Introduction to Clinical Trials - Day 2

Session 1 - Introduction

Presented July 24, 2018

Susanne J. May
Department of Biostatistics
University of Washington

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

Goals of Clinical Trial Design

Clinical trials

- Experimentation in human volunteers
- Investigation of a new treatment or preventive agent
  - *Safety* : Are there adverse effects that clearly outweigh any potential benefit?
  - *Efficacy* : Can the treatment alter the disease process in a beneficial way?
  - *Effectiveness* : Would adoption of the treatment as a standard effect morbidity in the population?

Goals of Clinical Trial Design

A trial must meet minimum scientific standards

- It must address a meaningful question
  - Discriminate between viable hypotheses (Science)
- Trial results must be credible to the scientific community
  - Valid materials, methods (Science, Statistics)
  - Valid measurement of experimental outcome (Science, Clinical, Statistics)
  - Valid quantification of uncertainty in experimental procedure (Statistics)

Goals of Clinical Trial Design

Individual Ethics

- Conducted in human volunteers, the clinical trial must be ethical for participants on the trial
  - Minimize harm and maximize benefit for participants in clinical trial
  - Avoid giving trial participants a harmful treatment
  - Do not unnecessarily give trial participants a less effective treatment

Goals of Clinical Trial Design

Group Ethics

- The clinical trial must ethically address the needs of the greater population of potential recipients of the treatment
  - Approve new beneficial treatments as rapidly as possible
  - Avoid approving ineffective or (even worse) harmful treatments
  - Do not unnecessarily delay the new treatment discovery process

Goals of Clinical Trial Design

Optimality criteria

- A good procedure will

1. Minimize “false positives"
   - Any treatment recommended for adoption will have a high probability of being a truly effective therapy
2. Minimize “false negatives"
   - Any truly effective therapy will have a high probability of being recommended for adoption
3. Be highly safe and ethical
   - Minimize the number of patients exposed to inferior treatments while investigations proceed
4. Be efficient
   - Minimize costs (patients, calendar time, money)

Goals of Clinical Trial Design

Common statistical approach

- Optimality criteria (1) and (2) speak directly to the need for achieving high PPV and low NPV
- Design an RCT to answer relevant question
  - Treatment, patient population, intervention, comparator, outcome
    - There is an underlying probability of our hypotheses being correct: “Prevalence of effective therapies"
- Fix probability of making wrong decisions
  - Erroneously decide against status quo < 2.5%
  - But: erroneously decide against status quo 2.5%
- Design trial to fix sensitivity of study
  - Power: High probability to detect beneficial treatment

Goals of Clinical Trial Design

Positive predictive value in research

- Positive predictive value: probability that a statistically significant trial indicates a truly effective treatment.
- Negative predictive value: probability that a non-significant trial indicates a truly non-effective treatment.
- Relationship to type I error, power, and prevalence of truly effective therapies

$$PPV = \frac{\text{Power} \times \text{Prev}}{\text{Power} \times \text{Prev} + (\text{Type I Error}) \times (\text{1-Prev})}$$

$$NPV = \frac{(\text{1-Type I Error}) \times (\text{1-Prev})}{(\text{1-Type I Error}) \times (\text{1-Prev}) + (\text{1-Power}) \times \text{Prev}}$$

Goals of Clinical Trial Design

Predictive value of statistically significant result depends on

1. Probability hypothesis is true to begin with (start with “good ideas")
   - Fixed when hypothesis is formulated
2. Type I error (Specificity)
   - Fixed by level of significance
3. Power (Sensitivity)
   - Statistical power made as high as possible by design

Goals of Clinical Trial Design

The later two elements are improved by

1. Minimizing bias
   - Remove confounding and account for effect modification
2. Decreasing variability of measurements
   - Homogeneity of population, appropriate endpoints, appropriate sampling strategy, more precise measuring device

Goals of Clinical Trial Design

Common pitfalls of studies

- Common pitfalls of experimentation are:
  - Data driven hypotheses (↑ Type I error)
  - Multiple comparisons (↑ Type I error)
  - Poor selection of subjects (↓ Power)
  - Over-fitting of data (↑ Type I error, (↓ Power)
  - Poor selection of subjects, outcomes (↓ Power)
  - Noncomparability of treatment groups (↑ Type I error)
- Each of these pitfalls leads to increases in variability and/or bias in clinical trials...

Course roadmap

Where are we going?

- Module 1: Design
  - Background
    - Phases of clinical trials
    - Interplay between science and statistics
    - Ethics and varying roles of oversight committees
  - Role screening studies in trial design
  - Fundamental design elements
    - Variability and bias
    - Identification of target population
    - Definition of intervention(s)
    - Choice of outcomes
    - Choice of comparison groups
    - Blinding
    - Brief introduction to randomization
  - Statistical tasks in trial design
    - Refinement of hypotheses
    - Probability models and summary measures
    - Determination of sample size
  - Focus on elements of a clinical trial protocol

Course roadmap

Where are we going?

- Module 2: Primarily implementation
  - Choice of outcome (surrogate outcomes vs. clinical outcomes)
  - Methods of randomization
  - Monitoring for quality and missing data
  - Role and function of IDMCs
  - Group sequential monitoring
  - Data management
  - Review of key elements of a clinical trial protocol
  - (Extra?) Further discussion on common endpoints: survival and change from baseline

Acknowledgments

- Many thanks to the following individuals for the use of some of their slides on the topics to be presented:
  - Scott S. Emerson, University of Washington
  - John M. Kittelson, University of Colorado

Introduction to Clinical Trials - Day 2

Session 2 - Surrogate Endpoints

Presented July 24, 2018

Susanne J. May
Department of Biostatistics
University of Washington

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

Choice of a Primary Outcome

Importance of primary outcome specification

- The goal of a RCT is to find effective treatment indications
  - The primary outcome is a crucial element of the indication
- Scientific basis:
  - A clinical trial is planned to detect the effect of a treatment on some outcome
  - Statement of the outcome is a fundamental part of the scientific hypothesis
- Ethical basis:
  - Generally, subjects participating in a clinical trial are hoping that they will benefit in some way from the trial
  - Clinical endpoints are therefore of more interest than purely biological endpoints

Choice of a Primary Outcome

Multiple comparison issues

- Type I error for each endpoint
  - In absence of treatment effect, will still decide a benefit exists with probability, say, .025
- Multiple endpoints increase the chance of deciding an ineffective treatment should be adopted:
  - This problem exists with either frequentist or Bayesian criteria for evidence
  - The actual inflation of the type I error depends on
    1. the number of multiple comparisons, and
    2. the correlation between the endpoints

Choice of a Primary Outcome

Multiple comparison issues

- Ex: Consider experiment-wise error rate when using level .05 per decision

Number Compared	Worst Case	Correlation 0.00	0.30	0.50	0.75	0.90
1	.050	.050	.050	.050	.050	.050
2	.100	.098	.095	.090	.081	.070
3	.150	.143	.137	.126	.104	.084
5	.250	.226	.208	.184	.138	.101
10	.500	.401	.353	.284	.193	.127
20	1.000	.642	.540	.420	.258	.154
50	1.000	.923	.806	.624	.353	.193

Choice of a Primary Outcome

Primary endpoint: Clinical

- Should consider (in order of importance)
  - The most relevant clinical endpoint (Survival, quality of life)
  - The endpoint the treatment is most likely to affect
  - The endpoint that can be assessed most accurately and precisely

Additional Endpoints

- Other outcomes are then relegated to a “secondary" status
  - Supportive and confirmatory
  - Safety
- Some outcomes are considered “exploratory"
  - Subgroup effects
  - Effect modification

Choice of a Primary Outcome

Primary endpoint: Clinical

- Should consider (in order of importance)
  - The phase of study: What is current burden of proof?
  - The most relevant clinical endpoint (Survival, quality of life)
    - Proven surrogates for relevant clinical endpoint (????) More later...
  - The endpoint the treatment is most likely to affect
    - Therapies directed toward improving survival
    - Therapies directed toward decreasing AEs
  - The endpoint that can be assessed most accurately and precisely
    - Avoid unnecessarily highly invasive measurements
    - Avoid poorly reproducible endpoints

Choice of a Primary Outcome

Multiple endpoints

- Sometimes we must consider multiple endpoints
- We then control experiment-wise error
- Possible methods include
  - Composite endpoint
    - AND: Individual success must satisfy all
    - OR: Individual success must only satisfy one
    - AVERAGE: Sum of individual scores
    - EARLIEST: e.g., event free survival
  - Co-primary endpoints
    - Must show improvement in treatment group on all endpoints
    - No guarantee that the same subjects are experiencing the improvement

Choice of a Primary Outcome

Competing risks

- Occurrence of some “nuisance" event precludes observation of the event of greatest interest, because
  - Further observation impossible
    - E.g., death from CVD in cancer study
  - Further observation irrelevant
    - E.g., patient advances to other therapy (transplant)
- Methods
  - Event free survival: time to earliest event
  - Time to progression: censor competing risks
  - “U statistics": define ranking based on both events

Competing risks caveats

- Competing risks produce missing data on the event of greatest interest
- As with all missing data problems, there is nothing in your data that can tell you whether your actions are appropriate
  - Are subjects with competing risk more or less likely to have event of interest?
  - (the term “competing risk" has become shorthand for a setting in which your results are in doubt)

Choice of a Primary Outcome

Issues with clinical outcomes

- Goal of clinical trial is to establish whether an experimental treatment will prevent a particular clinical outcome
  - Incidence of disease
  - Decreased quality of life
  - Mortality
- Relevant clinical outcomes are often relatively rare events that occur after a significant delay
  - Believe that earlier interventions have greater chance of benefit
- It can also be logistically difficult to measure a clinical outcome
  - Quality of life needs to be assessed over a sufficiently long period of time

Choice of a Primary Outcome

Impact on trial design

- Large sample size required to assess treatment effect on rare events
- Long period of follow-up needed to assess endpoints
- Isn't there something else that we can do?
- A tempting alternative is to move to "surrogate" endpoints...

Motivation for surrogate endpoints

- Hypothesized role of surrogate endpoints
  - Find a biological endpoint which
    - can be measured in a shorter timeframe,
    - can be measured precisely, and
    - is predictive of the clinical outcome
  - Use of such an endpoint as the primary measure of treatment effect will result in more efficient trials

Surrogate Endpoints

Identifying potential surrogates

- Typically use observational data to find risk factors for clinical outcome
- Treatments attempt to intervene on those risk factors
- Surrogate endpoint for the treatment effect is then a change in the risk factor

Surrogate Endpoints

Examples of surrogates

- Colon cancer prevention
  - Two-fold increase in risk of colon cancer for patients with adenomatous colon polyps
  - Prevention directed toward preventing colon polyps
  - Treatment effect measured by decreased incidence of colon polyps
  - True clinical outcome is preventing mortality

Examples of surrogates

- HIV/AIDS
  - HIV leads to suppression of CD4 cells
  - Decreased CD4 levels correlates with development of AIDS
  - Treatment effects measured by following CD4 counts
  - True clinical outcome is prevention of morbidity and mortality

Surrogate Endpoints

Examples of surrogates

- Coronary heart disease
  - Poor prognosis in patients with arrhythmias following heart attack
  - Therapies directed toward preventing arrhythmias
  - Treatment effects measured by prevention of arrhythmias
  - True clinical outcome is prevention of mortality

Surrogate Endpoints

Examples of surrogates

- Liver failure
  - Poor prognosis in patients who develop renal failure
  - Therapies directed toward treating renal failure (dialysis)
  - Treatment effects measured by creatinine, BUN
  - True clinical outcome is prevention of mortality

Surrogate Endpoints

Examples of surrogates

- Other examples that have been used historically include
  - Cancer: tumor shrinkage
  - Coronary heart disease: cholesterol, nonfatal MI, blood pressure
  - Congestive heart failure: cardiac output
  - Arrhythmia: atrial fibrillation
  - Osteoporosis: bone mineral density
- Future surrogates?
  - Gene expression
  - Proteomics

Surrogate Endpoints

Problem with surrogates

- Establishing biologic activity does not always translate into effects on the clinical outcome
- May be treating the symptom, not the disease
  - Concorde: ZDV improves CD4, not survival
  - CAST: encainide, flecainide prevents arrhythmias, worsens survival
- May be missing effect through other pathways
  - Intl CGD group: Gamma-INF no affect on biomarkers, decreases serious infections

Examples of Problems with Surrogate Endpoints

Ex: Concorde Trial (Lancet, 1993)

- Asymptomatic HIV positive patients
- Randomize to
  - Immediate ZDV (n = 877)
  - Placebo then progression to ZDV (n = 872)
- Mean follow-up: 3 years

Examples of Problems with Surrogate Endpoints

Ex: Concorde Trial (Lancet, 1993)

- Observed CD4 changes
- 3 mos relative to baseline
  - Immediate ZDV: +20 cells
  - Placebo: -10 cells
- Difference between treatment arms
  - 3 mos: 30 cells (P < .0001)
  - 6 mos: 35 cells (P < .0001)
  - 9 mos: 32 cells (P < .0001)

Examples of Problems with Surrogate Endpoints

Ex: Concorde Trial (Lancet, 1993)

- However, more deaths observed on ZDV arm with roughly equal 3-year survival rate

	ZDV (n = 877)	Placebo (n = 872)
AIDS / Death	175	171
Death	95	76
3 year survival	92%	93%

"Results cast doubt on the value of using changes over time in CD4 count as a predictive measure for effects of antiviral therapy on disease progression and survival."

Ex: HIV Meta-analysis

- Review of ZDV, ddI and ddC on Surrogate Markers and Clinical Endpoints
  - 16 trials reviewed by NIAID S.O.T.A. Panel, Jun 93

		AIDS/Death		Survival			
		+	-	+	-	--	?
CD4	+	7	6	2	6	3	2
Effect	-	1	2	2	1	0	0

Ex: CAST

Ex: Cardiac Arrhythmia Suppression Trial (CAST)

- Arrhythmia a risk factor for sudden death following a myocardial infarction
- Antiarrhythmic drugs (encainide and flecainide) successfully decrease incidence of arrhythmias
- CAST
  - Placebo controlled trial using mortality as outcome
  - Encainide and flecainide TRIPLE the death rate

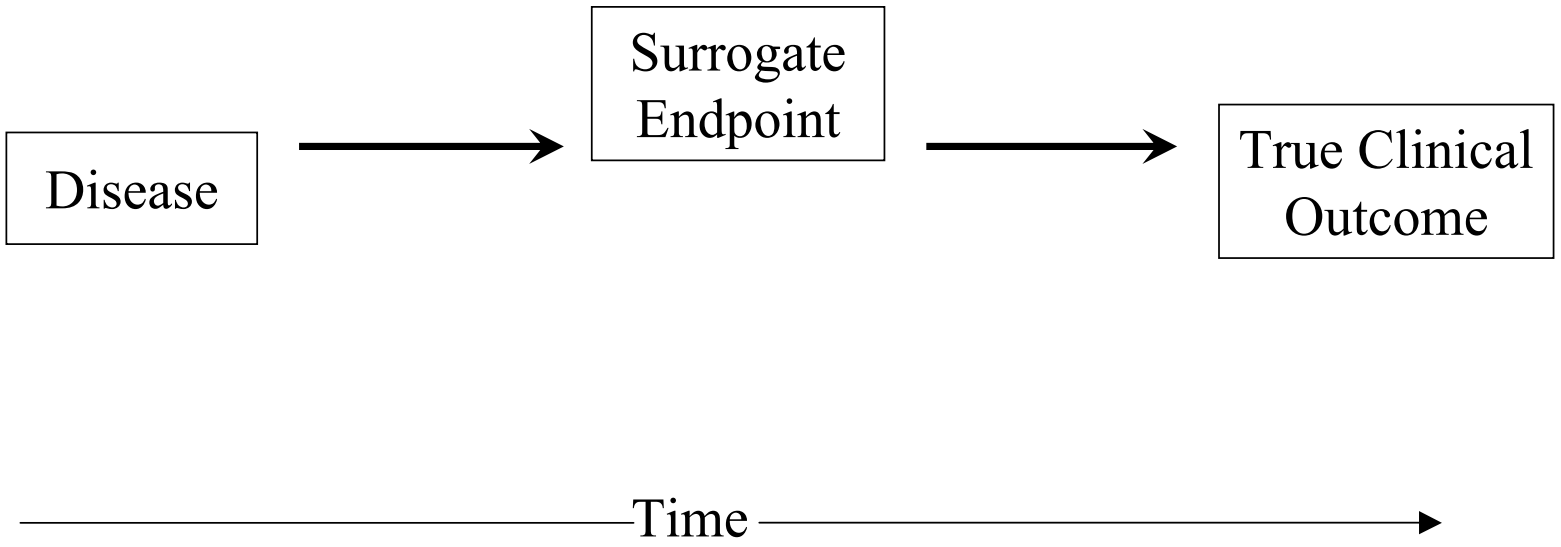
Ex: Chronic Granulomatous Disease (CGD)

- CGD leads to recurrent serious infections
- Gamma interferon increases bacterial killing and superoxide production?
- International CGD Study Group Trial of Gamma-INF
  - 70% reduction in recurrent serious infections
  - Essentially no effect on biological markers

Surrogate Endpoints

Scenario 1: Ideal Surrogate

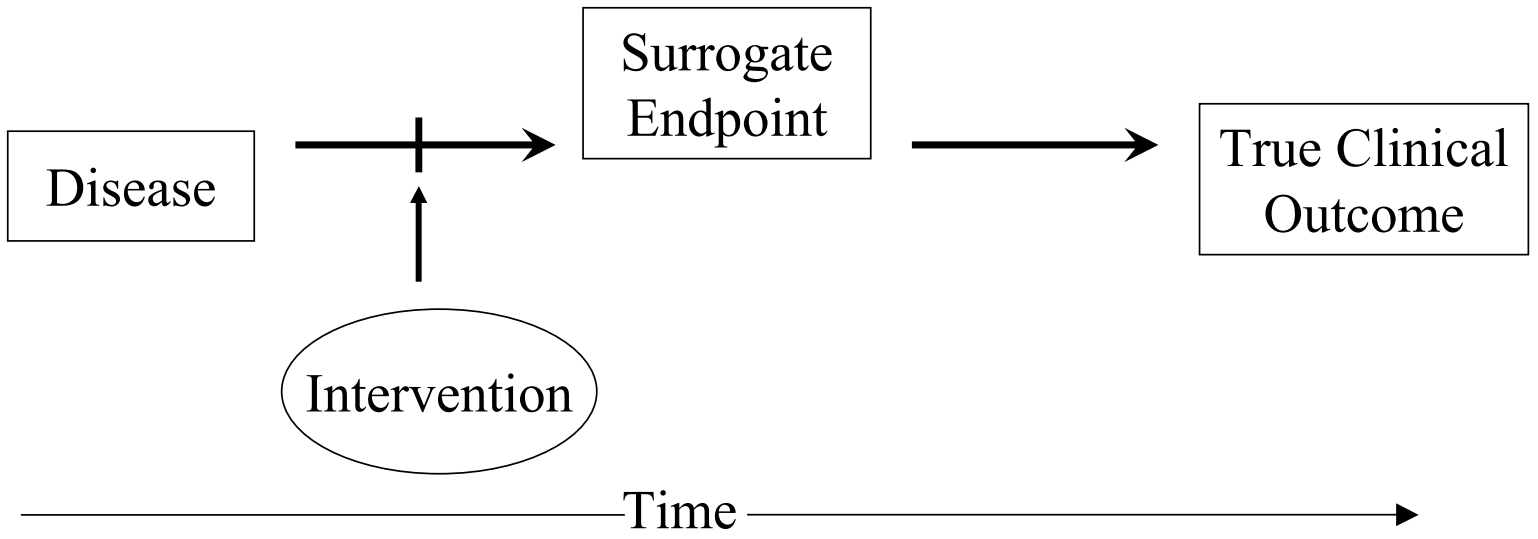
- Disease progresses to Clinical Outcome only through the Surrogate Endpoint



