Interim Analyses and Sequential Testing in Clinical Trials

Nicole Solomon
BIOS 790

November 12, 2015

Clinical Trials

- Goal: compare two clinical therapies or treatments on a clinically relevant endpoint
- Statistical test utilized depends on the endpoint
- Standard approach: single-stage design
  - Shortcomings:
    - Ethical concerns
    - Financial concerns
  - Alternative solution: interim analyses that allow for early termination

Repeated testing

- Clinical trial design to test $H_0 : \delta = 0$ vs $H_1 : \delta \neq 0$:
  - clinically meaningful difference δ
  - Power $= 1 - \beta$
  - α
  - N
- Single-stage trial: one analysis $\leftrightarrow$ all alpha spent
- Multiple testing inflates the actual Type I error rate as each analysis contributes substantially to the total probability of committing a Type I error
  - Armitage et al. (1969): $\alpha = 0.05 \rightarrow \alpha \approx 0.14$
  - Solution: Sequential testing methods

Motivating example: BHAT

Beta-blocker Heart Attack Trial (BHAT): effectiveness of beta-blocker drug in reducing mortality in patients who had recently suffered an MI.

- 2-sided $\alpha = 0.05$
- Power $= 90\%$
- 20% reduction in 3-year mortality
- Target N $= 4000$
- Endpoint: overall survival
- Sequential test design: OBF, 7 analyses, 6 months apart

Motivating example: BHAT analyses

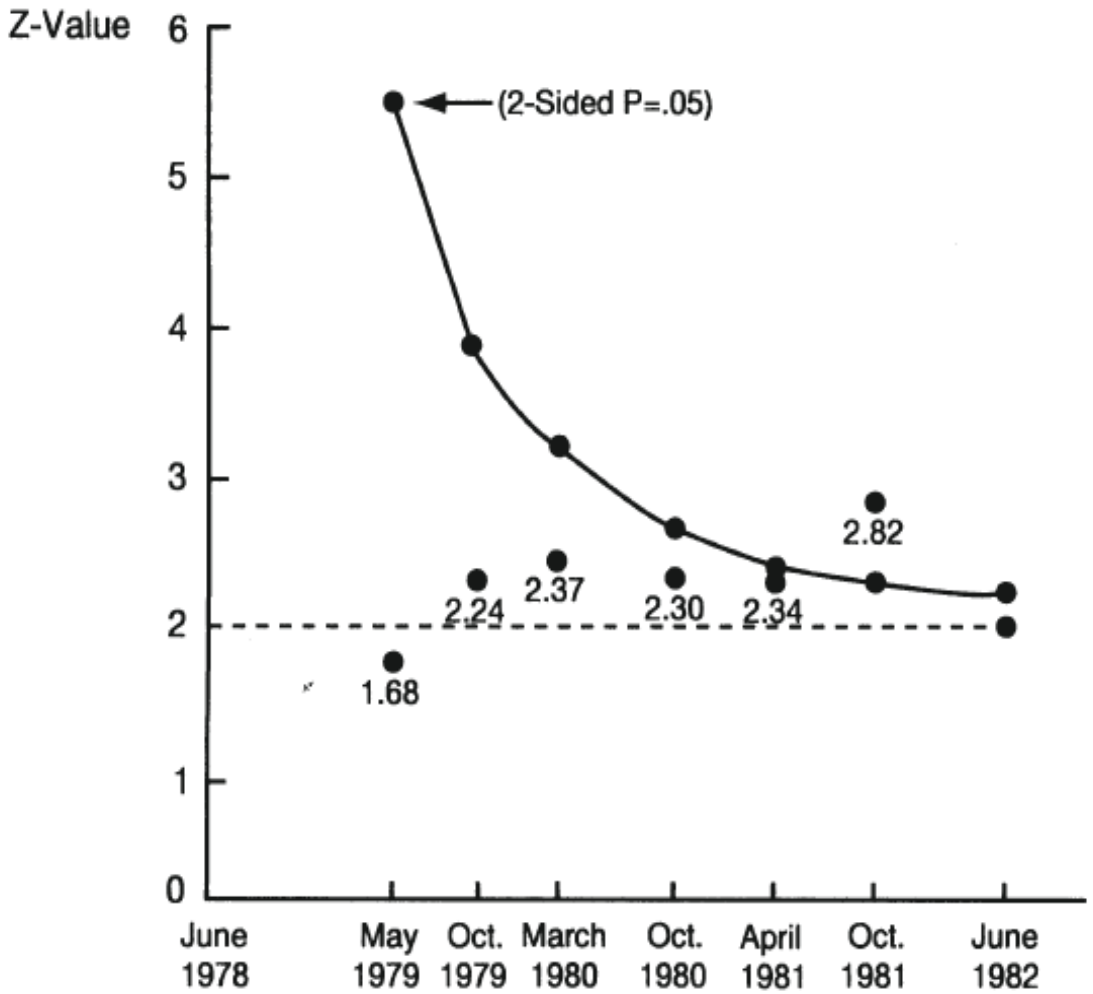


Fig. 16.8 Six interim log-rank statistics plotted for the time of data monitoring committee meetings with a two-sided O'Brien–Fleming significance level boundary in the Beta-Blocker Heart Attack Trial. *Dashed line* represents $Z=1.96$ [74]

Motivating example: BHAT sequential tests

The standard OBF sequential testing procedure was applied for BHAT.

