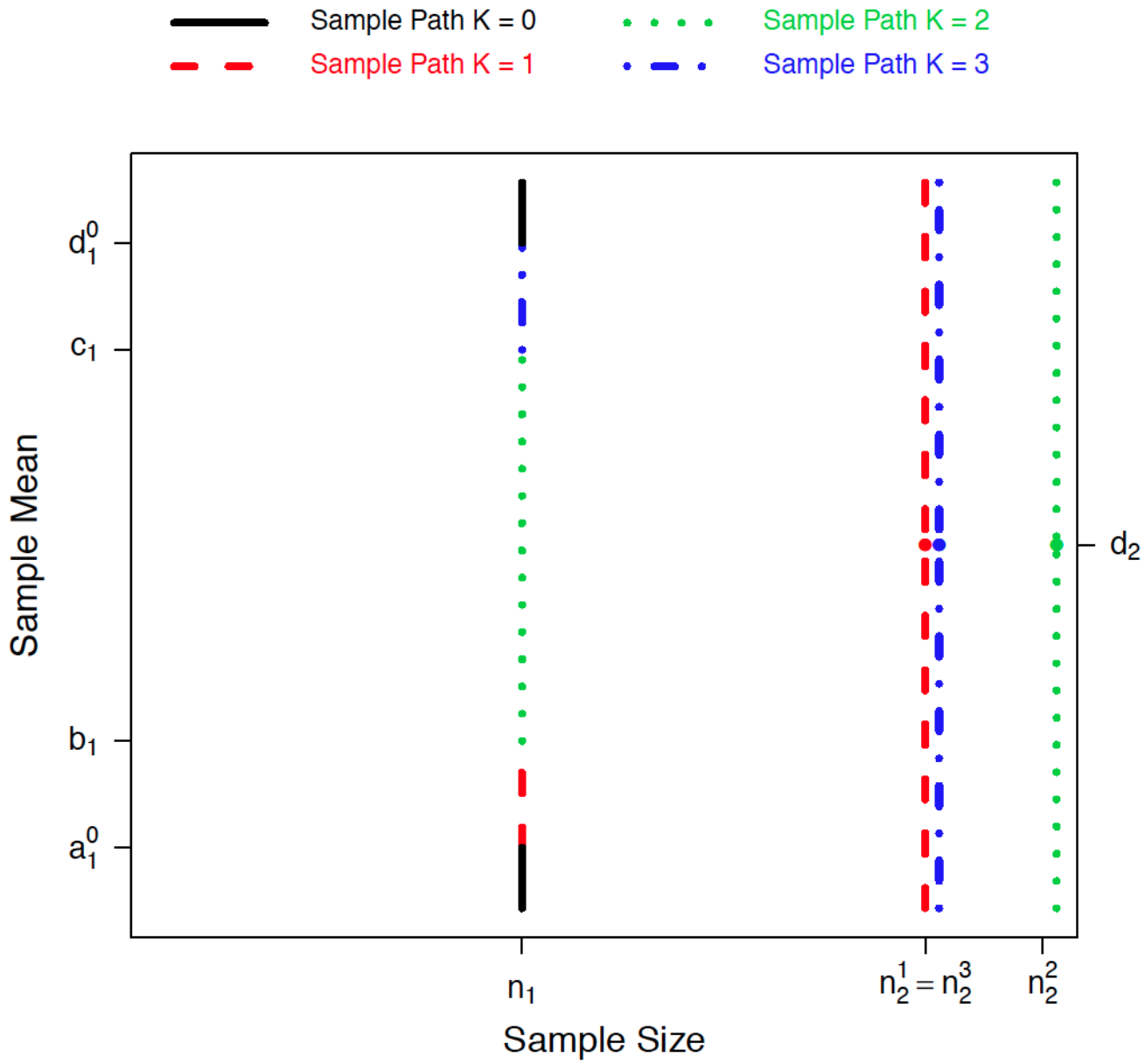


How about that promised elaboration?

Group Sequential Design with Adaptation

- Similar to group sequential design with no adaptation
- Adaptation occurs at interim analysis time $j = h$
  - For $j \in \{1, 2, ..., h-1\}$, essentially the same as with no adaptation
- Based on T_h, determine future analysis times and boundaries

Picture, please!



totally not lifted from the paper...

What about inference?

Recall that $H_0 : \theta \leq 0$, and $H_1 : \theta > 0$

$(1 - \alpha) \times 100\%$ Confidence intervals for θ

- Invert hypothesis test with type I error probability α
- Define acceptance region of "non-extreme" results for the test statistic
  - Fixed sample design
    - Neyman-Pearson lemma, Karlin-Rubin theorem applicable
  - Group sequential design, with or without adaptation
    - Likelihood ratio not monotone, so Neyman-Pearson lemma, Karlin-Rubin theorem not applicable
    - Need some ordering of sample space to determine "extreme" values

What orderings?

Three orderings focused on in paper

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

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

  then outcome 1 ordered higher than outcome 2, with $t^{(i)}$ the sample mean from outcome i
- Conditional Error Ordering: ???

What about the point estimates and p-values?

Three point estimates considered

- Sample mean $\hat{\theta}$ (MLE)
- Bias adjusted mean $\check{\theta}$
  - Whitehead (1986)
- Median-unbiased estimate $\tilde{\theta}$
  - Given the observed outcome, and an ordering, $\tilde{\theta}$ satisfies $P\left(\text{observed} \succ \text{all outcomes} | \theta = \tilde{\theta}\right) = \frac{1}{2}$

Given an ordering, upper one-sided p-value calculated as
p-value $= P(\text{observed} \succ \text{all outcomes} | \theta = 0)$

So...what do the authors do in the paper?

Evaluate by simulation the behavior of group sequential designs with one sample size adaptation, under different scenarios
In particular, looking at

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

Varying parameters:

- Type of stopping boundaries
- Power, at some clinically meaningful effect size $\theta = \Delta$
- Maximum number of interim analyses J
- Timing of the adaptation
- Maximum allowable sample size
- Rule for determining sample size
- True θ

What's next?

- do some serious background reading
- run simulations for days

Questions?

(Please be gentle)