Introduction to the Design and Evaluation of Group Sequential Clinical Trials

Session 1 - Scientific Setting

Presented July 26, 2017

Daniel L. Gillen
Department of Statistics
University of California, Irvine

John M. Kittelson
Department of Biostatistics & Informatics
University of Colorado Denver

Overview

Module 9: Introduction to the Design and Evaluation of Group Sequential Clinical Trials

Session 1: Scientific setting
Session 2: Fixed-sample design
Session 3: Evaluation of group sequential clinical trial designs
Session 4: Bayesian evaluation of group sequential clinical trial designs

Clinical Trials: Definition

A clinical trial is

- A planned experiment which involves patients that is designed to elucidate the most appropriate treatment of future patients. (Pocock, 1983)
- A planned experiment designed to assess the efficacy of a treatment in humans by comparing the outcomes in a group of patients treated with those observed in a comparable group of patients receiving a control treatment, where patients in both groups are enrolled, treated, and followed over the same time period. (Minert, 1986)

Clinical Trials: Definition

Types of questions that can be evaluated in clinical trials

- Therapeutic intervention studies:
  - Safety: Is risk of treatment-related toxicities suitably low?
  - Efficacy: Treatment benefits on disease processes?
  - Effectiveness: Does the treatment offer benefits when used as part of standard routine practice?
- Non-therapeutic intervention studies:
  - Mechanistic studies: Studies of drug mechanism of action.
  - Behaviorial interventions: Examples: Smoking cessation; diabetes prevention.
  - Prevention studies: Examples: Women's Health Initiative (HRT for prevention of cardiovascular disease); lung cancer screening trial.
  - Community intervention studies: Interventions on schools to promote healthy lifestyles.
- Our focus in this course is on medical intervention studies.

Clinical Trials: Definition

Clinical trials as experiments

- As scientific experiments clinical trials must:
  - Answer a scientifically meaningful questions.
    - Must discriminate between viable hypotheses
  - Provide results that inform (convince) medical practice.
    - Use valid materials and methods
    - Use valid measurement of the experimental outcome
    - Provide a valid quantification of uncertainty in the experiment.

Clinical Trials: Definition

As experiments on humans clinical trials must:

- Be ethically justifiable for the individuals entering the trial:
  - As much as possible, minimize harm and maximize benefit for individuals in the trial.
  - Avoid giving individual participants harmful treatments.
  - Avoid giving individual participants inferior treatments.
- Maintain the ethical responsiveness to all likely future recipients of the therapy under evaluation:
  - Identify (and approve) new beneficial therapies.
  - Avoid approving ineffective or harmful treatments.
  - Avoid unnecessary delays in the evaluation process.

Public Heath Objective in Clinical Trials

Public Health Objective

- Evidence-based medicine requires:
  - Results in the sample reflect effects in standard practice.
    - Does the trial population reflect the target population?
    - Do diagnostic procedures reflect standard practice?
    - Does ancillary/rescue therapy reflect standard practice?
- (Reiterating) Our ultimate goal should be to:
  - Identify (and approve) new beneficial therapies.
  - Avoid approving ineffective or harmful treatments.

The Public Health Objective

Our objective is to have trials with high positive predictive value

- Positive predictive value (PPV):
  - *Diagnostic testing:* prevalence of diseased individuals among those with a positive diagnostic test.
  - *Clinical trials:* prevalence of truly beneficial therapies among those which are identified by a positive clinical trial.
  - PPV is calculated using Bayes rule:

$$\begin{aligned} PPV &= \frac{\beta\pi}{\beta\pi + \alpha(1-\pi)} \\ \textit{where}: & \\ \beta &= \textit{sensitivity} \\ 1-\alpha &= \textit{specificity} \\ \pi &= \textit{prevalence} \end{aligned}$$

The Public Health Objective

Clinical trials as diagnostic tests

- The statistical hypothesis test is a diagnostic test for beneficial treatments.
  - α-level: probability of observing a positive (statistically significant) test in absence of a true treatment effect:
    - Level of significance is 1 - specificity.
    - Choosing $\alpha = 0.05$ gives 95% specificity.
  - Statistical power (β): Probability of observing a positive (statistically significant) test when there is a true treatment effect:
    - Power is sensitivity.
    - 80% sensitivity is a common (though not ideal) choice.
  - Prevalence (π_0): the percentage of effective treatments among all tested treatments.
  - Positive predictive value: probability that a statistically significant trial indicates a truly effective treatment.

$$PPV = \frac{\beta\pi_0}{\beta\pi_0 + \alpha(1 - \pi_0)}$$

The Public Health Objective: How does the design determine PPV?

PPV is increased through good experimental practice

$$PPV = \frac{\beta\pi_0}{\beta\pi_0 + \alpha(1-\pi_0)}$$

* Increase π_0:
  - Careful planning of preliminary studies
  - Avoid "novel" and "innovative" ideas
  - Careful specification of hypothesis-driven research
* Increase β:
  - Good practice (no missing data, low variation in outcome assessment, good adherence, etc.)
  - Increase sample size.
* Reduce α:
  - Pre-specify outcomes
  - Pre-specify all analyses
  - Avoid multiple comparisons
  - Avoid surrogate outcomes.
  - Avoid subgroups

The Public Health Objective: How does the design determine PPV?