Limitations:

- Necessary to pre-specify number of interim analyses
- Requirement for equal number of events between each analysis (pre-specified timing)

Group sequential methods

1. *Classical* sequential methods are executed after each pair of participants entered the study.
2. *Group* sequential methods are executed on entire groups of participants.
    1. Motivation: monitor the accumulating data so as to potentially terminate early
    2. Key feature: define unique significance levels at each interim analysis so as to maintain the overall Type I error rate

Group sequential features

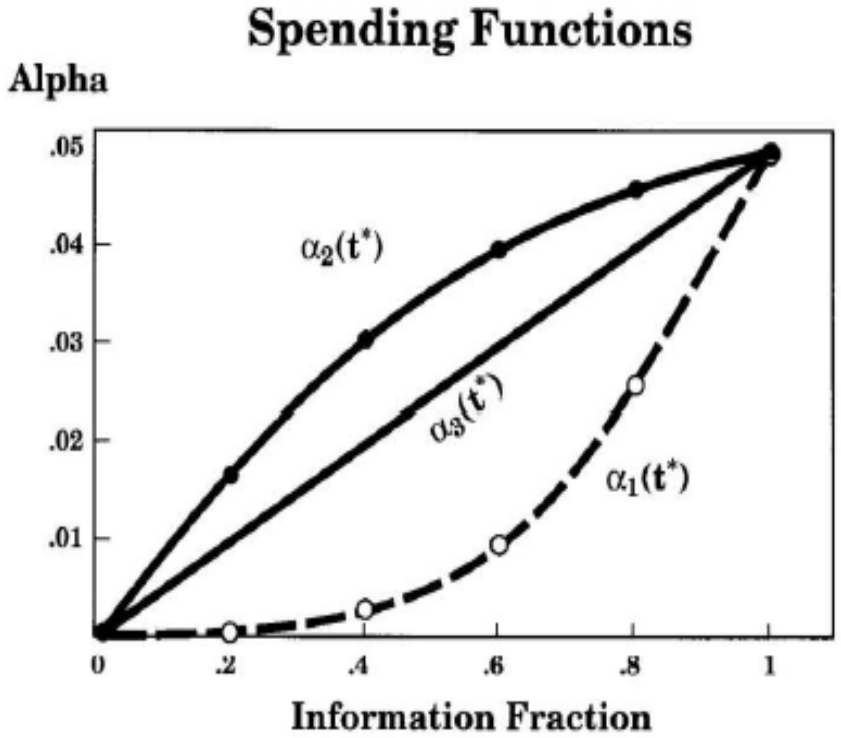
- Strengths of GS design:
  - possiblility of early stopping under H_1
    - maintain Type I error rates at nominal design-stage values
    - reduce trial sample size N
  - simple:
    - process depends only on K total number of analyses
    - standard test statistics are employed
- Weaknesses:
  - do not permit early stopping under H_0
    - may have a larger maximum N than a single-stage design (Pocock)

Common group sequential methods

- Requirements of GS methods:
  - specification of total number of analyses
  - specification of timing of analyses
- Common procedures:
  - O'Brien & Fleming (OBF) (1979)
    - conservative; increasing expenditure
  - Pocock (1977)
    - aggressive; decreasing expenditure

Group sequential behaviors

- Alpha spending behaviors of classical methods:
  - OBF: increasing significance levels $(\alpha_1(t^*))$
    - $\alpha^*_{1:5} = 5e^{-5}, 0.004, 0.012, 0.025, 0.04$
  - Pocock: constant significance levels $(\alpha_2(t^*))$
    - $\alpha^*_{1:5} = 0.016, 0.016, 0.016, 0.016, 0.016$



Flexible group sequential methods

- A critical weakness of classical GS methods is the requirement to pre-specify
  - total number of analyses
  - timing of the analyses
- Flexible GS methods:
  - Lan & DeMets (LDM) (1983)
  - Wang & Tsiatis (WT) (1987)
- Strength of flexible methods:
  - interim analyses can be performed at any time and as frequent as desired (abuse)
  - additional alpha control strategies possible
  - how?
    - alpha spending functions

Alpha spending notation

Notation:

- K: total number of analyses desired; $k = 1, \ldots, K$
- T: max length of the trial
- t_k: time of k^{th} analysis; $0 < t_1 < t_2 < \ldots < t_K = T$
- τ_k: fraction of expected information observed at time t_k
- $\alpha^*(\tau_k)$: significance level applied at k^{th} analysis
- $Z_C(k)$: boundary statistic utilized at k^{th} analysis
- $Z(k)$: test statistic computed on all available data at k^{th} analysis