Surrogate Endpoints

Scenario 1a: Ideal Surrogate Use

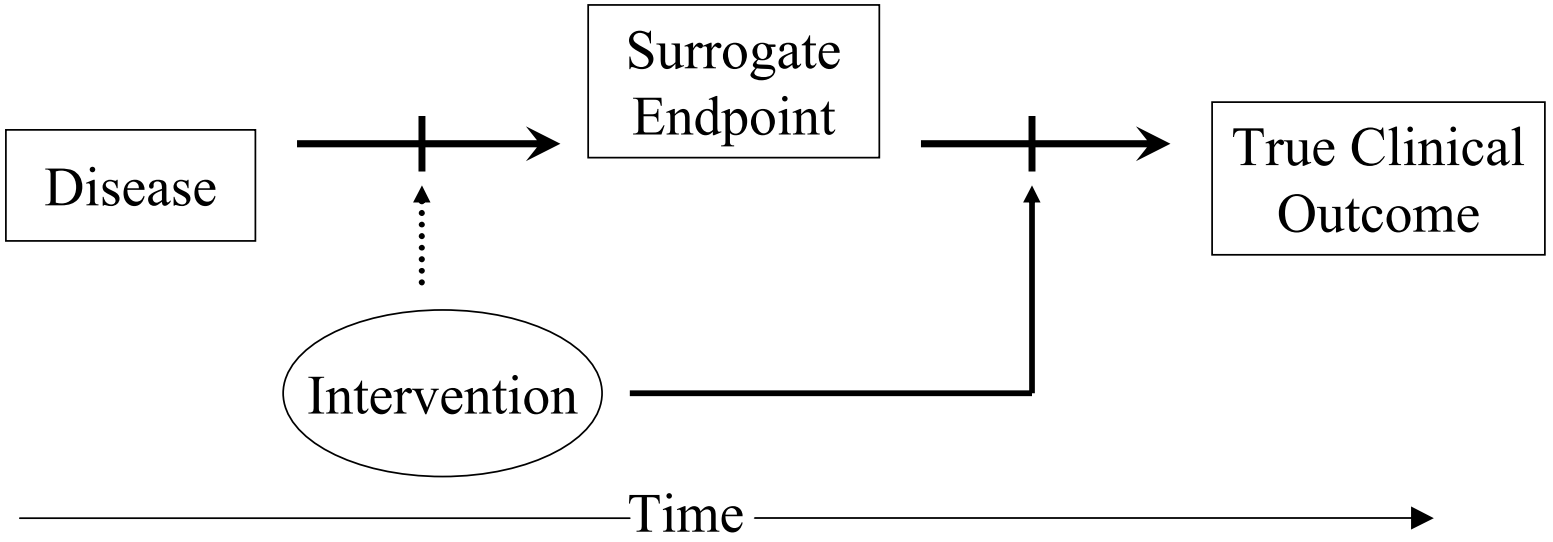
- The intervention's effect on the Surrogate Endpoint accurately reflects its effect on the Clinical Outcome



Surrogate Endpoints

Scenario 1b: Inefficient Surrogate

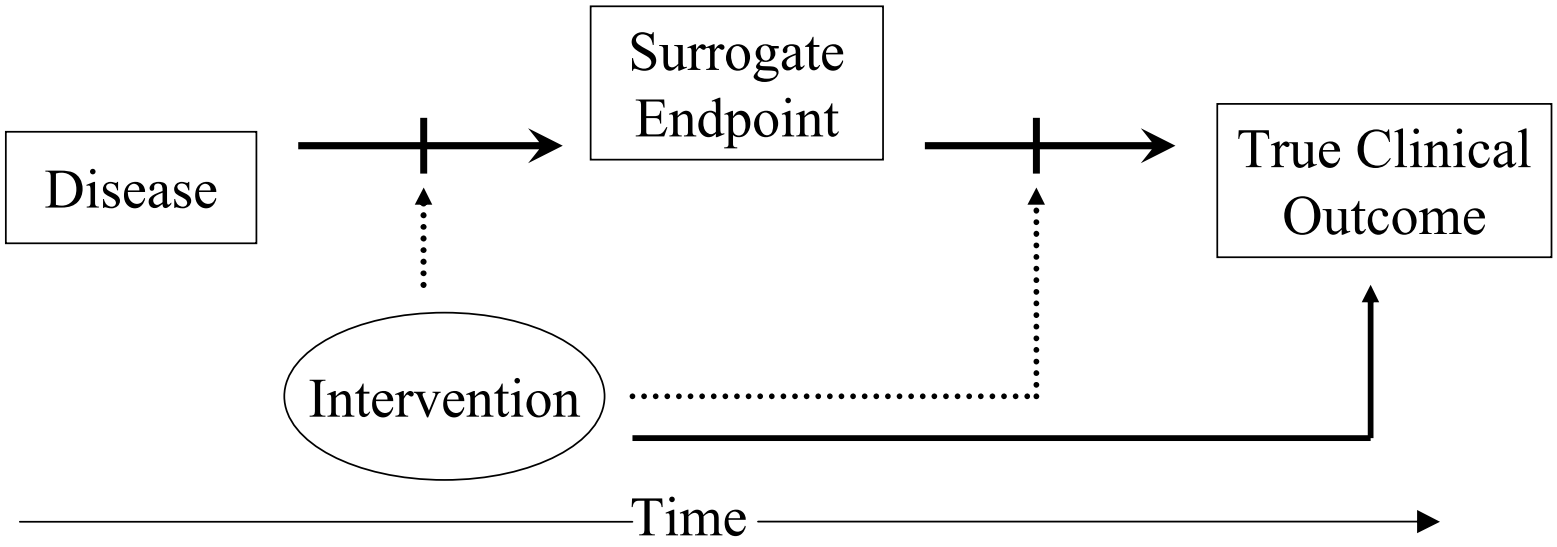
- The intervention's effect on the Surrogate Endpoint understates its effect on the Clinical Outcome



Surrogate Endpoints

Scenario 1d: Dangerous Surrogate

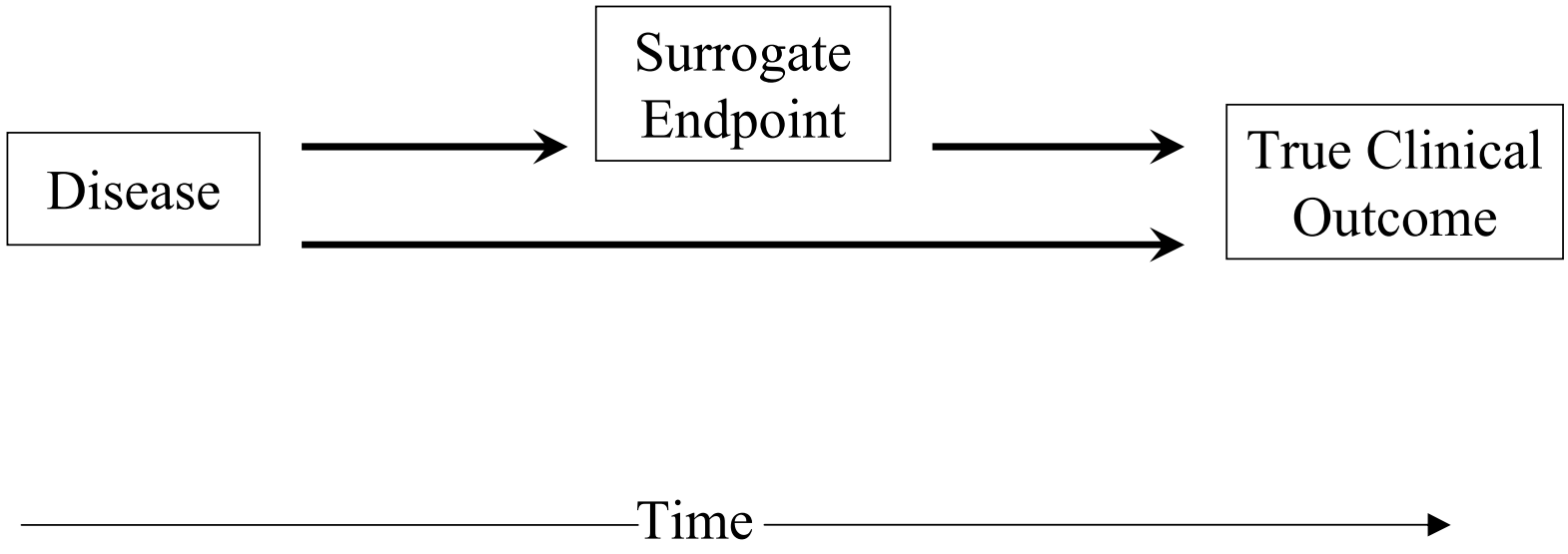
- Effect on the Surrogate Endpoint may overstate its effect on the Clinical Outcome (which may actually be harmful)



Surrogate Endpoints

Scenario 2: Alternate Pathways

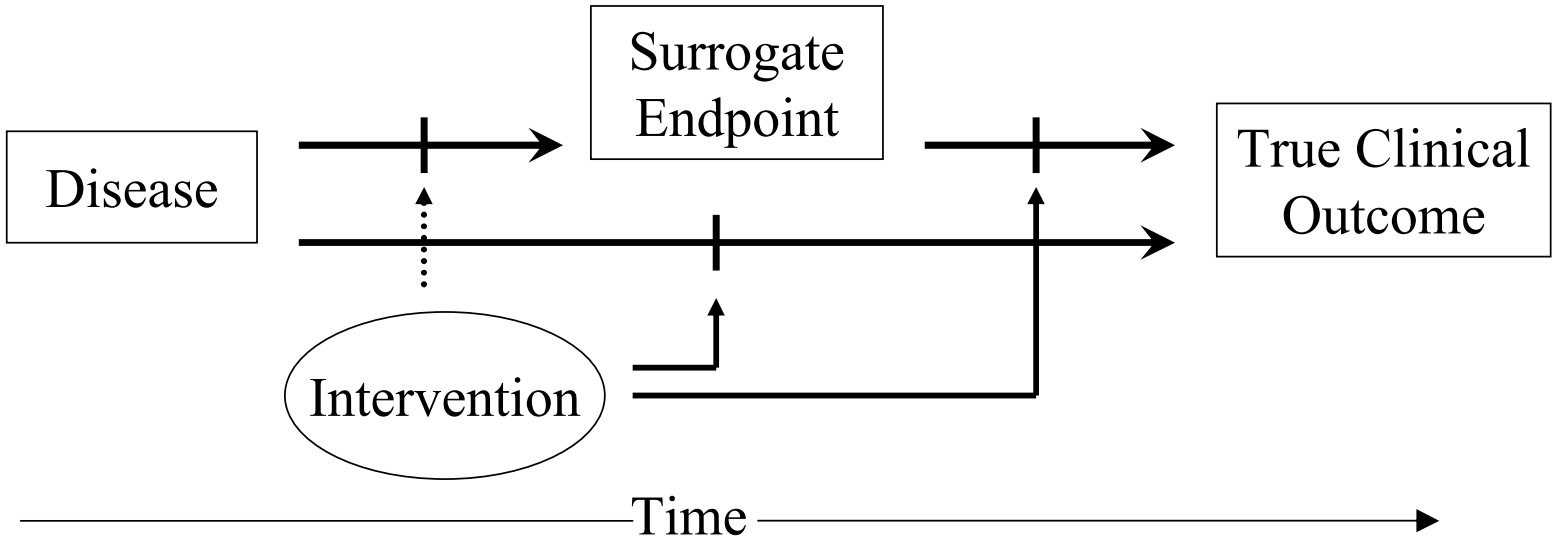
- Disease progresses directly to Clinical Outcome as well as through Surrogate Endpoint



Surrogate Endpoints

Scenario 2b: Inefficient Surrogate

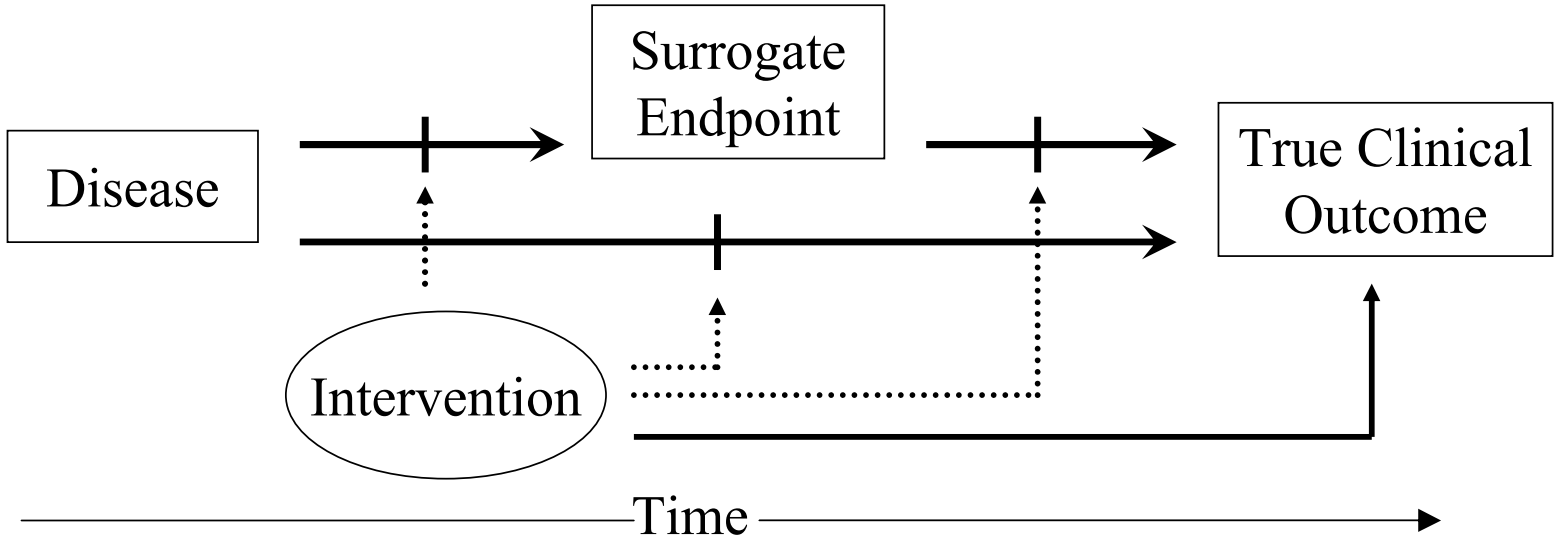
- Treatment's effect on Clinical Outcome is greater than is reflected by Surrogate Endpoint



Surrogate Endpoints

Scenario 2d: Dangerous Surrogate

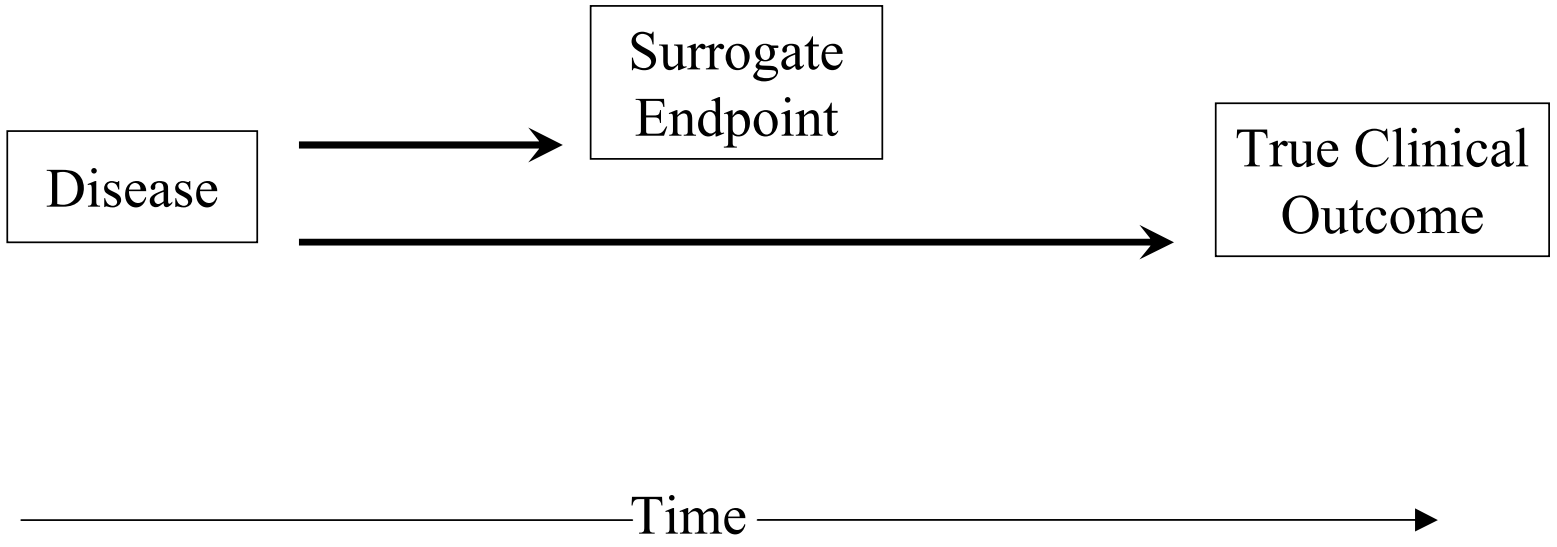
- The effect on the Surrogate Endpoint may overstate its effect on the Clinical Outcome (which may actually be harmful)



Surrogate Endpoints

Scenario 3: Marker

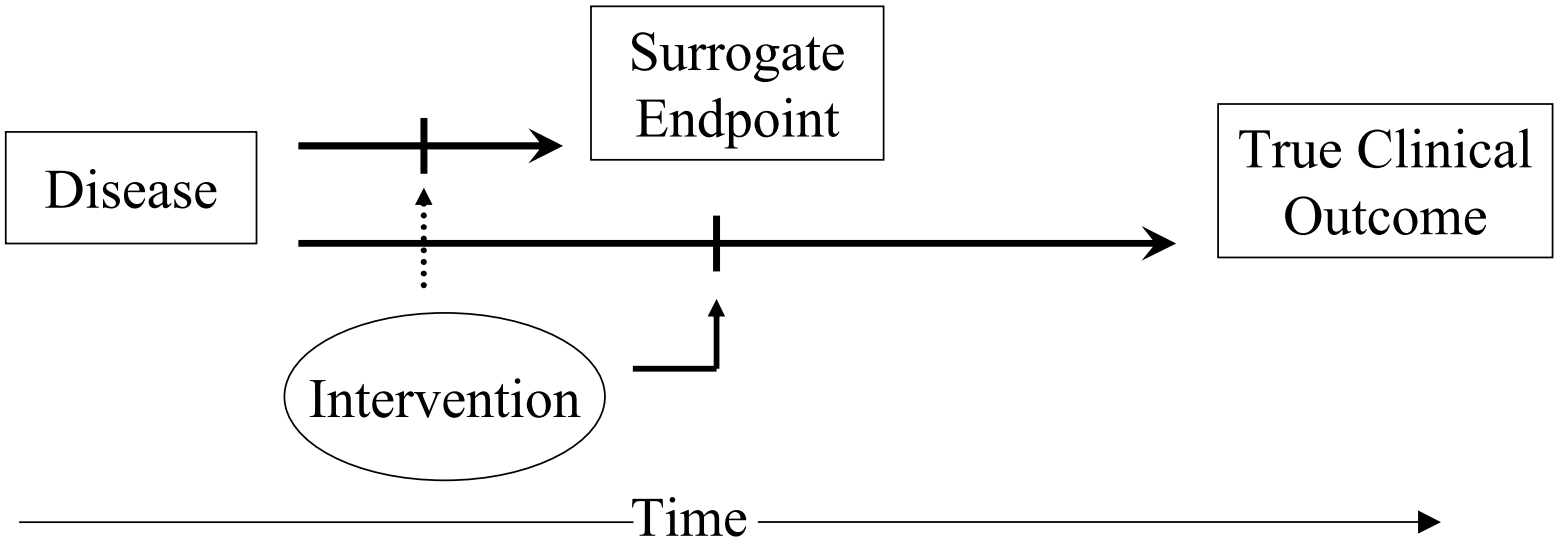
- Disease causes Surrogate Endpoint and Clinical Outcome via different mechanisms