References: PPV as a function of π_0, α, and β

** *EfficiencyForTargetedTX.pdf* discusses effect of power (sensitivity) and α-level (specificity) on the PPV of phase II/III clinical trials.

Scenario	Number of trials	π_0	Phase II trials α_2	Phase II trials β_2	Phase III trials α_3	Phase III trials β_3	Pos	Pos	PPV
1	1000	0.10	*	*	0.05	0.975	98	45	0.685
2	12500	0.10	0.05	0.15	0.05	0.800	150	28	0.842
3	11765	0.20	0.05	0.15	0.05	0.800	282	24	0.923
4	13245	0.01	0.05	0.15	0.05	0.800	16	33	0.327
5	9091	0.10	0.05	0.15	0.05	0.975	133	20	0.867
6	15385	0.10	0.05	0.15	0.05	0.500	115	35	0.769
7	6780	0.10	0.20	0.15	0.05	0.800	81	61	0.571
8	6780	0.10	0.20	0.15	0.10	0.800	81	122	0.400

***Evaluation of Strategies for the Phase II to Phase III Progression in Treatment Discovery*:

(Sanchez, 2014) http://rctdesign.org/TechReports/SanchezThesis201404.pdf

The Public Health Objective
How do clinical trials determine PPV?

Summary remarks

- A wide range of situations/therapies are studied in trials.
- Globally, clinical trials need to assure:
  - Scientific credibility
  - Ethical experiments
  - Efficient experiments:
    - Minimize time
    - Minimal number of extra subjects
    - Minimize cost
  - A high prevalence of truly beneficial therapies among all therapies used in routine care.