Alpha spending functions

- Alpha spending function $\alpha^*(\tau)$ defines how much alpha is spent at a fraction τ of the total information expected in the study
  - $0 = \alpha^*(0) < \alpha^*(\tau_1) < \ldots < \alpha^*(t_K) = \alpha$
  - $\sum_{k=1}^{K}\{\alpha^*(\tau_k) - \alpha^*(\tau_{k-1})\} = \alpha$
- approximate OBF & Pocock
  - OBF: $\alpha^*(\tau) = 2 * [1 - \Phi(z_{\alpha/2}/\sqrt{\tau})]$
  - Pocock: $\alpha^*(\tau) = \alpha * \log(1 + (\exp -1)^{\tau})$
- Information fraction may be approximated by
  - quantitative endpoint: $\tau_k \approx n_k/N$
  - censored endpoint: $\tau_k \approx d_k/D$

LDM design process

1. Choose $\alpha^*(\tau)$ and specify δ, σ^2, α, β
2. Fix K and define τ_k at equally spaced increments
3. Calculate $Z_C(k)$ for each k based on $\alpha, \alpha^*(\tau)$
4. Determine Δ, the hypothetical center of the test statistic's distribution under H_1 such that power $= 1 - \beta$
5. Compute total sample size N

BHAT with alpha spending

More complicated computation of the boundary values would have been necessary to account for the fact that the timing and event count requirements were not strictly met.

Had the alpha spending procedure been applied to BHAT with the OBF-equivalent function the final results would not have changed:

Analysis	Z	OBF Z_c	d/D	MI Z_c	t/T	MD Z_c
1	1.68	5.45	0.14	5.88	0.23	4.53
2	2.24	3.83	0.19	5.04	0.33	3.73
3	2.37	3.30	0.32	3.79	0.43	3.24
4	2.30	2.80	0.44	3.19	0.58	2.74
5	2.34	2.40	0.62	2.64	0.70	2.49
6	2.82	2.29	0.82	2.30	0.83	2.27

Early stopping under H_0

Methods which permit early stopping under H_0: (Pampallona, 1994)

- Beta-spending procedure
  - defined comparably to $\alpha^*(\tau)$; controls spending of β
  - Boundary procedure
  - boundary choice depends on both α and β

Strengths: permit early stopping under either H_1 or H_0

Weaknesses:

- increased max sample size vs single-stage design (Pocock)
- less tractable mathematically

Design points

- Choice of sequential design procedure
- Sidedness of testing
  - research question
  - expected treatment effect
- Allowance for early stopping under H_0
  - more precise estimate of treatment effect
  - greater power

Precautions

- Reduction in N only possible for quick endpoints
- Avoid unplanned interim analyses, especially data-driven
  - Assumption: analysis times are independent of 'behavior of data'
- Interim result is not the only stopping criteria
  - accrual rate
  - SAE rates
  - external related publications

SMART

- Dynamic Treatment Regimes (DTR) are treatments tailored to individuals
  - How to customize these? Treatment order, timing of adjustments, etc?
- SMART = sequential multiple assignment randomized trial
  - Goal: inform construction of DTR
  - multi-stage trial where each stage involves decision (and randomization) on intensity modification and treatment type
  - Structure:
    - Initial trt $\rightarrow$ intermediate outcome $\rightarrow$ secondary trt

Sequential Testing References

Armitage P, McPherson CK, Rowe BC (1969). Repeated significance testing on accumulating data. *JRSS Series A*, **132**: 235-44.

Friedman L, Furberg C, DeMets D. (2010). *Fundamentals of Clinical Trials* (4th ed). New York: Springer.

Lan K & De Mets D. (1983). Discrete sequential boundaries for clinical trials. *Biometrika*, **70**: 659-63.

O'Brien P & Fleming T. (1979). A multiple testing procedure for clinical trials. *Biometrics*, **35**: 549-56.

Pampallona S & Tsiatis A. (1994). Group sequential designs for one-sided and two-sided hypothesis testing with provision for early stopping in favor of the null hypothesis. *J Stat Plann Inf*, **42**: 19-35.

Pocock S. (1977). Group sequential methods in the design and analysis of clinical trials. *Biometrika*, **64**: 191-9.

Sebille V and Bellisant E. (2003). Sequential methodology for phase III trials. *Fundamental & Clin Pharm*, **17**: 505-16.

SMART References

Murphy S. (2007). SMART experimental designs for developing dynamic treatment regimes.
http://dept.stat.lsa.umich.edu/ samurphy/seminars/?C=M;O=D.

Adaptive Interventions. (2015). Retrieved November 8, 2015, from https://methodology.psu.edu/ra/adap-inter