Surrogate Endpoints

Scenario 3b: Inefficient Surrogate

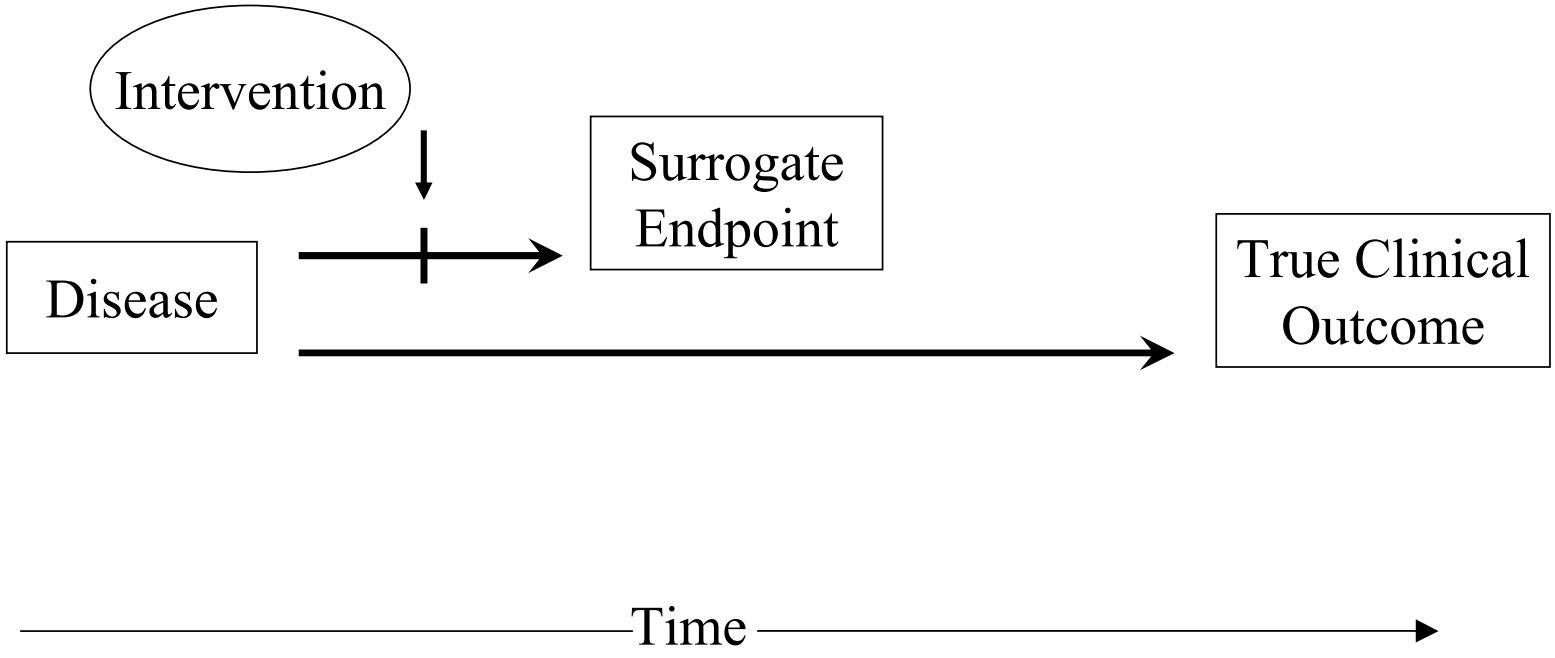
- Treatment's effect on Clinical Outcome is greater than is reflected by Surrogate Endpoint



Surrogate Endpoints

Scenario 3c: Misleading Surrogate

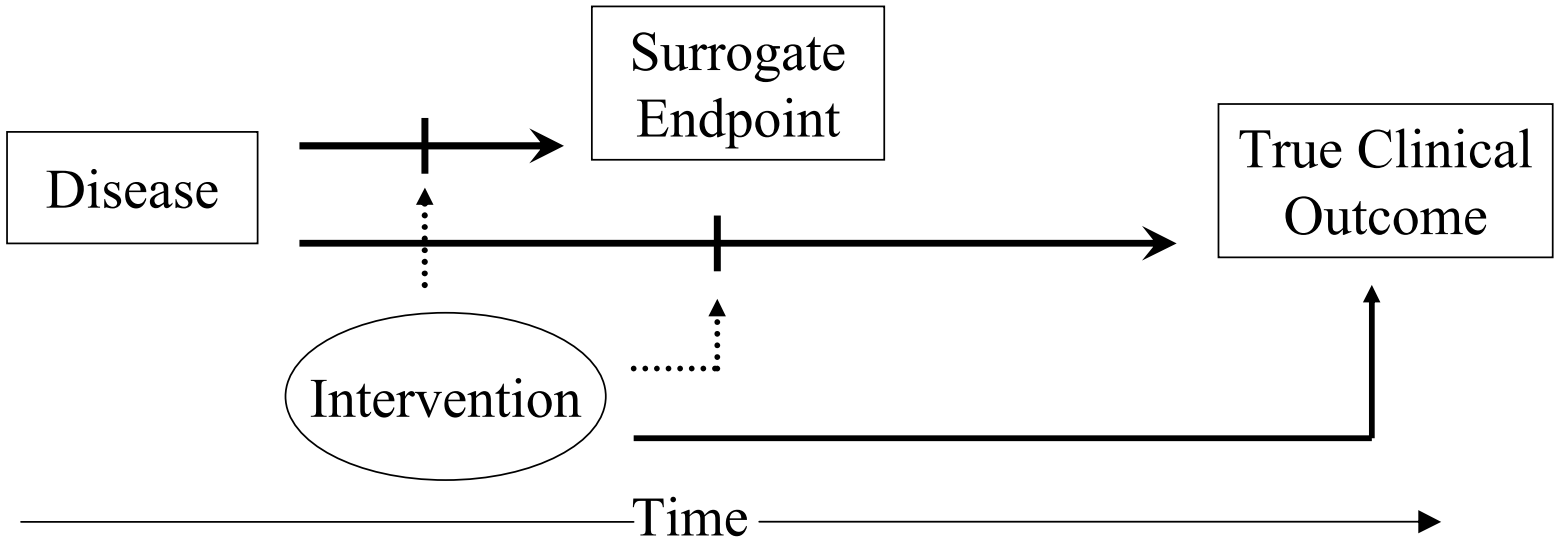
- Effect on Surrogate Endpoint does not reflect lack of effect on Clinical Outcome



Surrogate Endpoints

Scenario 3d: Dangerous Surrogate

- Effect on the Surrogate Endpoint may overstate its effect on the Clinical Outcome (which may actually be harmful)



Ex: HIV Meta-Analysis

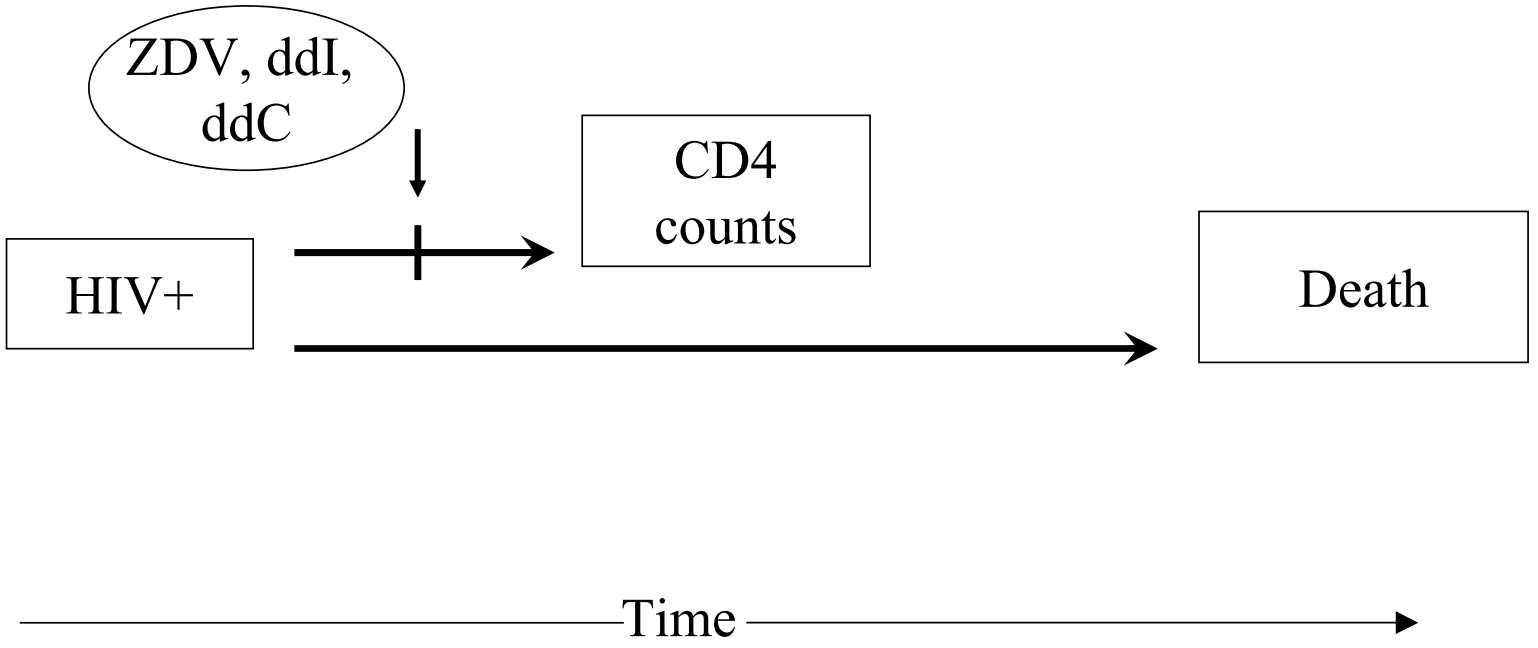
Ex: HIV Meta-analysis

- Review of ZDV, ddI and ddC on Surrogate Markers and Clinical Endpoints
  - 16 trials reviewed by NIAID S.O.T.A. Panel, Jun 93

		AIDS/Death		Survival			
		+	-	+	-	--	?
CD4	+	7	6	2	6	3	2
Effect	-	1	2	2	1	0	0

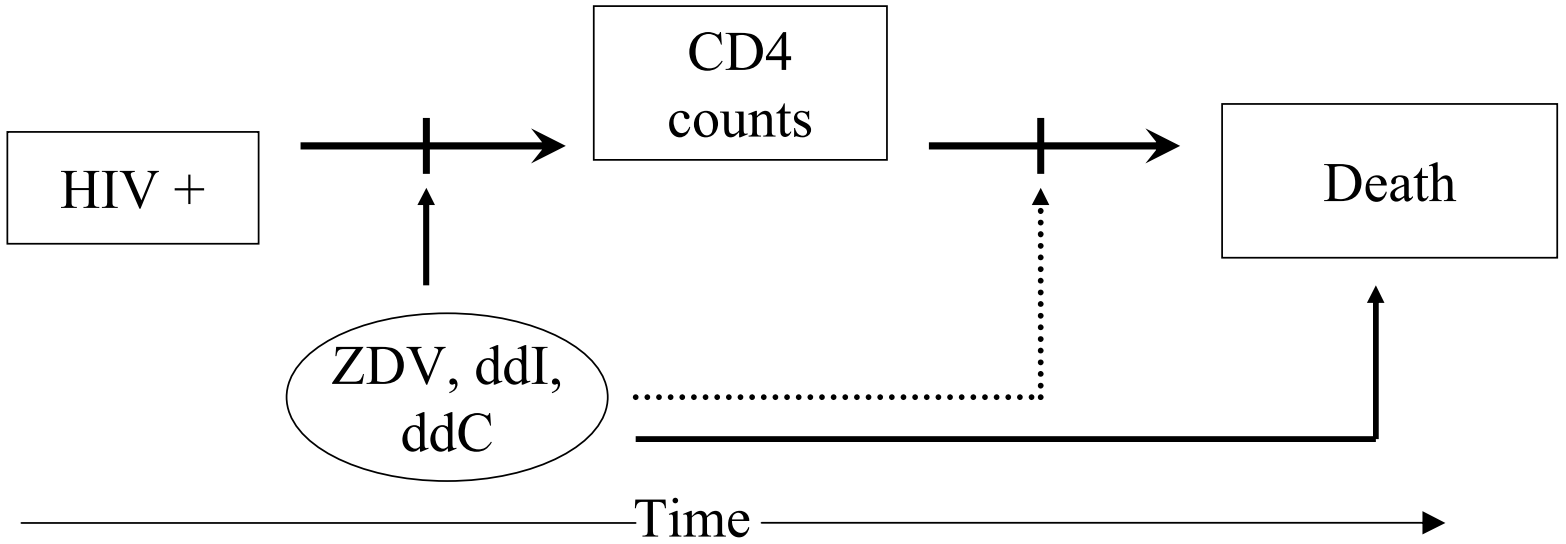
Ex: HIV Meta-Analysis

Scenario 3c: Misleading Surrogate



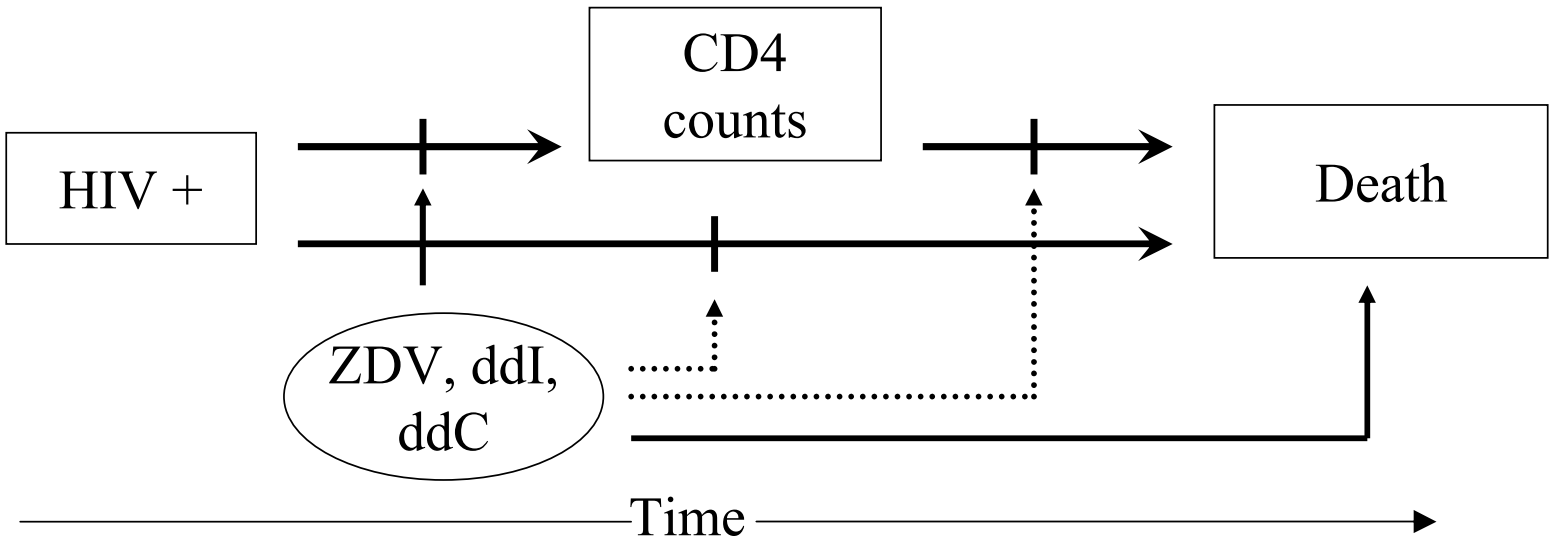
Ex: HIV Meta-Analysis

Scenario 1d: Dangerous Surrogate



Ex: HIV Meta-Analysis

Scenario 2d: Dangerous Surrogate

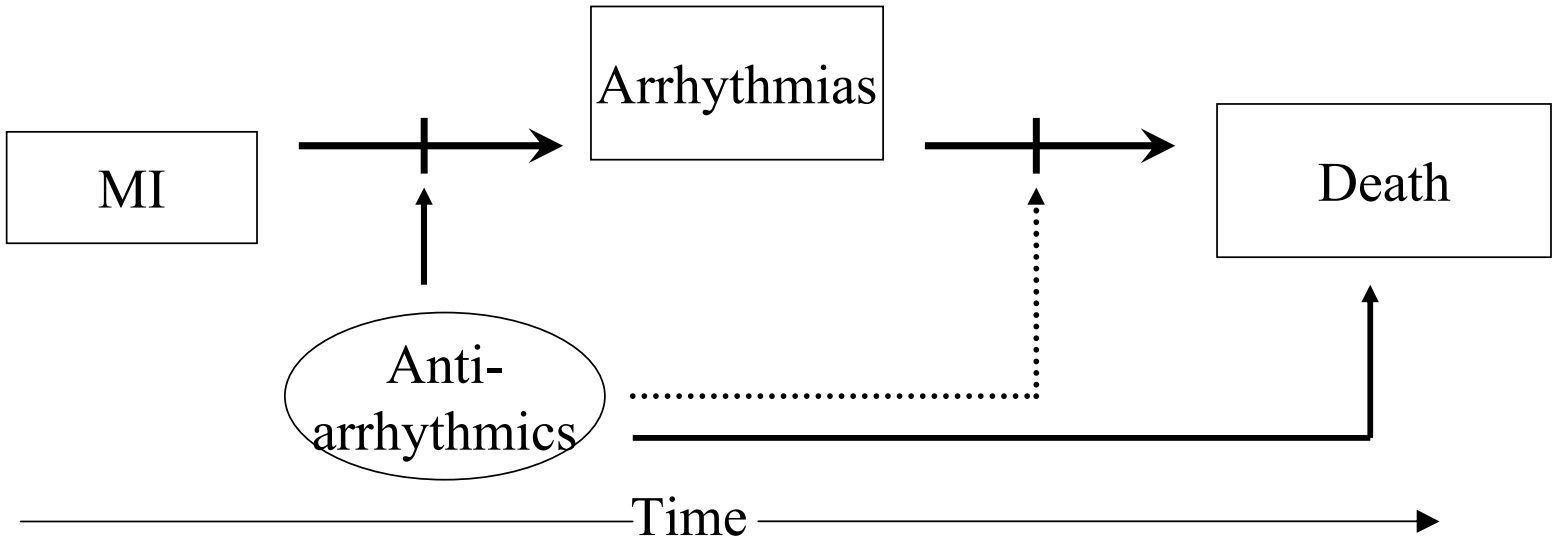


Ex: CAST

Ex: Cardiac Arrhythmia Suppression Trial (CAST)