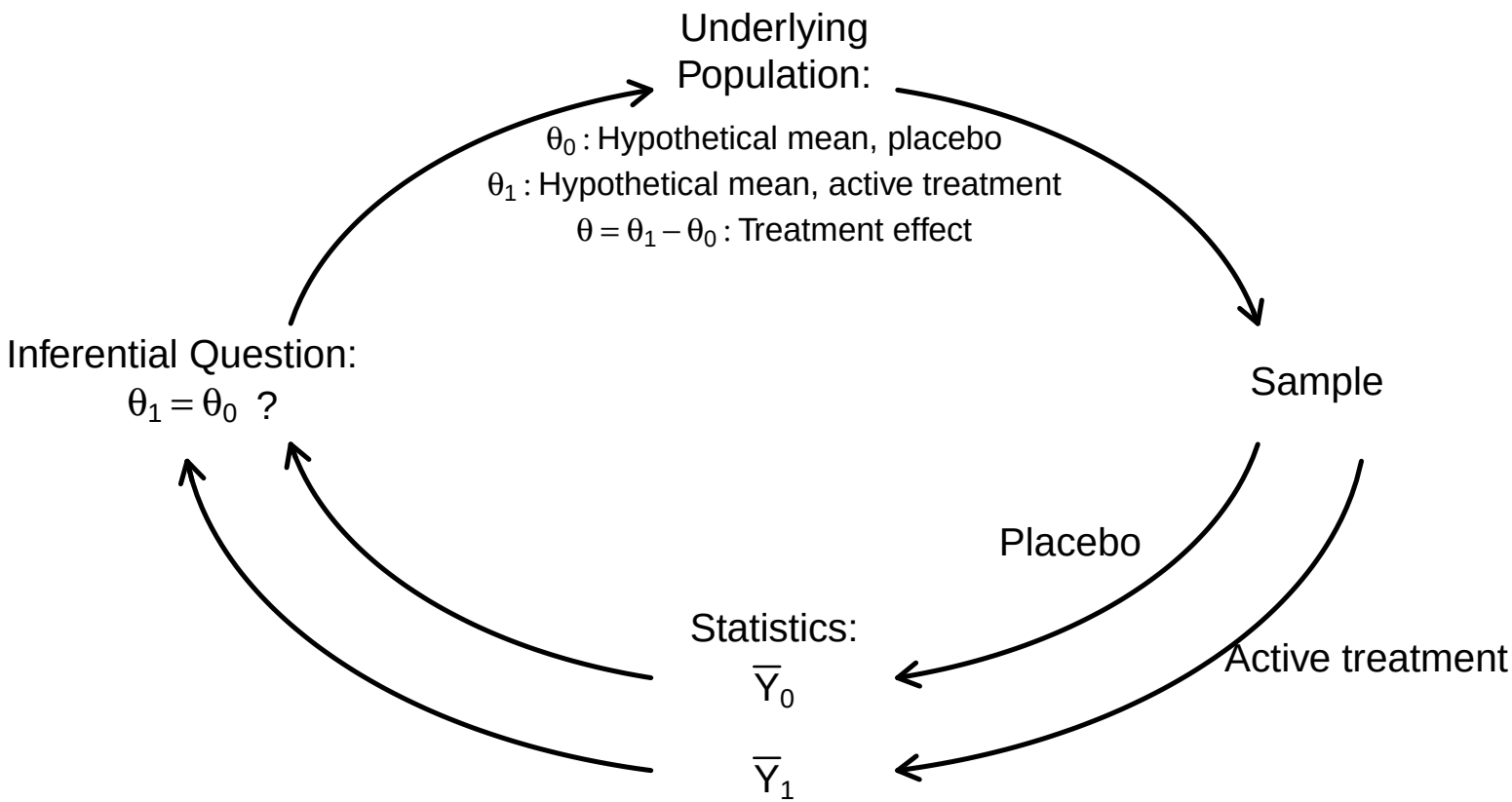
Statistical foundations

Evidence-based practice

- As a scientific experiment, the results of a clinical trial are used to rule out (or rule in) hypotheses about treatment effects.
- The standards for rejecting (or accepting hypotheses) are based on statistical criteria.
- Clinical trial designs should be evaluated by the potential inference upon trial completion.

Recall Empirical Objective:

Use observed trial result ($\hat{\theta}$) to make inference about underlying population θ



Statistical foundations

Four main inferential elements

1. Point estimate: $\hat{\theta}$ is the “best" estimate of θ.

2. Interval estimate: Values of θ that are consistent with the trial results.

3. Expression of uncertainty (p-value): To what degree is a particular hypothesis (the “null" hypothesis) consistent with the observed trial results?

4. Decision: Based on the above measures, what decision should be reached about the use of a new therapy?

Trial monitoring

Why monitor a clinical trial?

- Monitoring for quality control; for example,
  - Patient accrual.
  - Data quality/completeness.
  - Unanticipated adverse events.
- Monitoring study endpoint(s); for example,
  - Treatment benefits.
  - Toxicity differences.
- Good quality control should be part of every study to ensure that the study achieves its goals.
- Monitoring study endpoints is not applicable in every study, and requires special statistical methods to avoid increased statistical errors.

Elements of trial monitoring

Reasons to monitor trial endpoints

- To maintain the validity of the informed consent for:
  - Subjects currently enrolled in the study.
  - New subjects entering the study.
- To ensure the ethics of randomization.
  - Randomization is only ethical under equipoise.
  - If there is not equipoise, then the trial should stop.
- To identify the best treatment as quickly as possible:
  - For the benefit of all patients (i.e., so that the best treatment becomes standard practice).
  - For the benefit of study participants (i.e., so that participants are not given inferior therapies for any longer than necessary).

Elements of trial monitoring

Monitoring trial endpoints

Monitoring endpoints must be done properly to avoid bias:

- Data driven analyses cause bias:
  - Analyzing study results because they look good leads to an overestimate of treatment benefits.
- Publication/presentation of 'preliminary results' can affect:
  - Ability to accrue subjects.
  - Type of subjects that are referred and accrued.
  - Treatment of patients not in the study.
- Failure to design for interim analyses can lead to hasty decisions subject to:
  - Inadequate consideration of trade-offs between competing endpoints (toxicity versus benefit).
  - External pressures from study investigators or sponsors.
  - Lack of objectivity by study monitors.

Elements of trial monitoring

Monitoring trial endpoints

Thus:

- Monitoring of study endpoints is often required for ethical reasons.
- Monitoring of study endpoints must carefully planned as part of study design to:
  - Avoid bias
  - Assure careful decisions
  - Maintain desired statistical properties

Elements of trial monitoring

How are trials monitored?

- “Data Safety and Monitoring Boards (DSMB)" are used to avoid biased decisions:
  - DSMB members are *independent* of the study investigators
  - The DSMB reviews unblinded data in the midst of a trial to:
    - Assure the trial is safe to continue.
    - Make decisions about early termination based on the statistical monitoring plan (“group-sequential clinical trial design").
- DSMB composition:
  - Subject-matter specialists (2-4)
  - Biostatistician (1-2).

Elements of trial monitoring

Trial monitoring plan

Trial monitoring plan is typically pre-specified in 2 documents:

- DSMB charter:
  - Defines scope of trial monitoring
  - Defines DSMB responsibilities
  - Defines sponsor responsibilities
  - Pre-specifies monitoring plans and decisions (reasons for stopping)
- Interim Statistical Analysis Plan (ISAP):
  - Defines monitoring endpoint(s)
  - Pre-specifies analysis timing, decision criteria, and rationale
  - Pre-specifies methods for implementation (changes to analysis timing)
  - Pre-specifies adjustments to statistical inference about treatment effects

Monitoring the primary endpoint: Illustration of statistical implications

Illustration setting (trial design):

Consider a clinical trial evaluating *superiority* of a new agent:

- Measure of treatment effect ($\theta = \theta_1 - \theta_0$) defined based on fixed-sample design:
  - Primary endpoint
  - Probability model
  - Functional
  - Contrast
  - Statistical hypotheses
  - Statistical standards for decisions (i..e., frequentist or Bayes)
- Suppose large values of θ denote superiority of the new agent.