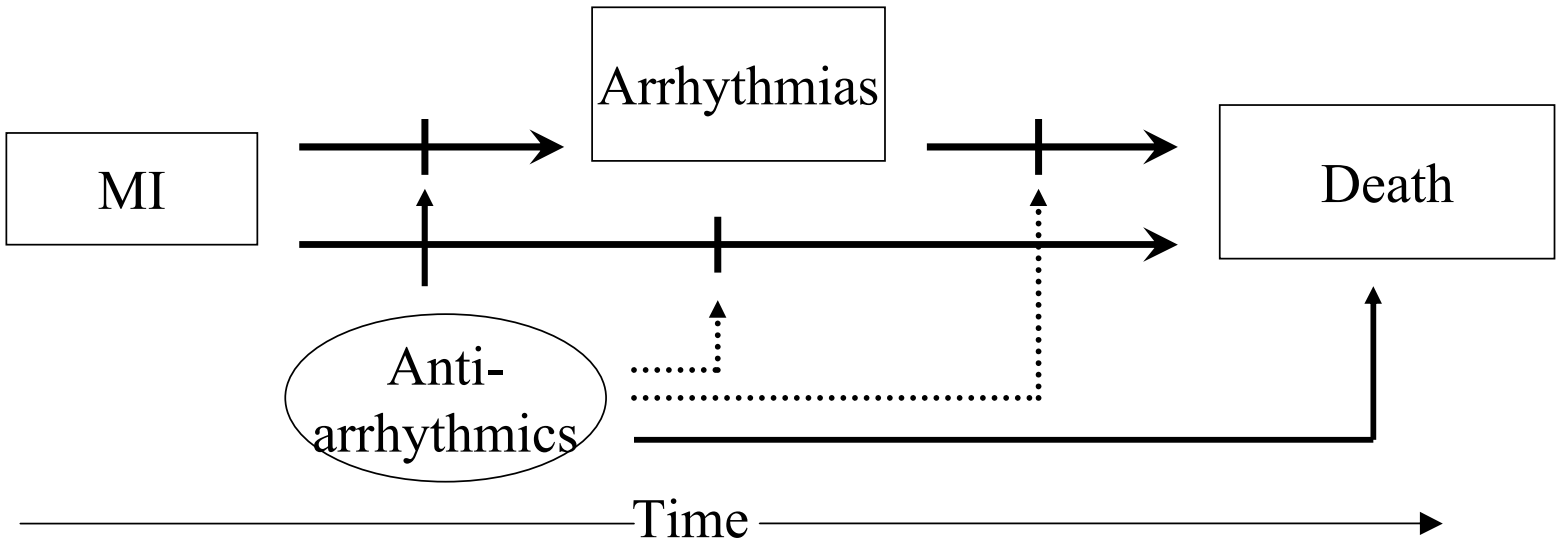
- Arrhythmia a risk factor for sudden death following a myocardial infarction
- Antiarrhythmic drugs (encainide and flecainide) successfully decrease incidence of arrhythmias
- CAST
  - Placebo controlled trial using mortality as outcome
  - Encainide and flecainide TRIPLE the death rate

Scenario 1d: Dangerous Surrogate



Ex: CAST

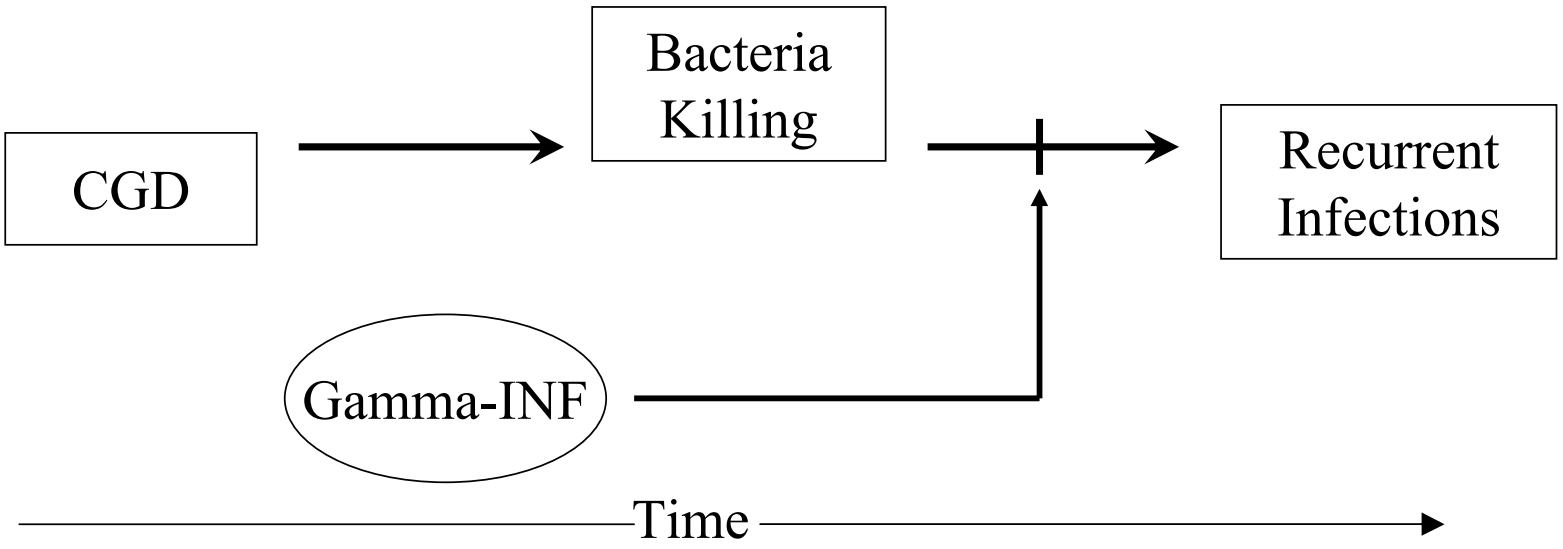
Scenario 2d: Dangerous Surrogate



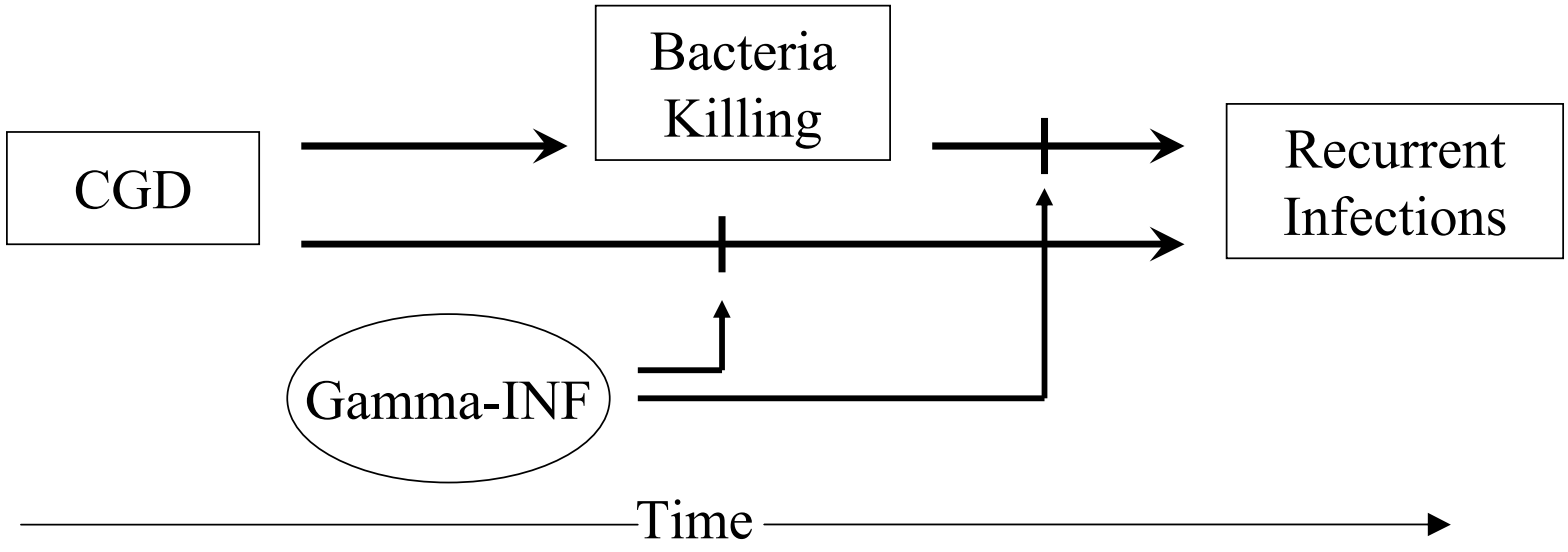
Ex: Chronic Granulomatous Disease (CGD)

- CGD leads to recurrent serious infections
- Gamma interferon increases bacterial killing and superoxide production?
- International CGD Study Group Trial of Gamma-INF
  - 70% reduction in recurrent serious infections
  - Essentially no effect on biological markers

Scenario 1b: Inefficient Surrogate



Scenario 2b: Inefficient Surrogate



Validation of Surrogate Outcomes

Can we validate a surrogate endpoint?

- Many proposed fixes for surrogate outcomes revolve around “validation" of particular surrogate outcomes
  - This is generally very difficult to do
- Is there a way to validate a surrogate endpoint by establishing which causal pathway holds?
- What doesn’t work...
  - It is not sufficient to establish that the surrogate endpoint predicts the clinical outcome in each treatment group separately
  - Treatment can affect the distribution of the surrogate endpoint while increasing mortality in every level

Validation of Surrogate Outcomes

What doesn't work...

- Consider the following hypothetical example

	Treatment		Control	
Surrogate	n	% die	n	% die
Low	30	50%	10	30%
Medium	40	60%	30	40%
High	30	70%	60	50%
Total	100	60%	100	45%

Validation of Surrogate Outcomes

Ex: CARET

- Beta-carotene supplementation for prevention of cancer in smokers
- Treatment group had excess cancer incidence and death
- Within each group, subjects having higher beta-carotene levels in their diet had better survival

Validation of Surrogate Outcomes

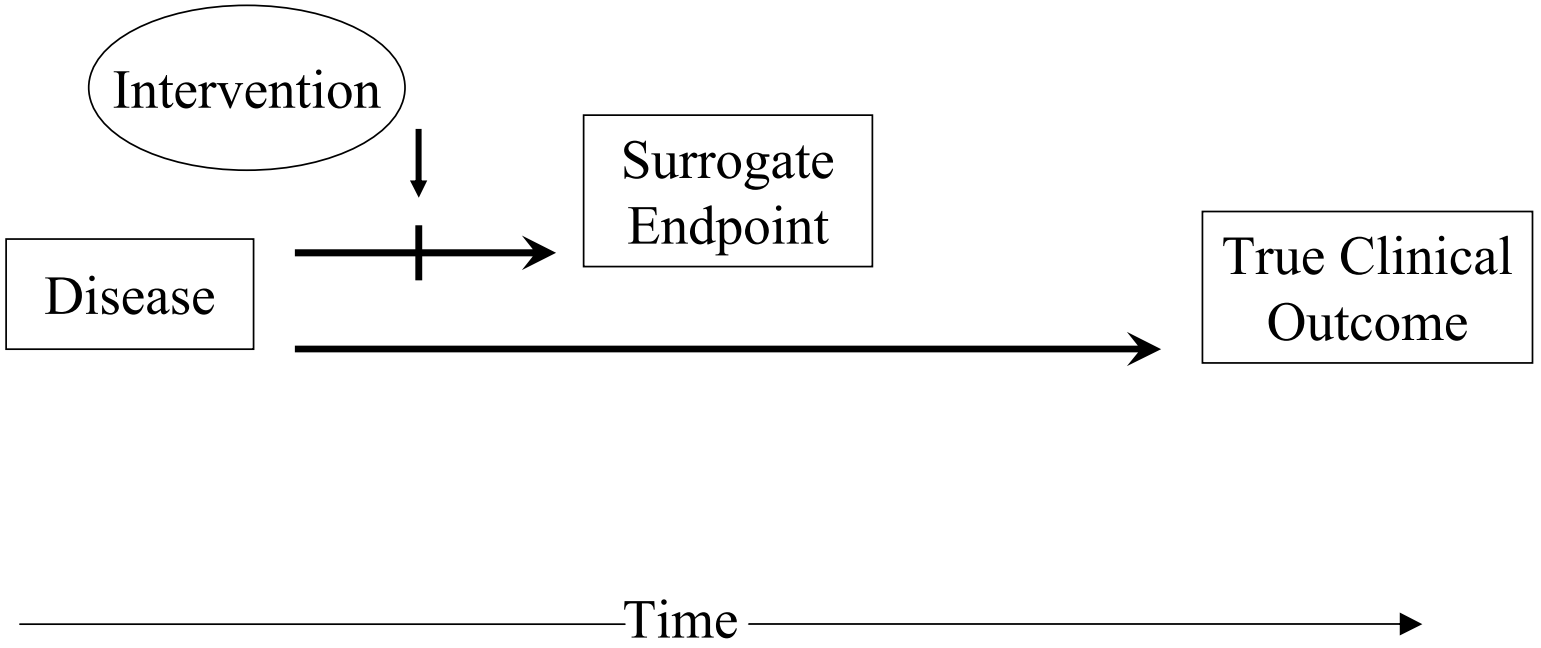
Prentice's Criteria (SIM, 1989)

- To be a direct substitute for a clinical benefit endpoint on inferences of superiority and inferiority
  - The surrogate endpoint must be correlated with the clinical outcome
  - The surrogate endpoint must fully capture the net effect of treatment on the clinical outcome

Validation of Surrogate Outcomes

Does Not Satisfy Criterion

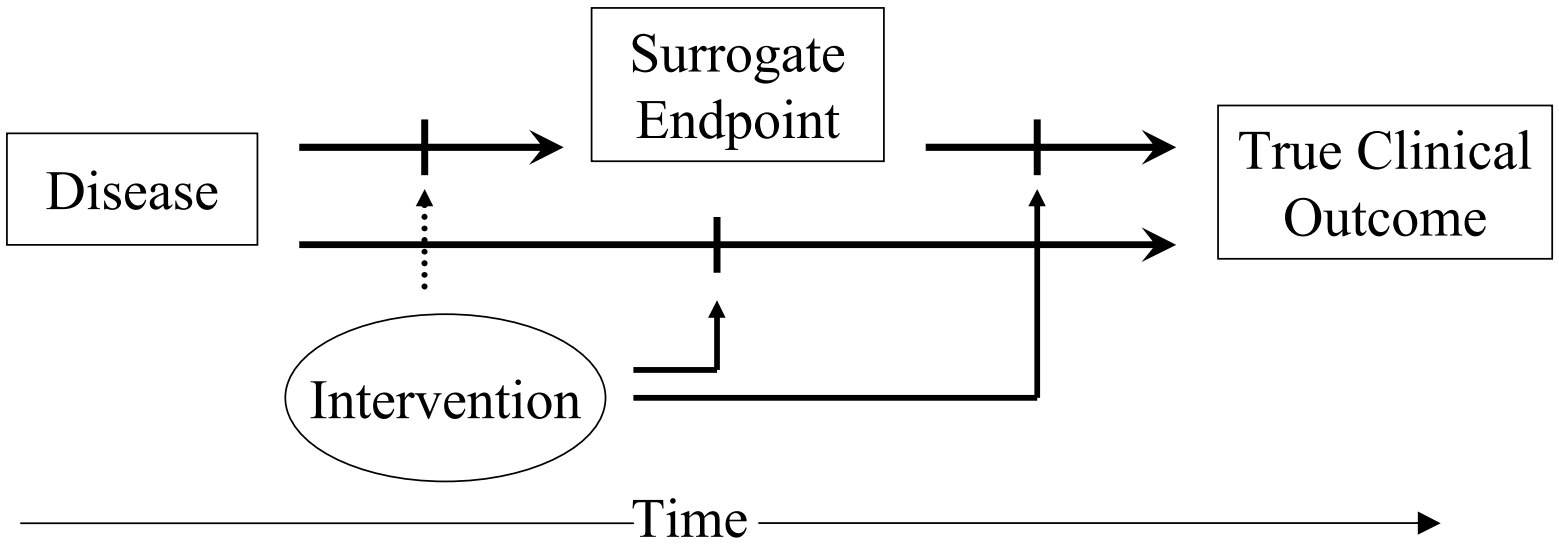
- Treatment has no effect on Clinical Outcome



Validation of Surrogate Outcomes

Does Not Satisfy Criterion

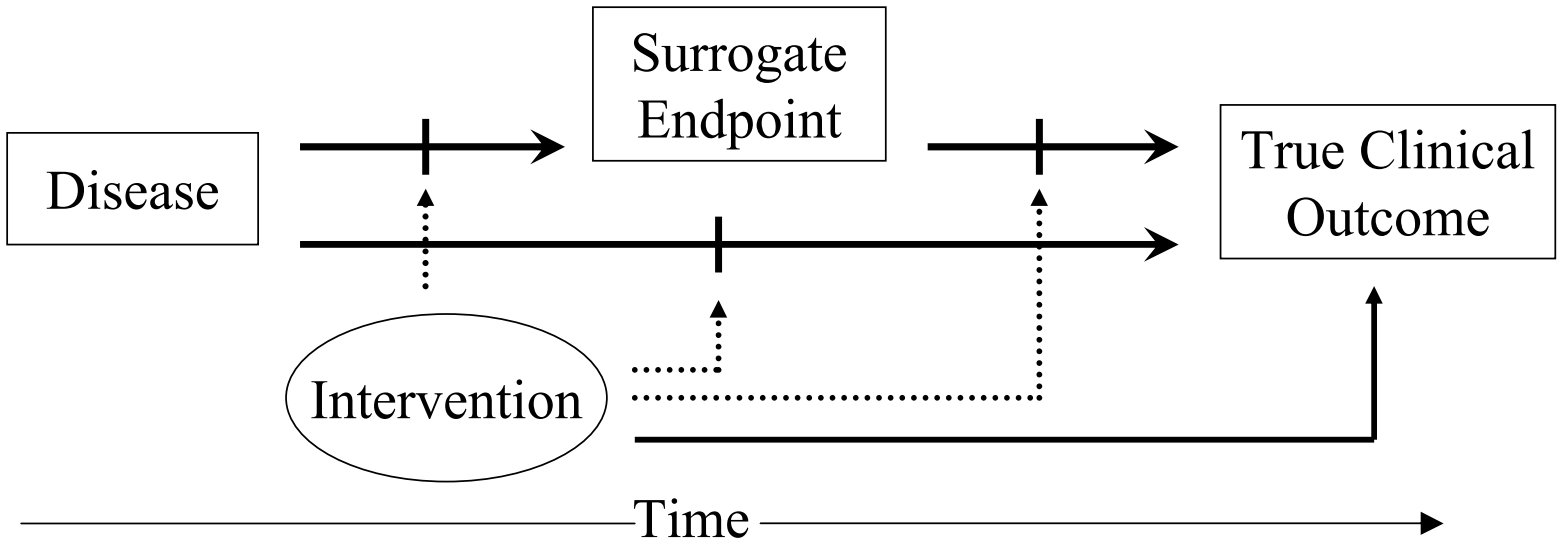
- Adjusting for Surrogate Endpoint will not capture all of Treatment effect



Validation of Surrogate Outcomes

Does Not Satisfy Criterion

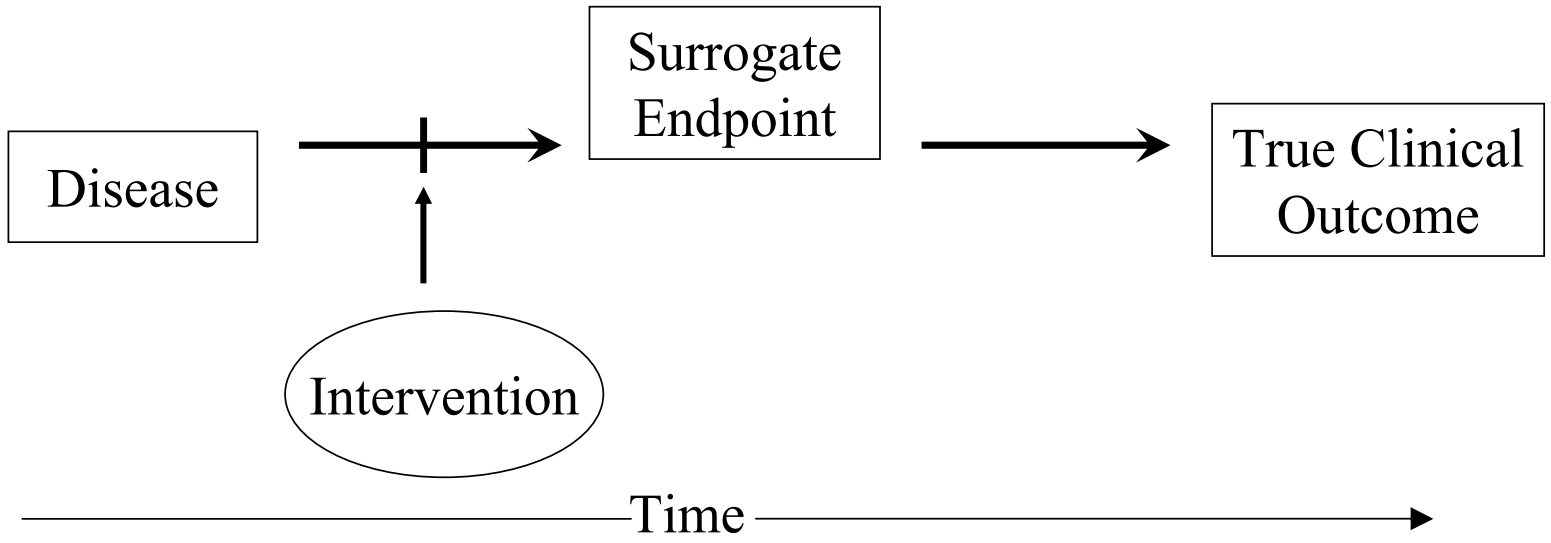
- Adjusting for Surrogate Endpoint will not capture all of Treatment effect on Clinical Outcome



Validation of Surrogate Outcomes

Satisfies Criterion

- Adjusting for Surrogate Endpoint will remove effect of Treatment on Clinical Outcome



Validation of Surrogate Outcomes

What is the implication?

- The validity of a surrogate endpoint is dependent upon
  1. the disease
  2. the clinical outcome
  3. the treatment
- Thus it is not possible to validate a surrogate endpoint for every combination of treatment and disease without doing a trial looking at the clinical outcome

Validation of Surrogate Outcomes

What is the implication?

- When considering a number of treatments that can be presumed to act in a similar manner, meta-analyses of clinical trial results can sometimes be used to establish the suitability of a surrogate endpoint for other treatments in that class
  - Even then, we must watch for outliers within such a meta-analysis
  - Such outliers suggest that the presumption of similar action is violated

Validation of Surrogate Outcomes

At the end of the day

- Surrogate endpoints have a place in screening trials where the major interest is identifying treatments which have little chance of working
- But for confirmatory trials meant to establish beneficial clinical effects of treatments, use of surrogate endpoints can (AND HAS) led to the introduction of harmful treatments

Introduction to Clinical Trials - Day 2

Session 3 - Methods of Randomization

Presented July 24, 2018

Susanne J. May
Department of Biostatistics
University of Washington

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

Why randomization?

Consider the scientific objective

- ICH guidelines (www.ich.org) part E9 Statistical Principles

  “The most important design techniques for avoiding bias in clinical trials are blinding and randomisation, and these should be normal features of most controlled clinical trials intended to be included in a marketing application."

- Similar criteria are required in the CONSORT guidelines.

Bias

What is bias?

- Bias is a tendency of a statistical estimate to deviate in one direction from a“true value"
- What defines the “truth" is dictated by the scientific goal
- Randomization is the primary tool of a clinical trialist for reducing bias
- In order to illustrate the role in which bias arises in clinical studies and motivate the role of randomization, it is useful to review the components of a statistical model in order to:
  1. Develop a standard nomenclature
  2. Illustrate the goals and impact of randomization
- To this end, we can begin withe role of adjustment variables in statistical models

Example - FEV Data

Is there an association between smoking and lung function in children?

- Scientific justification
  - Longterm smoking is associated with lower lung function
  - Are similar effects observed in short term smoking in children?
- Causal pathway of interest
  - Interested in whether smoking will cause a decrease in lung function

Smoking ⟶ Lung function

Example - FEV Data

Study design

- Observational study
  - Measurements obtained on a sample of 654 healthy children
  - Children were sampled while being seen for a regular checkup
  - Data available on smoking, age, gender, and height
  - Predictor of interest: Self-reported smoking
  - Response: FEV (Forced Expository Volume)

FEV Data

SMOKERS

1.953 2.236 3.428 3.208 1.694 3.957 4.789 2.384 3.074 2.387 3.835 2.599 4.756 3.086 4.309 3.413 2.975 3.169 3.343 3.751
2.216 3 .078 3.186 3.297 2.304 3.102 2.677 3.297 3.498 2.759 2.953 3.785 2.276 4.637 3.038 3.120 3.339 3.152 3.104 4.045
4.763 3.069 4.506 3.519 3.688 2.679 2.198 3 .345 3.082 2.903 3.004 3.406 3.122 3.330 2.608 3.799 4.086 4.070 2.264 4.404
2.278 4.872 4.270 3.727 2.795

NONSMOKERS

1.708 1.724 1.720 1.558 1.895 2.336 1.919 1.415 1.987 1.942 1.602 1.735 2.193 2.118 2.258 1.932 1.472 1.878 2.352 2.604
1.400 1 .256 0.839 2.578 2.988 1.404 2.348 1.755 2.980 2.100 1.282 3.000 2.673 2.093 1.612 2.175 2.725 2.071 1.547 2.004
3.135 2.420 1.776 1.931 1.343 2.076 1.624 1 .344 1.650 2.732 2.017 2.797 3.556 1.703 1.634 2.570 3.016 2.419 1.569 1.698
2.123 2.481 1.481 1.577 1.940 1.747 2.069 1.631 1.536 2.560 1.962 2.531 2.715 2 .457 2.090 1.789 1.858 1.452 3.842 1.719
2.111 1.695 2.211 1.794 1.917 2.144 1.253 2.659 1.580 2.126 3.029 2.964 1.611 2.215 2.388 2.196 1.751 2.165 1.682 1 .523
1.292 1.649 2.588 0.796 2.574 1.979 2.354 1.718 1.742 1.603 2.639 1.829 2.084 2.220 1.473 2.341 1.698 1.196 1.872 2.219
2.420 1.827 1.461 1.338 2.090 1 .697 1.562 2.040 1.609 2.458 2.650 1.429 1.675 1.947 2.069 1.572 1.348 2.288 1.773 0.791
1.905 2.463 1.431 2.631 3.114 2.135 1.527 2.293 3.042 2.927 2.665 2 .301 2.460 2.592 1.750 1.759 1.536 2.259 2.048 2.571
2.046 1.780 1.552 1.953 2.893 1.713 2.851 1.624 2.631 1.819 1.658 2.158 1.789 3.004 2.503 1.933 2.091 2 .316 1.704 1.606
1.165 2.102 2.320 2.230 1.716 1.790 1.146 2.187 2.717 1.796 1.335 2.119 1.666 1.826 2.709 2.871 1.092 2.262 2.104 2.166
1.690 2.973 2.145 1 .971 2.095 1.697 2.455 1.920 2.164 2.130 2.993 2.529 1.726 2.442 1.102 2.056 1.808 2.305 1.969 1.556
1.072 2.042 1.512 1.423 3.681 1.991 1.897 1.370 1.338 2 .016 2.639 1.389 1.612 2.135 2.681 3.223 1.796 2.010 1.523 1.744
2.485 2.335 1.415 2.076 2.435 1.728 2.850 1.844 1.754 1.343 2.303 2.246 2.476 3.239 2.457 2 .382 1.640 1.589 2.056 2.226
1.886 2.833 1.715 2.631 2.550 1.912 1.877 1.935 1.539 2.803 2.923 2.358 2.094 1.855 1.535 2.135 1.930 2.182 1.359 2.002
1.699 2 .500 2.366 2.069 1.418 2.333 1.514 1.758 2.535 2.564 2.487 1.591 1.624 2.798 1.691 1.999 1.869 1.004 1.427 1.826
2.688 1.657 1.672 2.015 2.371 2.115 2.328 1 .495 2.884 2.328 3.381 2.170 3.470 3.058 1.811 2.524 2.642 3.741 4.336 4.842
4.550 2.841 3.166 3.816 2.561 3.654 2.481 2.665 3.203 3.549 3.222 3.111 3.490 3 .147 2.520 2.292 2.889 2.246 1.937 2.646
2.957 4.007 2.386 3.251 2.762 3.011 4.305 3.906 3.583 3.236 3.436 3.058 3.007 3.489 2.864 2.819 2.250 4.683 2.352 3 .108
3.994 4.393 2.592 3.193 2.346 3.515 2.754 2.720 2.463 2.633 3.048 3.111 3.745 2.094 3.183 3.977 3.354 3.411 3.171 3.887
2.646 2.504 3.587 3.845 2.971 2 .891 1.823 2.417 2.175 2.735 4.273 2.976 4.065 2.318 3.596 3.395 2.751 2.673 2.556 2.542
2.608 2.354 1.458 3.795 2.491 3.060 2.545 2.993 3.305 3.774 2.855 2 .988 2.498 3.169 2.887 2.704 3.515 3.425 2.287 2.434
2.365 2.696 2.868 2.813 3.255 4.593 4.111 1.916 1.858 3.350 2.901 2.241 4.225 3.223 5.224 4.073 4.080 2 .606 4.411 3.791
3.089 2.465 3.200 2.913 4.877 2.358 3.279 2.581 2.347 2.691 2.827 1.873 2.538 2.758 3.050 3.079 2.201 1.858 3.403 3.501
2.578 1.665 2.081 2 .974 4.073 4.448 3.984 2.250 2.752 3.680 2.862 3.023 3.681 3.255 3.692 2.356 4.591 3.082 3.258 2.216
3.247 4.324 2.362 2.563 3.206 3.585 4.720 3.331 5.083 2 .417 2.364 2.341 3.231 3.078 3.369 3.529 2.866 2.891 3.022 3.127
2.866 2.605 3.056 2.569 2.501 3.320 2.123 3.780 3.847 3.924 2.132 2.752 2.449 3.456 3.073 2 .688 3.329 4.271 3.530 2.928
2.689 2.332 2.934 3.110 2.894 2.435 2.838 3.035 4.831 2.812 2.714 3.086 3.519 4.232 2.770 3.341 3.090 2.531 2.822 2.935
2.568 2 .387 2.499 4.130 3.001 3.132 3.577 3.222 3.280 2.659 2.822 2.140 4.203 2.997 2.562 3.082 3.806 2.458 2.391 3.141
2.579 2.100 2.785 4.284 2.906 5.102 4.429 4 .279 4.500 2.635 3.082 3.387 5.793 3.985 4.220 4.724 3.731 3.500 3.674 5.633
3.645 2.887 3.960 4.299 2.981 4.504 5.638 2.853 3.211

Example - FEV Data

Interpretation of smoking effect in unadjusted analysis

- Restrict sample to children 9 years and above (age of youngest smoker in sample)
- Consider log-transformation of FEV based upon past studies
  - Scientific focus on median FEV
  - Distribution of log-transformed FEV approximately symmetric

Example - FEV Data

Unadjusted association between smoking and FEV

- Consider an unadjusted comparison of FEV between smokers and non-smokers
  - Unadjusted Result: The median FEV of a smoker is estimated to be 10.8% higher than that of a non-smoker (95% CI: 1.04, 1.18). This difference is statistically significant $p = 0.002$.

Example - FEV Data

Adjustment for age

- Consider adjustment for age in a linear regression model
  - Age-adjusted result: The median FEV of a smokers is estimated to be 5.0% lower than that of non-smokers *similar in age* (95% CI: 0.90, 1.01). This difference is not statistically significant at the .05 level ($p = 0.093$).

Example - FEV Data

Adjustment for age and height

- After adjustment for age, height should have little association with smoking status but is still likely to have an association with FEV.
- Consider additional adjustment for height...
  - Age and height-adjusted result: The median FEV of smokers is estimated to be 5.2% lower than that of non-smokers *similar in age and height* (95% CI: 0.91, 0.99). This difference is statistically significant at the .05 level ($p = 0.011$).

Example - FEV Data

Comparison of age and age-height adjusted analyses

- Notice that there is little difference in estimated effect of smoking between age adjusted models with and without height
- Effect of height adjustment on precision
  - Lower Root MSE (.144 vs .209) in height adjusted model resulting in increased precision of estimate of smoking effect
  - Net effect: Much greater precision (SE 0.021 vs 0.031)

Example - FEV Data

Take-home message

- Our scientific question was not

  "Is there a difference between smokers' and nonsmokers' median FEV?"

- But rather

  "Do smokers have lower median FEV than otherwise comparable nonsmokers?"

Example - FEV Data

Take-home message

- This example highlights:
  1. How a scientific question should dictate a chosen statistical model
  2. The role of a *confounding* variable on association estimates
  3. The impact that adjustment has on the precision of association estimates
- These ideas provide the motivation for randomization, as well as the types and implementation of various randomization methods
- However, before going there, it is useful to define the statsitical role of variables and to generalize the observations that were made in the FEV example...

Statistical role of variables

Effect modifiers (interaction terms)

- Suppose that we are interested in modeling the association between an outcome variable Y and a predictor X
- Consider four broad categories of variables (this terminology is not universal)
- Effect modifiers (interaction variables)
  - An effect modifier (W) is a covariate for which the association between the predictor of interest (X) *and* the outcome of interest (Y) differs with each level of W

Statistical role of variables

Example: Effect modification

- Example: The association between gender and the risk of chd differs by systolic blood pressure

sbpgrp	Odds Ratio	chi2(1)	P>chi2	[95% Conf.	Interval]
1	0.394493	86.23	0.0000	0.32186	0.48351
2	0.429583	56.59	0.0000	0.34243	0.53892
3	0.597384	9.91	0.0016	0.43193	0.82621
4	0.741269	1.75	0.1858	0.47495	1.15693

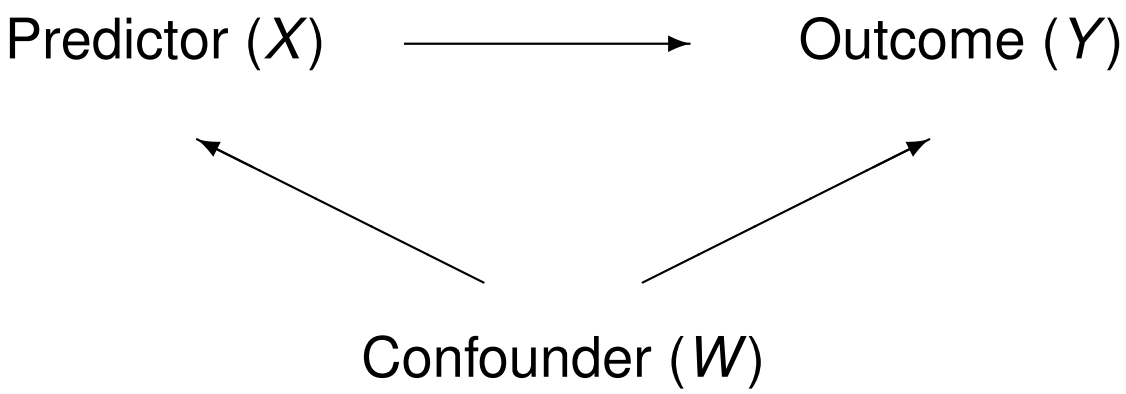
How do we deal with effect modifiers?

- When the scientific question involves effect modification, analyses must be within each stratum separately

Statistical role of variables

Confounders

- One definition: A **confounder** is a variable that is causally related to the predictor of interest (X) *and* the outcome of interest (Y).



Statistical role of variables

Example: Confounding

- Example: Age in the FEV example:
  - Older kids tend to smoke
  - Older kids tend to have larger lungs

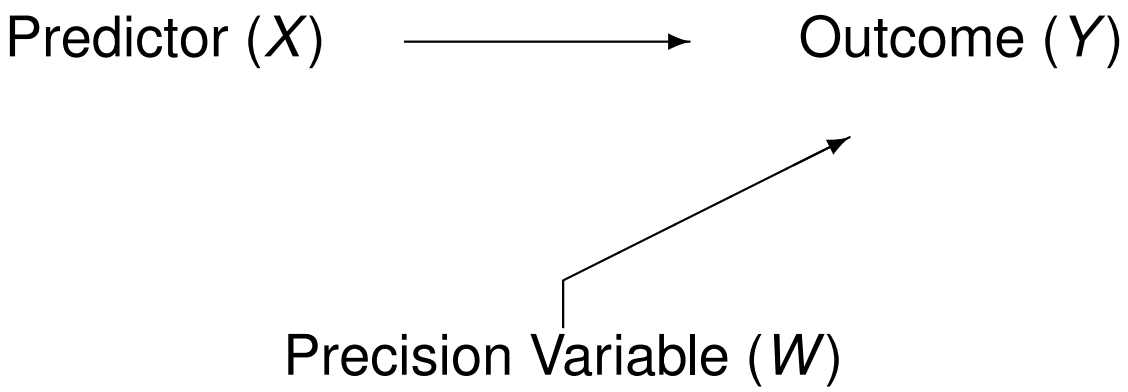
How do we deal with confounding?

- Adjust for the confounder

Statistical role of variables

Precision variables

- I define a **precision variable** as a covariate that is related to the outcome Y, but independent of the predictor of interest X.



Statistical role of variables

Example: Precision variable

- Example: Height (after adjustment for age) in the FEV example:
  - Conditional on age, little difference in prevalence of smoking by height
  - Conditional on age, taller kids tend to have larger lungs

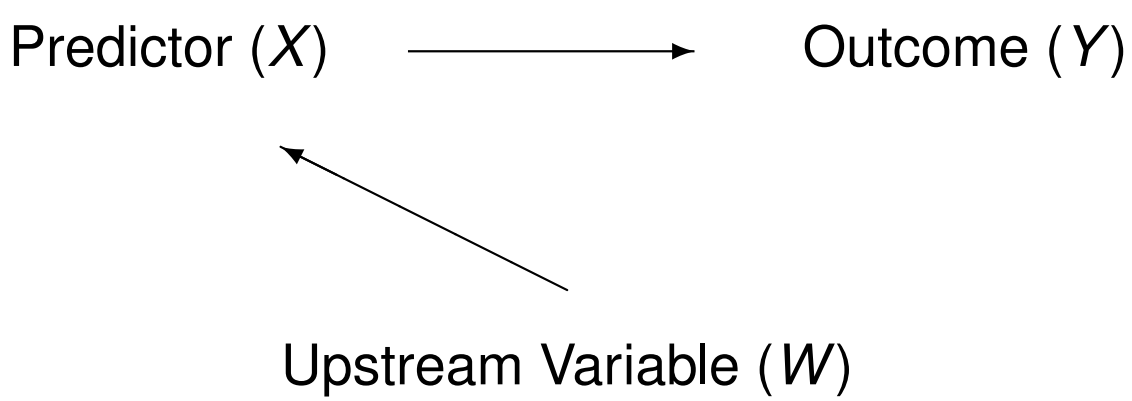
How do we deal with precision variables?