Monitoring the primary endpoint: Illustration of statistical implications

Illustration setting (interim analysis plan):

- Suppose that the trial is planned with interim analyses after outcomes are measured on $N_1 < N_2 < ... < N_J$ participants.
- Let $\hat{\theta}_j$ denote the estimated treatment effect at the jth analysis ($j = 1, ..., J$).
- Consider stopping criteria $a_j < d_j$ with:

$$\begin{aligned} \hat{\theta}_j \geq d_j &\Rightarrow \text{Decide new treatment is superior} \\ \hat{\theta}_j \leq a_j &\Rightarrow \text{Decide new treatment is not superior} \\ a_j < \hat{\theta}_j < d_j &\Rightarrow \text{Continue trial} \end{aligned}$$

Set $a_J = d_J$ so that the trial stops by the Jth analysis.

Illustration of statistical implications

Example: O'Brien-Fleming (OBF) 2-sided design

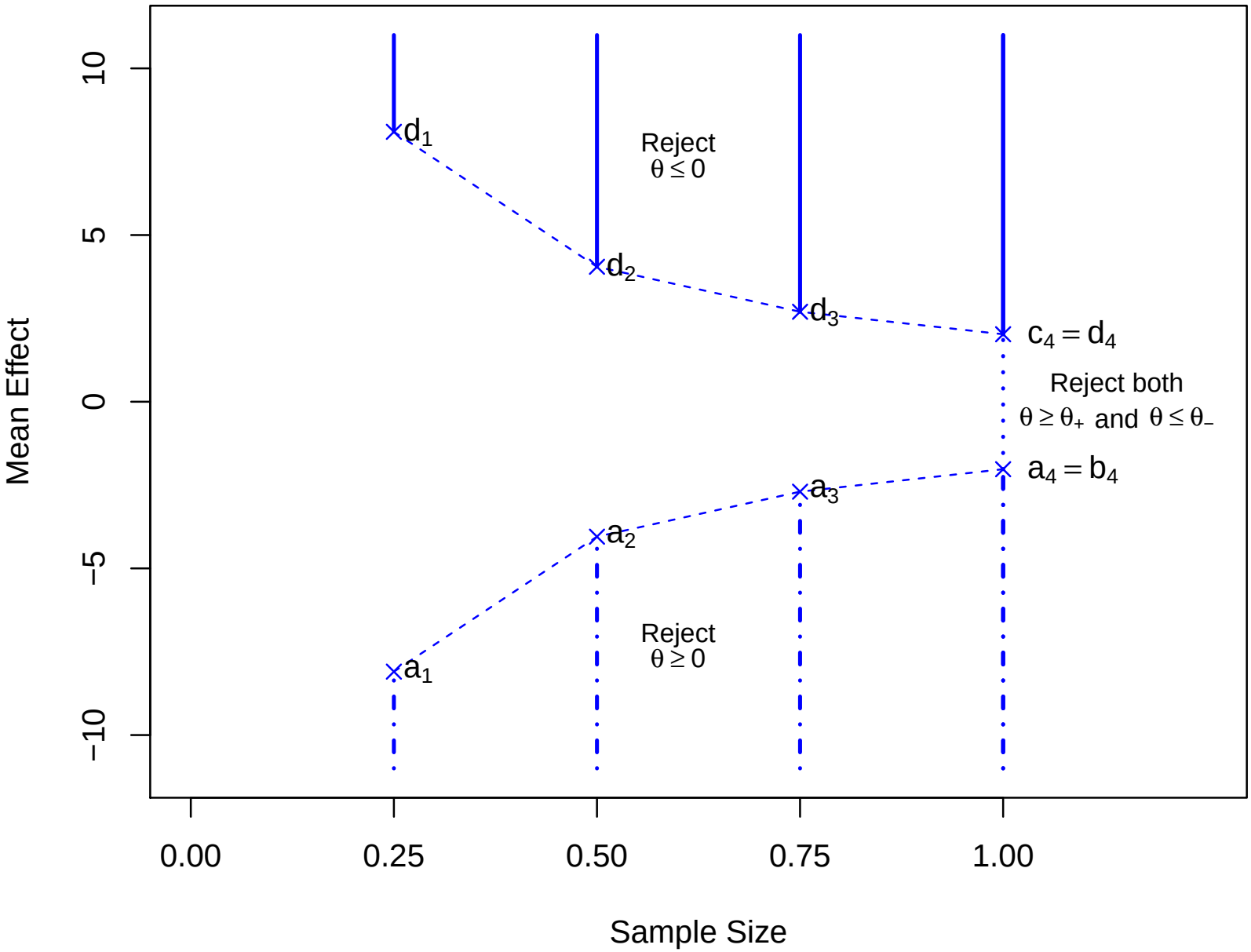
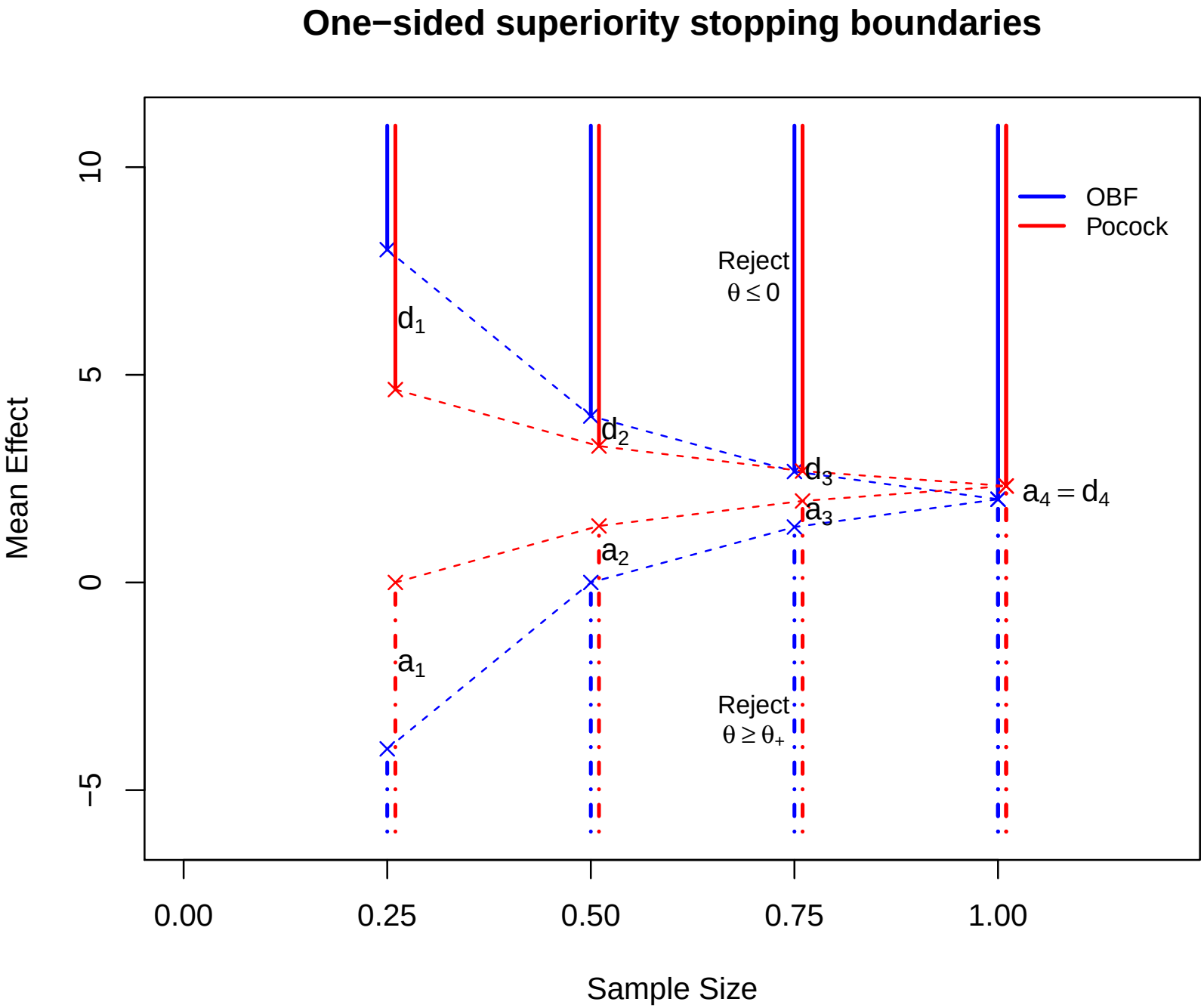