- Often a good idea to control for them
- For example, in a two sample comparison of means, we might control some variable in order to decrease the within group variability

Statistical role of variables

“Upstream" variables

- I define an **upstream variable** as a covariate that is independent of the outcome Y, but may or may not be related to the predictor of interest X.



- Generally a bad idea to adjust for “upstream" variables

Statistical role of variables

Why randomize?

- The fundamental statistical distinctions between unadjusted and adjusted regression models are central to the goals of randomization
- We thus want to be able to consider the relationships between
  - unadjusted and adjusted parameters, and
  - the standard errors of the two parameter estimates
- This is easily done in the context of linear regression and that will be the setting for our discussion
  - Results are less straightforward for non-linear models (eg. logistic regression or proportional hazards)
  - However, the general principles still apply

Statistical role of variables

Adjusted vs. unadjusted covariate effects

- Consider the following linear regression models:

1. Unadjusted model: $\mathrm{E}[Y_i] = \beta_0 + \beta_1 X_i$
   - β_1 is the difference in the mean of Y for groups differing by 1-unit in X
2. Adjusted model: $\mathrm{E}[Y_i] = \gamma_0 + \gamma_1 X_i + \gamma_2 W_i$
   - γ_1 is the difference in the mean of Y for groups differing by 1-unit in X, but agreeing in their value of W

Adjusted vs. unadjusted covariate effects

- Proposition 1: Let $\hat{\beta}_1$ denote the OLS estimate of β_1. Then under the adjusted model,

$$\mathrm{E}[\hat{\beta}_1] = \gamma_1 + \frac{\mathrm{cov}(X, W)}{\mathrm{var}(X)}\gamma_2$$

$$= \gamma_1 + r_{XW}\sqrt{\frac{\mathrm{var}(W)}{\mathrm{var}(X)}}\gamma_2$$

where r_{XW}, $\mathrm{var}(X)$, and $\mathrm{var}(W)$ are the sample correlation between X and W, sample variance of X, and sample variance of W, respectively.

Statistical role of variables

The implication...

- $\hat{\beta}_1$ is biased (and inconsistent) for γ_1 unless at least one of the following hold

  1. $r_{XW} = 0$: X and W are uncorrelated (in the sample), OR
  2. $\gamma_2 = 0$: W is not related to Y

- In either case, $\hat{\beta}_1$ is unbiased (and consistent) for β_1

- Implication for confounders?

  - By definition, a confounder is related to the predictor of interest and the response

  - This implies that if W is a confounder, then both conditions above fail

  - Hence the parameter from the reduced model is biased for the adjusted estimate

Precision of Estimators

Relationship between the precision of unadjusted and adjusted estimates

- Consider the following linear regression models:

1. Unadjusted model: $\mathrm{E}[Y_i] = \beta_0 + \beta_1 X_i$
2. Adjusted model: $\mathrm{E}[Y_i] = \gamma_0 + \gamma_1 X_i + \gamma_2 W_i$

Precision of Estimators

Relationship between the precision of unadjusted and adjusted estimates

- Proposition 2:
  1. For the unadjusted model,

$$\text{Var}[\hat{\beta}_1] = \frac{\sigma^2_{Y|X}}{n\text{var}(X)}$$

  2. For the adjusted model,

$$\text{Var}[\hat{\gamma}_1] = \frac{\sigma^2_{Y|X,W}}{n\text{var}(X)(1 - r^2_{XW})}$$

  where $\sigma^2_{Y|X,W} = \sigma^2_{Y|X} - \gamma_2^2 \text{var}(W|X)$

- Hence, if $\gamma_2 \neq 0$ then $\sigma^2_{Y|X,W} < \sigma^2_{Y|X}$

Statistical role of variables

Implications of Propositions 1 & 2 (generalizeable to p coviarate case)

- Case 1: $r_{XW} = 0$ (X and W uncorrelated) and $\gamma_2 = 0$ (W and Y unrelated)
  - From Proposition 1, $\hat{\beta}_1$ unbiased for γ_1
  - From Proposition 2, $\text{Var}[\hat{\beta}_1] = \text{Var}[\hat{\gamma}_1]$
  - Conclusion: Lose 1 degree of freedom for hypothesis tests and CIs if adjusting for W

Statistical role of variables

Implications of Propositions 1 & 2 (generalizeable to p coviarate case)

- Case 2: $r_{XW} \neq 0$ (X and W correlated) and $\gamma_2 = 0$ (W and Y unrelated)
  - From Proposition 1, $\hat{\beta}_1$ unbiased for γ_1
  - From Proposition 2, $\text{Var}[\hat{\beta}_1] < \text{Var}[\hat{\gamma}_1]$
  - Conclusion: Mathematically estimating the same quantity but *lose* precision when adjusting for W (nuisance variable)

Statistical role of variables

Implications of Propositions 1 & 2 (generalizeable to p coviarate case)

- Case 3: $r_{XW} = 0$ (X and W uncorrelated) and $\gamma_2 \neq 0$ (W and Y related)
  - From Proposition 1, $\hat{\beta}_1$ unbiased for γ_1
  - From Proposition 2, $\text{Var}[\hat{\beta}_1] > \text{Var}[\hat{\gamma}_1]$
  - Conclusion: Mathematically estimating the same quantity but *gain* precision when adjusting for W (precision variable)

Statistical role of variables

Implications of Propositions 1 & 2 (generalizeable to p coviarate case)

- Case 4: $r_{XW} \neq 0$ (X and W correlated) and $\gamma_2 \neq 0$ (W and Y related)
  - From Proposition 1, $\hat{\beta}_1$ biased for γ_1
  - From Proposition 2, no definitive statement about the variances
  - Conclusion: W is a confounder and decision to adjust should be based on what you are trying to estimate.

Statistical role of variables

Why do we care?

- The above results provide the fundamental motivation for

  1. The use and types of randomization (balance of confounders)

  2. The consideration of analytic methods under various types of randomization

Methods of Randomization

Cause and Effect

- Necessary conditions for establishing cause and effect of a treatment

  1. The treatment should precede the effect

     - Beware protopathic signs (eg. Marijuana and risk of MI within 3 hours)

  2. When comparing groups differing in their treatment, the groups should be comparable in every other way (at baseline) (see previous discussion on confounding)

Methods of Randomization

Cause and Effect

Randomization is the major way in which cause and effect is established

- Ensures comparability of populations
  - Each treatment group drawn from same population
  - Differences in other prognostic factors will only differ by random sampling
    - Provides balance on the total effect of all other prognostic factors
    - May not provide balance on each individual factor
- Note: Sequential allocation of patients is not randomization
  - Possible time trends in recruitment, treatments, etc.

Nonadaptive Randomization

General statements on randomization

- Randomization is our friend...
  - If we randomize, we do not (on average) need to worry about differences between the treatment groups with respect to factors present at time of randomization
    - Any difference in outcomes can be attributed to treatment
    - However, recognize that treatment can lead to differential use of other ancillary treatments
- But like all friends, we must treat it with respect.
  - We must analyze our data in groups defined at the time of randomization
    - Discarding or missing data on randomized subjects may lead to bias (It certainly leads to diminished scientific credibility)

Nonadaptive Randomization

Impact on data analysis

In presence of randomized treatment assignment

- Intent to treat analysis (ITT)
  - Based on randomization
- Confounding not an issue (on average)
  - P value measures probability of observed effects occurring due only to randomization imbalance
- Gain precision if
  - Control important predictors, or
  - Adjust for stratification variables
- Subgroup analyses
  - If effect modification is concern
  - Pre-specification

Nonadaptive Randomization

Randomization strategies

- Complete randomization (CRD)
- Blocked randomization
  - Ensure balance after every k patients
  - Ensure closer adherence to randomization ratio
  - Undisclosed block sizes to prevent bias
- Stratified randomization
  - Separately within strata defined by strong risk factors
    - Lessens chance of randomization imbalance
  - Need to consider how many variables can be used
- Dynamic randomization
  - Adaptive randomization to achieve best balance on marginal distribution of covariates

Nonadaptive Randomization

Complete randomization

- The simplest form of randomization is independent randomization of each individual
- With each accrued subject a (possibly biased) coin is tossed to determine which arm
  - Probability of treatment arm = $r/(r+1)$
  - Independence of successive randomizations
- Possible issues with complete randomization include
  - Bias,
  - Face validity, and
  - Precision

Nonadaptive Randomization

Complete randomization

- On average (across repeated experiments)
  - No correlation between treatment variable and other covariates
  - Individual type I errors come from samples in which other covariates are imbalanced

$$\mathrm{E}[\hat{\beta}_1] = \gamma_1 + \frac{\mathrm{cov}(X, W)}{\mathrm{var}(X)}\gamma_2$$

$$= \gamma_1 + r_{xw}\sqrt{\frac{\mathrm{var}(W)}{\mathrm{var}(X)}}\gamma_2$$

Nonadaptive Randomization

Complete randomization

- Typical to consider face validity of randomization in a “Table 1"

	Methotrexate Arm		Placebo Arm	
	n	Mean (SD; Min – Max)	n	Mean (SD; Min – Max)
Age (yrs)	132	50.4 (8.5; 32 - 69)	133	52.2 (8.5; 26 - 67)
Female	132	92.4%	133	92.5%
Pruritus score	116	7.7 (3.8; 4 - 16)	124	6.9 (3.8; 4 - 20)
Splenomegaly	131	8.4%	133	10.5%
Telangiectasia	132	4.6%	133	11.3%
Edema	132	6.1%	133	3.0%
Alkaline phosphatase	132	242.6 (145.9; 53 - 933)	133	245.0 (187.6; 66 - 1130)
ALT	131	54.5 (41.7; 12 - 202)	132	50.6 (41.4; 12 - 311)
Total bilirubin	132	0.7 (0.4; 0.1 - 2.7)	133	0.7 (0.4; 0.1 - 2.4)
Albumin	132	4.0 (0.3; 3.1 - 6.0)	133	4.0 (0.3; 3.0 - 4.8)
Prothrombin time INR	124	1.0 (0.1; 0.7 - 1.3)	132	1.0 (0.1; 0.7 - 1.3)
Mayo score	128	3.8 (0.8; 1.6 - 6.3)	133	3.9 (0.8; 1.6 - 6.1)

Nonadaptive Randomization

Complete randomization

- Consider differences in baseline stoke severity in a multi-center randomized clinical trial comparing tissue plasminogen activator (tPA) for the treatment of acute ischemic stroke
  - Percentage of patients (N = 320) in the 91 to 180-minute subgroups with a specific baseline National Institutes of Health Stroke Scale (NIHSS) score (Marler et al., *Neurology*, 2000)

Baseline NIHSS score	*tPA-treated patients, % (n = 153)*	*Patients given placebo, % (n = 167)*
0-5	19.0	4.2
6-10	24.2	27.5
11-15	17.0	21.0
16-20	21.6	19.8
>20	18.3	27.5
tPA = tissue plasminogen activator		

"The marked imbalance in baseline stroke severity in the 91 to 180-minute groups of the NINDS trial suggests that the NINDS trial lacks internal validity." -Mann, *West J. of Med* (2002)

Nonadaptive Randomization

Complete randomization

- Table 1: Potential for imbalance in covariates
  - Depends on number of covariates and correlations among them
  - Probability of at least one “significant" imbalance

Number	Worst	Correlation			
Displayed	Case	0.00	0.30	0.50	0.75
1	.050	.050	.050	.050	.050
2	.100	.098	.095	.090	.081
3	.150	.143	.137	.126	.104
5	.250	.226	.208	.184	.138
10	.500	.401	.353	.284	.193
20	1.000	.642	.540	.420	.258
50	1.000	.923	.806	.624	.353

Nonadaptive Randomization

Complete randomization

- Of course, statistical significance is not the issue
- The real concern is “conditional confounding"
  - How does unadjusted estimate compare to adjusted estimate?
  - Product of sample correlation between X (treatment) and W (potential confounder) and adjusted association between Y (outcome) and W

$$\mathrm{E}[\hat{\beta}_1] = \gamma_1 + \frac{\mathrm{cov}(X, W)}{\mathrm{var}(X)}\gamma_2$$

$$= \gamma_1 + r_{XW}\sqrt{\frac{\mathrm{var}(W)}{\mathrm{var}(X)}}\gamma_2$$

Nonadaptive Randomization

Complete randomization

- Spurious results due to covariate imbalance
  - Unconditionally: Unbiased so no problem
    - CONSORT Item 15 : “Although proper random assignment prevents selection bias, it does not guarantee that groups are equivalent at baseline. Any differences in baseline characteristics are, however, the result of chance rather than bias."
  - Conditional on obtained randomization:
    - IF covariates are strongly predictive of outcome, then covariate imbalance increases type I error
    - But need to consider that combined effect of other measured and unmeasured covariates may provide balance
- Ultimately, however, we need to have credible results
  - We do not always get to choose what others believe

Nonadaptive Randomization

Precision

- Impact of completely randomized design on precision of inference
  - Impact of imbalance in sample sizes
    - The number accrued to each arm is random
  - Impact of imbalance in covariates
    - “One statistician’s mean is another statistician’s variance"

Nonadaptive Randomization

Randomization ratio

- Most efficient
  - When test statistics involve a sum, choose ratio equal to ratio of standard deviations
- Most ethical for patients on study
  - Assign more patients to best treatment
    - Many sponsors / patients presume new treatment
    - (Adaptive randomization: Play the winner)
- Most ethical for general patient population
  - Whatever is most efficient (generally not adaptive)
- Other goals
  - Attaining sufficient patients exposed to new treatment
  - Maintaining DSMB blind

Nonadaptive Randomization

Randomization ratio : Optimal r (fixed n)

- Suppose we are constrained by maximal sample size $n = n_1 + n_2$
- Smallest standard error when

$$r = \frac{n_1}{n_2} = \frac{s_1}{s_2}$$

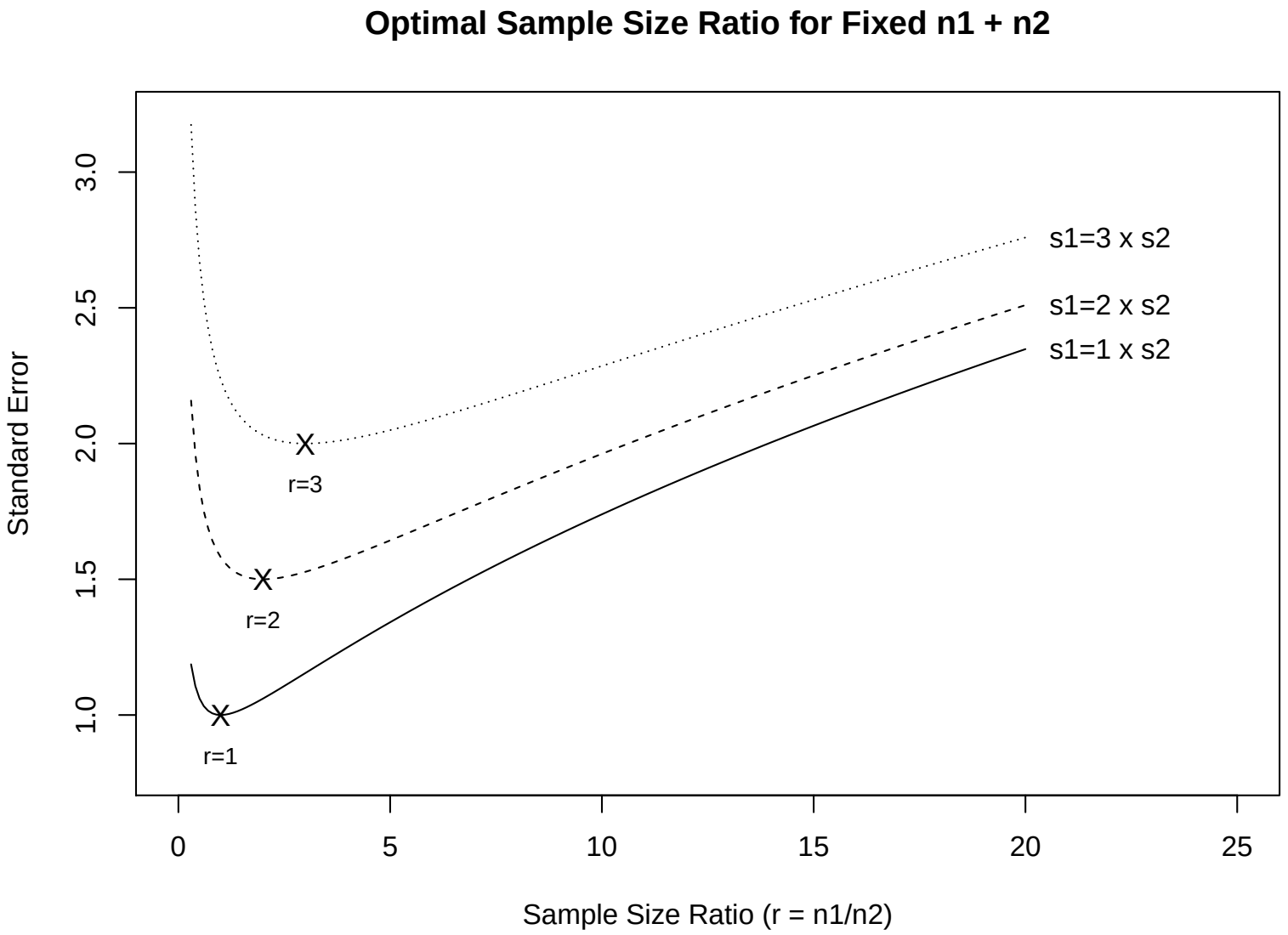
where s_i is the standard deviation of response in group i,

$i = 1, 2$

- When we are unconstrained by maximal sample size we still hit a point of diminishing returns
  - Often quoted: $r = 5$
  - Really depends on ratio of standard deviations...