Illustration of statistical implications

Example: OBF versus Pocock 1-sided designs



Effect of stopping boundaries on the sampling density

[Simulated sample paths]

Illustration of statistical implications

Example: O'Brien-Fleming (OBF) 2-sided design

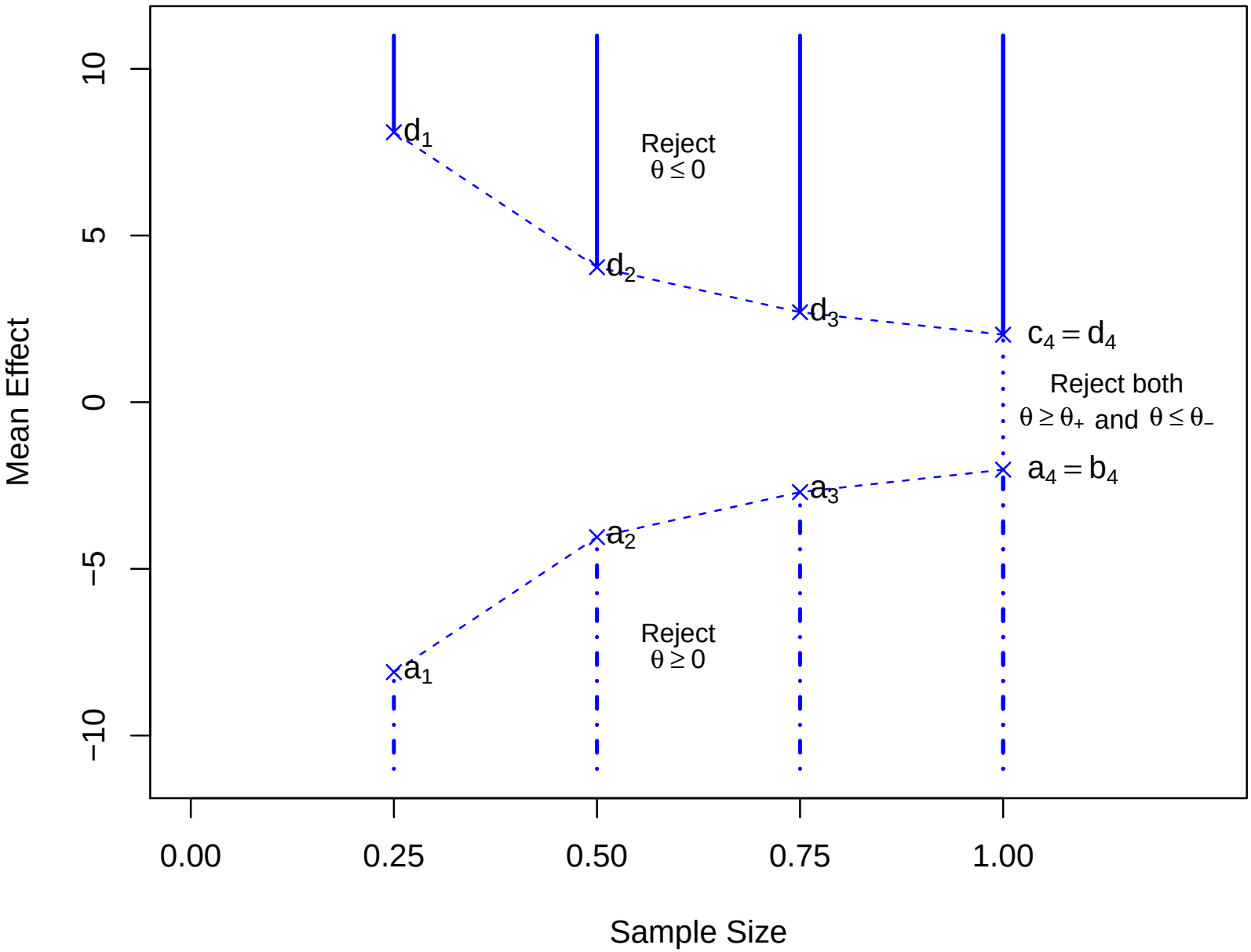


Illustration of statistical implications

Sampling density for OBF boundaries with $\theta = 0$ and $\theta = 3.92$ (corresponding Normal sampling density for comparison):