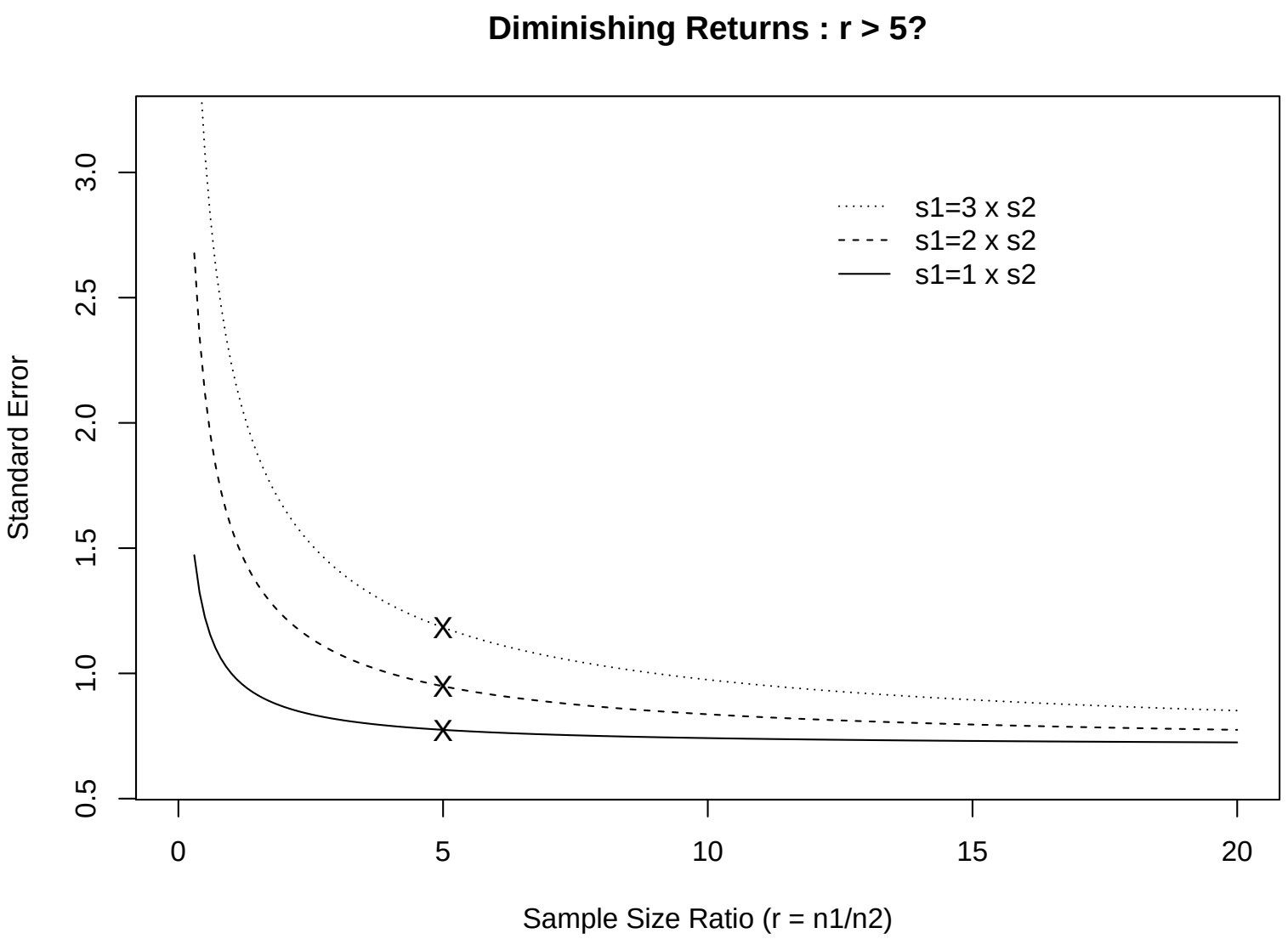
Nonadaptive Randomization

Randomization ratio : Optimal *r* (fixed *n*)



Nonadaptive Randomization

Randomization ratio : Diminishing returns



Nonadaptive Randomization

Complete randomization

- It is possible, in smaller studies, that a completely randomized design with high randomization ratio may not randomize at least two subjects to each arm
- Consider the probability that a CRD may not randomize at least two subjects to each arm as a function of the total trial size and randomization ratio

N	r= 1	r= 2	r= 3	r= 5	r=10
20	0.0000	0.0033	0.0243	0.1304	0.4459
50	0.0000	0.0000	0.0000	0.0012	0.0511
100	0.0000	0.0000	0.0000	0.0000	0.0008
200	0.0000	0.0000	0.0000	0.0000	0.0000
500	0.0000	0.0000	0.0000	0.0000	0.0000
1000	0.0000	0.0000	0.0000	0.0000	0.0000

Efficiency loss from imbalance

- Covariates may be imbalanced across arms
  - Variability across replicated experiments increased if important predictor not controlled
  - Recall

$$\text{Var}[\hat{\beta}_1] = \frac{\sigma^2_{Y|X}}{n\text{var}(X)}$$

with

$$\sigma^2_{Y|X} = \gamma_2^2 \text{var}(W|X) + \sigma^2_{Y|X,W}$$

Nonadaptive Randomization

How to improve performance?

- If we adjust for important covariates, we will often gain precision
  - Face validity in Table 1 if readers recognize that adjustment accounts for any observed imbalance
- Caveats:
  - If covariate imbalance by arm, model misspecification can be an issue regarding conditional bias
  - If covariate imbalance by arm, lack of effect can be an issue regarding variance inflation
  - If adjustment not TOTALLY prespecified, “intent to cheat" analysis can be an issue
    - Loss of precision from imperfect model should not be too much of an issue in most situations

Nonadaptive Randomization

Issues with complete randomization

- Imbalance across arms in sample sizes
  - Not much of an issue with large sample sizes
  - Could be problematic with sequential sampling
    - Interim analyses of data early in the study
- Imbalance across arms in time trends
  - Outcome may be associated with time of accrual
- *Blocking* is sometimes used to ensure
  - Proper ratio of sample sizes across groups, and
  - Balance across arms over time

Nonadaptive Randomization

Mechanisms leading to time trends

- Patients accrued early may differ from those accrued later, because
  - Backlog of eligible patients
  - Startup of new clinical sites
  - Pressure to increase accrual
  - Secular trends in beliefs about intervention
    - (Made much worse if any interim results leak out)
  - Secular trends in diagnostic tools used for eligibility
  - Secular trends in ancillary treatments

Nonadaptive Randomization

Mechanisms leading to time trends

- Within every sequence of k patients, the ratio of treatment to control is exactly $r : 1$
  - Within each “block" ordering of treatments is random
- Important caveats:
  - Investigators must not know block size
    - Otherwise, decisions to enroll patients might be affected by knowledge of next assignment
  - Hence, often use “concealed blocks of varying sizes" (often termed a “random block design")

Alternative strategy : Urn Model

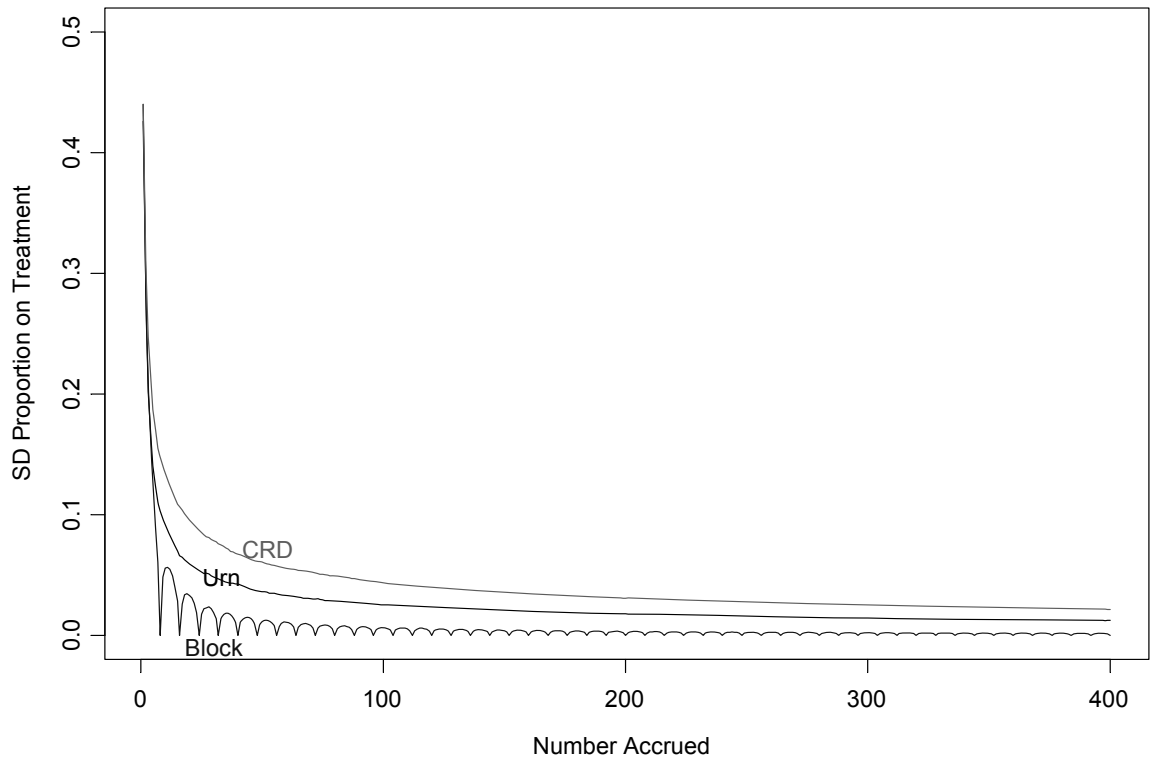
1. Begin with k white balls and $r \times k$ black balls in an urn
2. Upon accrual of a patient draw a ball from urn
   - White $\rightarrow$ control; black $\rightarrow$ treatment
   - After every white ball withdrawn, return 1 white ball and $r \times m$ black balls
   - After every r-th black ball withdrawn, return r black balls and m white balls

- Such a strategy tends to behave like small blocks early and complete randomization later, depending on k and m

Nonadaptive Randomization

Comparison of blocking strategies

- SD proportion on treatment for 3:1 randomization
  - Urn ($k = 1, m = 1$) vs Blocking (size = 8) vs CRD



Statistical inference after blocking

- Impact on statistical inference relative to CRD
  - Bias properties unchanged
  - Face validity largely unchanged
    - We rarely report accrual patterns over time
  - Precision slightly improved due to achieving closer to desired randomization ratio
  - Precision could be improved if adjust for blocks as a random effect in analysis
    - This is rarely done, except in re-randomization test

Nonadaptive Randomization

Issues with complete randomization

- Imbalance across arms in covariate distribution
  - Loss of face validity
  - Conditional bias
  - Not much of an issue with large sample sizes
  - Could be problematic with sequential sampling
    - Interim analyses of data early in the study
  - Could be problematic with subgroup analyses
    - Possibility of very inefficient randomization ratio in small subgroups
- *Stratified randomization* is often used to ensure proper ratio of sample sizes across subgroups defined by important covariates

Stratified randomization

- Strata are defined based on values of important covariates
  - E.g., sex, age, disease severity, clinical site
- Within each stratum defined by a unique combination of stratification variables, CRD or blocked randomization
- Important caveats:
  - Number of strata is exponential in number of stratification variables
    - E.g., 4 two level stratification variables $\Rightarrow$ 16 strata

Nonadaptive Randomization

Statistical inference

- Impact on statistical inference relative to CRD
  - Bias properties unchanged
  - Face validity improved for most important variables
  - Precision improved due to achieving closer to desired randomization ratio
  - Precision could be further improved if adjust for stratification variables in analysis
    - This should be done! (Without adjustment for strata, may even lose power for some alternatives)
    - Requires pre-specification of analysis model to avoid “intent to cheat" analysis

Additional advantages of stratified randomization

- Additional advantages of stratification
  - Balance within clinical center
    - Especially if quality control issues
  - Balance for interim analyse
  - Balance for subgroup analyses
- Also, stratified randomization does not preclude the use of blocking
  - Common to combine the two...blocking within strata

Adaptive Randomization

Issues with stratified randomization

- The need to stratify on all combinations of variables
  - Good news:
    - Balances on interactions as well as main effects
  - Bad news:
    - Effect of interactions might be quite small
    - Really only need to adjust on “counterfactual" outcome based on linear combination of all covariates
- Stratified randomizations has drawbacks in the presence of sparse data
- Because of this, some authors have described dynamic randomization processes that will allow balancing on more covariates

Adaptive Randomization

Dynamic randomization

- Subjects are assigned to the treatment arm that will achieve best balance
  - "Minimization": minimize the difference between the distribution of covariate effects between arms
    - Define a "distance" between arms for covariate vectors
    - Probability of assignment depends upon arm that would provide smallest difference
- Two arms are "distant" based on one of:
  - Randomization ratio very different from $r : 1$ in some stratum
  - Summary measure of distribution of $(W_{i1}, \ldots, W_{ip})$ differs
    - Mean, median, variance, ...
  - Distribution of $(W_{i1}, \ldots, W_{ip})$ differs
  - Contribution of covariates to the outcome differs

Conditional confounding

- Consider unadjusted and adjusted (linear) models for an outcome Y:

  1. Unadjusted model: $\mathrm{E}[Y_i] = \beta_0 + \beta_1 X_i$
  2. Adjusted model: $\mathrm{E}[Y_i] = \gamma_0 + \gamma_1 X_i + \vec{W}_i^T \vec{\delta}$

  or in matrix notation

  1. Unadjusted model: $\mathrm{E}[\vec{Y}] = \mathbf{X}\vec{\beta}$
  2. Adjusted model: $\mathrm{E}[\vec{Y}] = \mathbf{X}\vec{\gamma} + \mathbf{W}\vec{\delta}$

Adaptive Randomization

Conditional confounding

- Then it can be shown that

$$\mathrm{E}[\widehat{\vec{\beta}}] = \vec{\gamma} + (\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T\mathbf{W}\vec{\delta}$$

- This implies that

$$\beta_1 = \gamma_1 + \sum_{j=1}^{p} \left(\bar{W}_{1j\cdot} - \bar{W}_{0j\cdot}\right) \delta_j$$

with

$$\bar{W}_{kj\cdot} = \frac{1}{n_k} \sum_{i=1}^{n} W_{ij} 1_{[X_i=k]}$$

Adaptive Randomization

This provides reasonable ways to define distance metrics

- Based on contribution to confounding :

$$d(\vec{X}, \mathbf{W}) = \left| \sum_{j=1}^{p} \left(\bar{W}_{1j\cdot} - \bar{W}_{0j\cdot} \right) \delta_j \right|$$

- Weighted distance between standardized means :

$$d(\vec{X}, \mathbf{W}) = \sum_{j=1}^{p} c_j \left| \frac{\bar{W}_{1j\cdot} - \bar{W}_{0j\cdot}}{SD(W_j} \right|^{\lambda}$$

- Weighted imbalance in n across strata $\Omega_1, \ldots, \Omega_s$:

$$d(\vec{X}, \mathbf{W}) = \sum_{s=1}^{S} c_s \left| \sum_{i=1}^{n} 1_{[X_i=1]} 1_{[\vec{W}_i \in \Omega_s]} - \sum_{i=1}^{n} 1_{[X_i=0]} 1_{[\vec{W}_i \in \Omega_s]} \right|^{\lambda}$$

Adaptive Randomization

Conditional confounding

- Spurious associations will be minimized if means of important predictors are balanced across treatment arms
  - The greater the value of δ_j the more important it is for the means of the j-th covariate to be equal
    - (Presumes linear model reasonable approximation)
  - We could use estimates of the of δ_j's to define the distance between the arms (or just balance means)
- Balancing group sizes across covariates will tend to have means balanced by randomization
  - Group sizes within strata may matter for subgroup analyses

Adaptive Randomization

Dynamic randomization

- Subjects are assigned to the treatment arm that will achieve best balance
  - When i-th patient accrued, compute a randomization probability, π_i, where

$$\Delta_i = d(\vec{X}, \mathbf{W}|X_i = 1) - d(\vec{X}, \mathbf{W}|X_i = 0)$$

and

$$\pi_i = \Pr[X_i = 1] = f(\Delta_i),$$

with

  - $0 \leq \pi_i \leq 1$
  - $f(\Delta_i)$ monotonically decreasing in π_i
  - (generally seek to avoid $\pi_i = 0$ and $\pi_i = 1$)

Adaptive Randomization

Inference : Population model

- Impact on statistical inference relative to CRD
  - Bias properties unchanged
  - Face validity improved for most important variables
  - Precision improved due to achieving closer to desired randomization ratio
  - Precision could be further improved if adjust for stratification variables in analysis for population model
    - This should be done
    - Requires pre-specification of analysis model to avoid “intent to cheat" analysis

Adaptive Randomization

Advantages and disadvantages

- Advantages:
  - Typically improved face validity
  - Can handle an arbitrary number of covariates
    - Depending on distance metric
- Disadvantages:
  - Logistically more involved
  - Decreased credibility if too deterministic
    - Approaches sequential allocation
  - Some analytic strategies more complex (permutation tests for strong null)
  - Does not necessarily facilitate subgroup analyses
    - Unless distance metric chosen carefully

Adaptive Randomization

Ethics

- Clinical trials are experiments in human volunteers
  - Individual ethics:
    - Patients on trial: Avoid continued administration of inferior treatment
    - Patients not yet on trial: Avoid starting inferior treatment
  - Group ethics:
    - Facilitate rapid adoption of new beneficial treatments
    - Avoid prolonging study of ineffective treatments
- Some authors have described dynamic randomization processes that attempt to minimize exposure of patients to harmful treatments

Adaptive Randomization

Proposed solutions

- Most commonly used
  - Sequential sampling
    - Interim analyses of data
    - Terminate trials when credible decisions can be made
- Also proposed
  - Response adaptive randomization
    - Change randomization probabilities as evidence accumulates that one treatment might be best
    - “Play the winner"

Adaptive Randomization

Play the winner : Urn model

1. Begin with k white balls and k black balls in an urn
2. Upon accrual of a patient draw a ball from urn
   - White $\rightarrow$ control; black $\rightarrow$ treatment
3. Observe outcome
   - If outcome is good, return $m + 1$ balls of same color as withdrawn
   - If outcome is bad, return 1 ball of same color as withdrawn and m balls of opposite color

Adaptive Randomization

Bayesian methods

- An explicit Bayesian approach to dynamic randomization bases the randomization ratio on the current posterior probability that one treatment is superior
  - Ultimately, that posterior probability is based on the number of good outcomes on each treatment (in conjunction with a probability model for the response and a prior distribution)
- Advantage of using Bayesian posterior probability
  - Can easily handle continuous outcomes
  - Can easily handle continuous randomization probabilities

Adaptive Randomization

Analytic issues

- Treatment of successive patients is not independent of previous patients treatment and results
  - Possible bias in accrual of future patients
- Conditionally biased estimates of treatment effect in arm with lower sample sizes
  - Bad early results tend to preclude regression to mean

Response-Adaptive Randomization (Example)

ECMO: Extracorporeal Membrane Oxygenation in neo-natal respiratory failure

- Persistent pulmonary hypertension results in right to left shunt through the foramen ovale or ductus arteriosus causing hypoxemia. ECMO is used to maintain life until the condition resolves.
- Trial 1 (Play the winner absolutely): *Pediatrics* (1985) 76:479-487
  - First subject was randomized to conventional medical therapy (CMT); the infant died.
  - Second subject given ECMO; infant lived.
  - Next 8 subjects given ECMO; all lived.
  - Result:

    100% mortality with CMT
    0% with ECMO
    RR = 0.

Response-Adaptive Randomization (Example)

ECMO Example (con't):

- Trial 2 (Play the winner with higher probability): *Pediatrics* (1989) 84(6):957-63
  - Randomize until the 4th CMT death, then treat remainder with best approach.
  - 19 babies in first phase (4/10 die with CMT; 0/9 die with ECMO).
  - 20 babies on ECMO in second phase (1 death).
  - Result:
    40% (4/10) mortality with CMT;
    3% (1/29) with ECMO;
    RR = 0.086.

- Trial 3 (conventional RCT): *Pediatrics* (1998) 101(4):E1
  - Randomize 185 infants (92 to CMT, 93 to ECMO)
  - Result:
    59% (54/92) mortality with CMT;
    32% (30/93) with ECMO;
    RR = 0.55.

Response-Adaptive Randomization (Example)

ECMO Example (con't):

Implications of the ECMO example:

- ECMO looked better with response-adaptive randomization.
- Response-adaptive designs were not accepted as adequate justification for ECMO.
- Inadequate study designs can delay introduction of beneficial treatments or prolong use of inferior treatments.