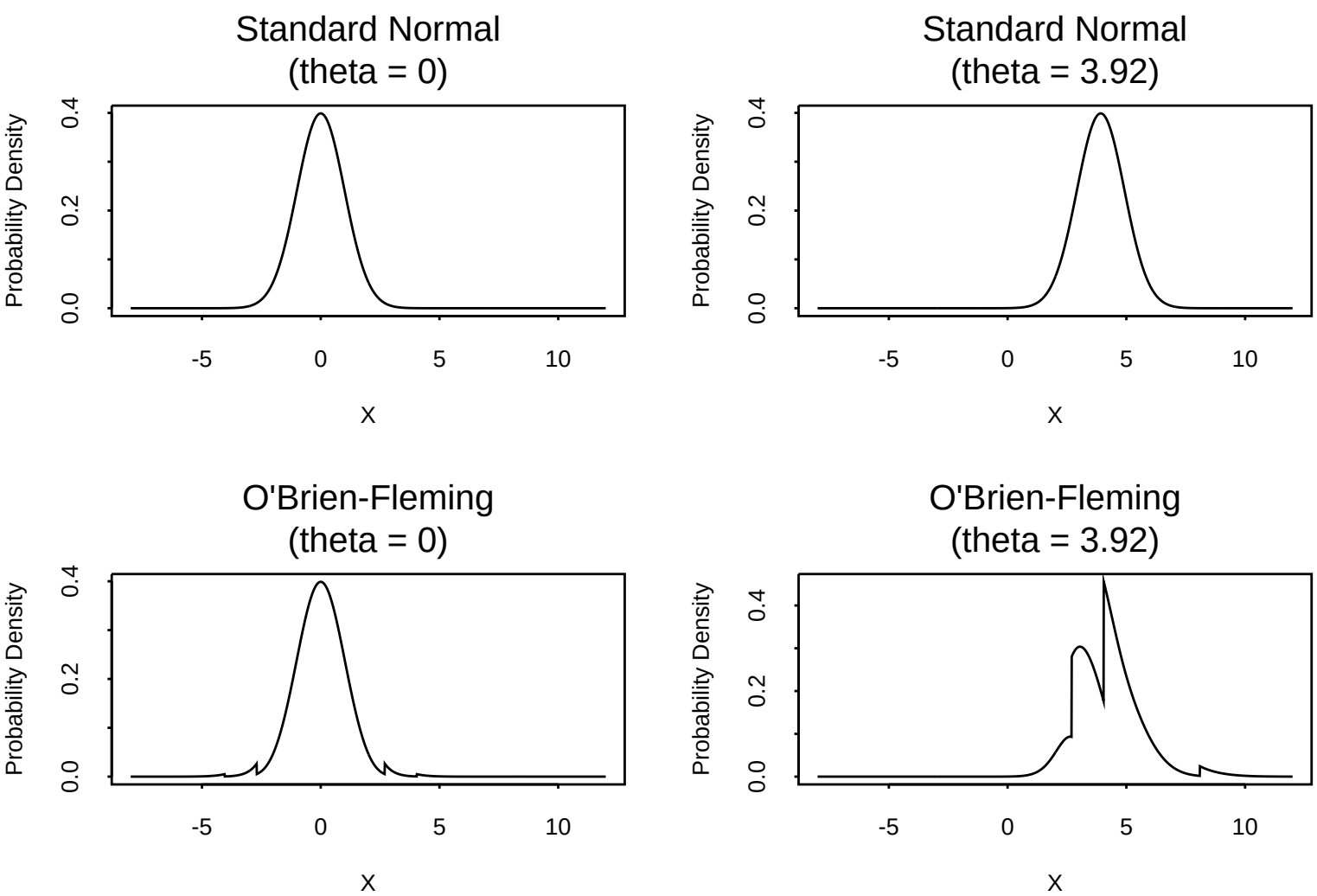


Illustration of statistical implications

Example: OBF versus Pocock 1-sided designs

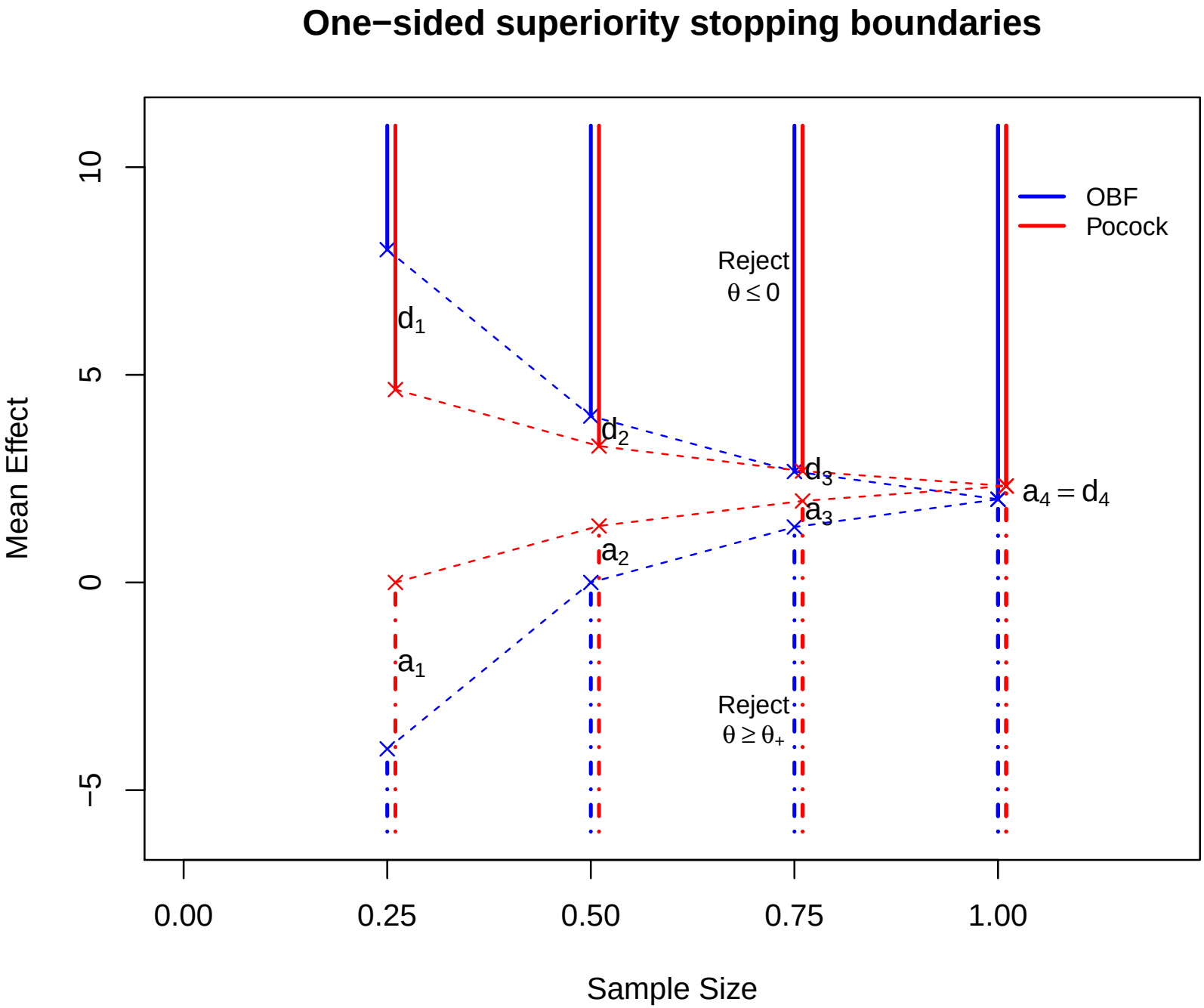


Illustration of statistical implications

Sampling density for OBF and Pocock 1-sided designs.

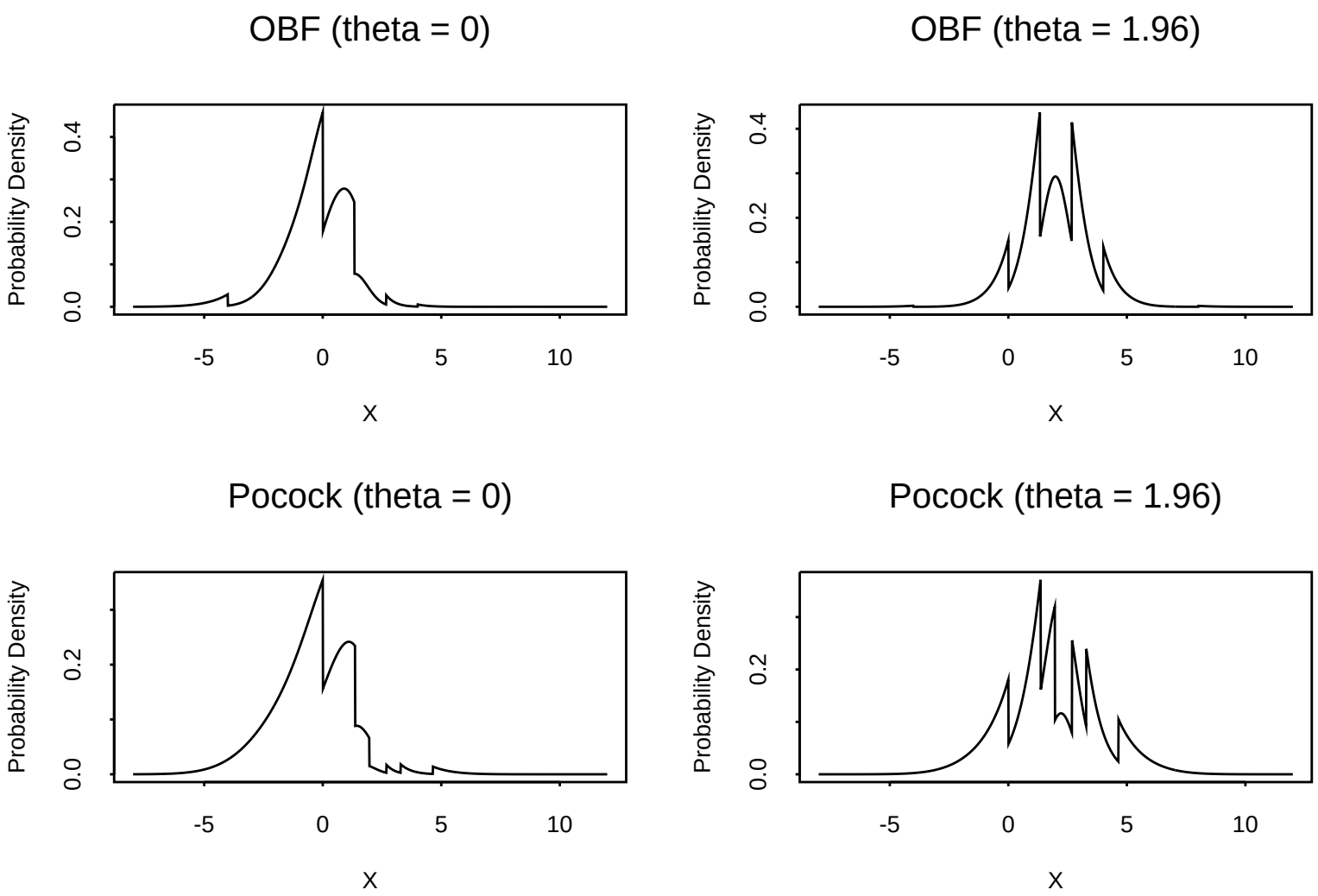


Illustration of statistical implications

Characteristics of the group sequential sampling density:

- Density is not shift invariant
- Jump discontinuities
- Requires numerical integration
- Sequential testing introduces bias:

	$E(\hat{\theta})$	
θ	OBF	Pocock
0.00	-0.29	-0.48
1.96	1.95	1.82
3.92	4.21	4.38

- (Recursive form of the sequential sampling density is computationally useful.)
- Fully discussed in sections 3 and 6 of this course.