"In fact, in the ECMO trial, the patient who failed on treatment B had the most extreme values on no fewer than four important covariates (Paneth & Wallenstein, 1985), and was clearly the sickest. In effect, the trial provides no information whatsoever regarding the treatment comparison. "

-Begg (1990)

Response-Adaptive Randomization (Example)

- The ECMO experience has tempered enthusiasm for randomized PTW
- This being said, there may be times were response-adaptive randomization will work, but
  - There needs to be a clear dilemma re individual ethics
  - There will tend to be decreased group ethics
  - It takes a lot of planning in order to obtain results that will be sufficiently credible

Methods: Logistics of Randomization

- Where to perform randomization:
  - Central randomization:
    - Phone calls to the coordinating center.
    - Sequences can be determined at the start of the study (except with adaptive randomization).
  - Distributed randomization: Computer programs, envelopes, or lists at pharmacies.
- Important principles:
  - Strong quality assurance must be in place to ensure proper randomization.
  - Ensure adequate concealment/blinding.
  - Provide for emergency unblinding.
  - Exact randomization scheme must be known for analysis.

Introduction to Clinical Trials - Day 2

Session 4 - Trial Monitoring for Quality Control

Presented July 24, 2018

Susanne J. May
Department of Biostatistics
University of Washington

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

Study Monitoring and Quality Control

Essential principle of good trial conduct

- Good trial conduct should include:

  1. Masking (blinding)
  2. Treatment allocation (randomization)
  3. Study quality control
     - Data management
     - Data quality monitoring
  4. Trial monitoring
     - Data quality
     - Safety
     - Interim decision and group sequential designs

Essential principle of good trial conduct

Study quality control

- Key elements of study quality control include:

1. Recruitment and retention
2. Ongoing (monitoring) trial quality
   - Quality control of data and study processes
   - Site monitoring
   - Anticipating the unanticipated...
3. Prevention and treatment of missing data

Study quality control

Recruitment, retention, and compliance

- Recruitment and retention:
  - Motivation
    - Most studies are only of scientific interest/relevance for a few years.
    - There is an ethical responsibility to participants to complete a trial once it is started.
    - One of the major reasons for closing studies is lack of accrual.
    - (One of the major reasons for suspending clinical research in an entire institution (closing the IRB) is old studies that are unlikely to be completed.)

"The most important part of good retention is good recruitment." (Richard Hamman, U Colorado)

Study quality control

Recruitment, retention, and compliance

- Recruitment and retention strategies:
  - Study design:
    - Choose intervention groups to encourage participation regardless of intervention group assignment.
    - Minimize trial burden
  - Sources for subjects:
    - Clinical practice
    - Previous trials
    - Patient registries
    - Health fairs (free screening, etc.)
    - Advertisements
  - Inducements:
    - Pens, coffee mugs,...
    - Reimbursement for time and inconvenience.
    - Payments beyond reimbursement are often considered unethical.

Study quality control

Recruitment, retention, and compliance

- Recruitment and retention strategies (Example: SLV HFP)
  - Study design:
    - Even ‘usual care’ group gets screening and education
    - Fasting blood measurements restricted to 12-month (i.e, not at 6 and 18 months)
  - Sources:
    - Medical practice records (groups and individuals)
    - Churches, parks and recreation.
    - Media
    - Health fair (diabetes screening)
    - Previous or ongoing diabetes studies
  - Inducements:
    - Some discussion of pens, coffee mugs,...

Study quality control

Recruitment, retention, and compliance

- Recruitment and retention: monitoring and problem solving
  - Monitoring:
    - Annual IRB reports must summarize accrual
    - Investigators might track accrual of particular types of subjects (especially if sub-group analyses are important).
  - Problem Solving:
    - *Accept a smaller number of subjects
    - More rigorous recruitment
    - Extend the number of centers
    - Extend study time
    - *Relax eligibility or exclusions
    - *Recycle previous subjects

    *Can have serious (adverse) effects on study interpretation or generalizability.

Study quality control

Recruitment, retention, and compliance

- However, the best strategy for recruitment and retention that I have seen is to have:
  - A dedicated study nurse on site
  - Far better recruitment/retention if this person is familiar with the patients (culturally and personally)
  - Far better recruitment/retention if financial reimbursements for the site are (at least partially) paid up front

Study quality control

Recruitment, retention, and compliance

- Compliance
  - Bias is decreased and power is increased when subjects complete the study and are fully compliant.
- It is important to design a study to maximize compliance:
  - Treatments should be defined/chosen to minimize the number of patients deemed non-compliant:
    - Define treatment as a single dose rather than multiple doses.
    - Incorporate ancillary treatments for adverse effects.
    - Modify treatments in presence of adverse effects.
  - Select compliant subjects:
    - Consider perception of potential benefit
    - Education level
    - Co-existing conditions (e.g., chronic conditions, drug abuse)
    - Questionnaires about patient beliefs, family support, etc.
    - Identify compliers with a run-in periods

Study quality control

Recruitment, retention, and compliance

- Methods for promoting compliance
  - Educating subjects:
    - Subjects who are informed of study goals will be better compliers.
    - Communication of potential problems before it is too late.
    - Establish difference between stopping treatment and quitting the study. (True for investigators as well!)
  - Minimize the trial burden:
    - Number and length of clinic visits.
    - Number of forms to be completed.
    - Number of painful procedures.

Recruitment, retention, and compliance

- Disadvantages to promoting compliance:
  - May lengthen trial.
  - Subjects may notice change in therapy (run-in period).
  - Loss of generalizability (efficacy vs. effectiveness).
  - Compliant subjects may have lower event rates and thus potentially lower power (Good thing?).

Study quality control

Demonstration of problems caused by poor compliance

- Compliance (adherence): The extent to which the subjects in a trial follow the treatment that was prescribed for them by the study protocol.
- Problem:
  - Subjects who do not comply with the treatment protocol will decrease statistical power of the study.
  - Non-compliance results in misclassification of some patients in each treatment group:
    - Drop-out: Non-compliant subjects on the new treatment arm.
    - Drop-in: Control subjects who take the new treatment.

Study quality control

Demonstration of problems caused by poor compliance

- Example: Clinical trial of fiber in prevention of colorectal polyps:
  - Endpoint: recurrent polyps within 3 years.
  - Hypotheses:
    - Low fiber: 45% recurrence
    - High fiber: 36% recurrence (20% reduction)
  - Sample size calculation:
    - (One-sided level $\alpha = 0.025$ test with power $\beta = 0.9$)

$$\begin{aligned} N &= \frac{(z_{0.975} + z_{0.90})^2}{(p_0 - p_1)^2}(p_0 q_0 + p_1 q_1) \\ &= \frac{(1.96 + 1.28)^2}{(0.45 - 0.36)^2}(0.45 \times 0.55 + 0.36 \times 0.64) \\ &= 620/arm \end{aligned}$$

Demonstration of problems caused by poor compliance

- Example (con't): Effect of drop-out
  - Suppose there is 75% compliance on the high fiber arm.
  - Attenuated treatment effect:
    - 75% have 36% recurrence
    - 25% have 45% recurrence
    - Overall $\approx$ 38% recurrence
  - Revised sample size:

$$
\begin{aligned}
N &= \frac{(z_{0.975} + z_{0.90})^2}{(p_0 - p_1)^2}(p_0 q_0 + p_1 q_1) \\
&= \frac{(1.96 + 1.28)^2}{(0.45 - 0.38)^2}(0.45 \times 0.55 + 0.38 \times 0.62) \\
&= 1035/arm
\end{aligned}
$$

Demonstration of problems caused by poor compliance

- Example (con't): Effect of drop-in
  - Suppose 10% of controls increase their fiber.
  - Attenuated treatment effect:
    - 10% have 36% recurrence
    - 90% have 45% recurrence
    - Overall $\approx$ 44% recurrence
  - Revised sample size:

$$\begin{aligned} N &= \frac{(z_{0.975} + z_{0.90})^2}{(p_0 - p_1)^2}(p_0 q_0 + p_1 q_1) \\ &= \frac{(1.96 + 1.28)^2}{(0.44 - 0.38)^2}(0.44 \times 0.56 + 0.38 \times 0.62) \\ &= 1406/arm \end{aligned}$$

Study quality control

Demonstration of problems caused by poor compliance

- Very naive solution: Treat non-compliant patients on the treatment arm as if they were on control.
  - Problem: Many studies have shown that non-compliant patients have lower survival than compliant patients (even on placebo).
  - Clearly this approach will tend to make any treatment look good.

Study quality control

Demonstration of problems caused by poor compliance

- Naive solution: Restrict analysis to compliant patients (“as treated analysis").
  - If non-comlpliant patients can be indentified and safely discarded from the analysis, then we would only need to inflate the sample sizes for each arm according to the rate of non-compliance.
- Example:
  - High fiber arm (25% drop-out)
    Accure 620/0.75 = 827
  - High fiber arm (10% drop-in)
    Accure 620/0.10 = 689
  - Compare the total of 1516 as opposed to $2 \times 1406 = 2812$ if the misclassified subjects are used.

Study quality control

Demonstration of problems caused by poor compliance

- Problems with naive solution:
  - Treatment may affect compliance:
    - Compliance is then an outcome of the treatment.
    - Can make bad treatments look good.
  - Non-compliers are different from compliers.
    - We can never know if the outcome in non-compliers would have been different if they had been compliant.
    - To leave them out of an analysis can create selection bias.

Failure of the As Treated Analysis

1. Drop-out is due to symptoms related to worsening of the disease; the treatment 'cures' the symptoms, but not the disease:
   - Control group will have more drop-outs, and those drop-outs will be the ones with bad disease.
   - As treated analysis will make the treatment look bad because the worst control patients are ignored.

Failure of the As Treated Analysis

2. Drop-out due to perception of getting the worse treatment:
   - Patients have a bias toward the new treatment.
   - Worsening condition on placebo leads to non-compliance.
   - Worsening condition on new treatment has no effect on compliance.
   - As-treated analysis makes new treatment look bad.
   - (Example: early AIDS trials.)

Study quality control

Failure of the As Treated Analysis

3. Drop-out due to adverse events, but concordance between adverse events and treatment outcome differs between treatment arms:
   - Adverse events might indicate better prognosis on the treatment arm and worse prognosis on the control arm
   - Example: Chemotherapy in cancer
     - Nausea and vomiting can be caused both by progressive disease and by the treatment.
     - Treatment arm: greater side effects tend to go with higher anti-tumor effects.
     - Control arm: greater side effects tend to go with disease progression.
     - As treated analysis can make treatment look bad.

Failure of the As Treated Analysis

4. Drop-out due to treatment harm:
   - Example: Chemotherapy in cancer
     - New chemotherapy cannot be tolerated by the patients with poor prognosis (or even worse, treatment causes adverse outcomes that lead to non-compliance).
     - Control arm has no tolerance problems and good compliance.
     - As treated analysis makes the treatment look good by ignoring its failures.

Study quality control

Demonstration of problems caused by poor compliance

- Solution:
  - Primary efficacy analysis should generally be based on intention-to-treat
    - Analyze patients according to the treatment they were randomized to
  - (discussed as part of Statistical Analysis Plan)
- See also: National Academies Panel on Prevention and Treatment of Missing Data (discussed below)

Study quality control

Monitoring study quality

- Although the trail must be designed to assure quality, that quality must be monitored as part of trial conduct.
  - Data QC
  - Monitoring accrual, compliance, and retention as discussed above
  - Problems must be discovered and corrected ASAP
  - Example of what I monitor for data quality
    - Data consistency monitoring (software checks)
    - Regular reports on missing data, protocol deviations, etc.
    - Reports on eligibility and exclusion criteria (and exceptions)
    - Randomization integrity (randomized subjects must receive treatment)
    - Adherence to visit schedules

Monitoring study quality

- Site monitoring:
  - Most multi-center trials send site monitors to all sites to confirm:
    - Treatments and procedures are following protocol.
    - Data in trial database matches information in patient charts.
    - Discrepancies are reported to sponsor and site PI must correct.

Prevention and treatment of missing data

How can there be missing data?

- Consider 3 mechanisms by which missing data in trials arise:
  - Non-compliance:
    - Subject stops the assigned treatment
    - Outcome measurements are obtained
    - Missing the outcome measure that would have been obtained if the subject had remained on treatment.
    - Solution: Intention-to-treat analysis
  - Withdrawal of consent:
    - Subject withdraws from the study (it is their right).
    - Outcome measurement cannot be obtained
    - Subjects should be offered the opportunity to remain on the study but stop all interventions and still return for outcome measurements (i.e., non-compliant).
  - Loss-to-followup:
    - Subjects have left the study and cannot be contacted.
    - Avoidable through good study management.
    - We should not accept loss-to-followup.

Prevention and treatment of missing data

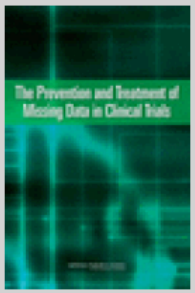
Impact of missing data

- Missing data decrease trial quality:
  - Cannot rule out bias due to differences between those who are observed and those who are not.
  - Avoid missing data through careful definition of endpoints.
    - Identify the most important endpoints and make sure they are measured.
    - Use outcomes that are easy to obtain (mortality vs tumor progression).
    - Define the endpoint so that data which are impossible to observe are assigned a meaningful value:
      E.g., Quality of life after death = 0.
  - Statistical adjustments are always based on untestable assumptions:
    - MNAR: missing not at random. Missing data mechanism differs from the relationships that are observed in the non-missing data.

Prevention and treatment of missing data

How big of a problem is missing data in clinical trials?

- The National Academies recently convened an expert panel of statisticians to discuss the prevention and treatment of missing data, including
  - Standardizing terminology
  - Enforcing the idea that the best way to deal with missing data is to not have missing data
  - Provide recommendations to avoid missing data
  - Provide recommendations for addressing missing data in trial analyses



The Prevention and Treatment of Missing Data in Clinical Trials

ISBN
978-0-309-15814-5

162 pages
6 x 9
PAPERBACK (2010)

Panel on Handling Missing Data in Clinical Trials; National Research Council

Prevention and treatment of missing data

- Contents of NRC report:

1. Introduction and background
2. Trial designs to reduce the frequency of missing data
3. Trial strategies to reduce the frequency of missing data
4. Drawing inference from incomplete data
5. Principles and methods of sensitivity analyses
6. Conclusions and recommendations:
   - *Trial Objectives*:
     Recommendation 1
   - *Reducing dropouts through trial design:*
     Recommendations 2, 3, 4, 5.
   - *Reducing dropouts through trial conduct:*
     Recommendations 6, 7, 8.
   - *Treating missing data:*
     Recommendations 9, 10, 11, 12, 13, 14, 15.
   - *Understanding the causes and degree of dropouts in clinical trials:*
     Recommendations 16, 17, 18.

Recommendations of the NRC report

Recommendation 1:

- The trial protocol should explicitly define the objective(s) of the trial; the associated primary outcome or outcomes; how, when, and on whom the outcome or outcomes will be measured; and the measures of intervention effects, that is, the causal estimands of primary interest.

- These measures should be meaningful for all study participants, and estimable with minimal assumptions. Concerning the latter, the protocol should address the potential impact and treatment of missing data.

Recommendations of the NRC report

Recommendation 2:

- Investigators, sponsors, and regulators should design clinical trials consistent with the goal of maximizing the number of participants who are maintained on the protocol-specified intervention until the outcome data are collected.
- (see previous discussion)

Recommendations of the NRC report

Recommendation 3:

- Trial sponsors should continue to collect information on key outcomes on participants who discontinue their protocol-specified intervention in the course of the study, except in those cases for which a compelling cost-benefit analysis argues otherwise, and this information should be recorded and used in the analysis.
- Treatment discontinuation does not equate to study discontinuation!

Recommendations of the NRC report

Recommendation 4:

- The trial design team should consider whether participants who discontinue the protocol intervention should have access to and be encouraged to use specific alternative treatments.
- Such treatments should be specified in the study protocol.

Recommendations of the NRC report

Recommendation 5:

- Data collection and information about all relevant treatments and key covariates should be recorded for all initial study participants, whether or not participants received the intervention specified in the protocol.

Recommendations of the NRC report

Recommendation 6:

- Study sponsors should explicitly anticipate potential problems of missing data. In particular, the trial protocol should contain a section that addresses missing data issues, including the anticipated amount of missing data, and steps taken in trial design and trial conduct to monitor and limit the impact of missing data.

Recommendations of the NRC report

Recommendation 7:

- Informed consent documents should emphasize the importance of collecting outcome data from individuals who choose to discontinue treatment during the study, and they should encourage participants to provide this information whether or not they complete the anticipated course of study treatment.

Recommendations of the NRC report

Recommendation 8:

- All trial protocols should recognize the importance of minimizing the amount of missing data, and, in particular, they should set a minimum rate of completeness for the primary outcome(s), based on what has been achievable in similar past trials.

Recommendations of the NRC report

Recommendation 9:

- Statistical methods for handling missing data should be specified by clinical trial sponsors in study protocols, and their associated assumptions stated in a way that can be understood by clinicians.

Recommendations of the NRC report

Recommendation 10:

- Single imputation methods like last observation carried forward and baseline observation carried forward should not be used as the primary approach to the treatment of missing data unless the assumptions that underlie them are scientifically justified.

Recommendations of the NRC report

Recommendation 11:

- Parametric models in general, and random effects models in particular, should be used with caution, with all their assumptions clearly spelled out and justified. Models relying on parametric assumptions should be accompanied by goodness-of-fit procedures.

Prevention and treatment of missing data

Recommendations of the NRC report

Recommendation 12:

- It is important that the primary analysis of the data from a clinical trial should account for the uncertainty attributable to missing data, so that under the stated missing data assumptions the associated significance tests have valid type I error rates and the confidence intervals have the nominal coverage properties.
- For inverse probability weighting and maximum likelihood methods, this analysis can be accomplished by appropriate computation of standard errors, using either asymptotic results or the bootstrap.
- For imputation, it is necessary to use appropriate rules for multiply imputing missing responses and combining results across imputed datasets because single imputation does not account for all sources of variability.

Recommendations of the NRC report

Recommendation 13:

- Weighted generalized estimating equations methods should be more widely used in settings when missing at random can be well justified and a stable weight model can be determined, as a possibly useful alternative to parametric modeling.

Recommendations of the NRC report

Recommendation 14:

- When substantial missing data are anticipated, auxiliary information should be collected that is believed to be associated with reasons for missing values and with the outcomes of interest.
- This could improve the primary analysis through use of a more appropriate missing at random model or help to carry out sensitivity analyses to assess the impact of missing data on estimates of treatment differences.
- In addition, investigators should seriously consider following up all or a random sample of trial dropouts, who have not withdrawn consent, to ask them to indicate why they dropped out of the study, and, if they are willing, to collect outcome measurements from them.

Recommendations of the NRC report

Recommendation 15:

- Sensitivity analyses should be part of the primary reporting of findings from clinical trials.
- Examining sensitivity to the assumptions about the missing data mechanism should be a mandatory component of reporting.

Prevention and treatment of missing data

Recommendations of the NRC report

- The NRC Panel recommendations have made an impact on funding agencies, regulatory agencies, and journals
- Since they have emerged, FDA has consistently required multiple sensitivity analyses be pre-specified in the Statistical Analysis Plan

Prevention and treatment of missing data

Recommendations of the NRC report

- Commonly requested sensitivity analyses include some combination of:

1. Multiple imputation
2. Inverse probability weighted estimator
3. “Worst case" scenario
   - Assume best observed outcome in control and worst observed outcome in treatment
4. Pattern mixture models
   - Semi-parametric (shift) model on differences in missing values between treatment and control subjects
   - Generally range from worst case scenario to no difference
5. “Tipping point" analysis
   - How bad do imputed differences between treatment and control have to be in order to change results?

Prevention and treatment of missing data

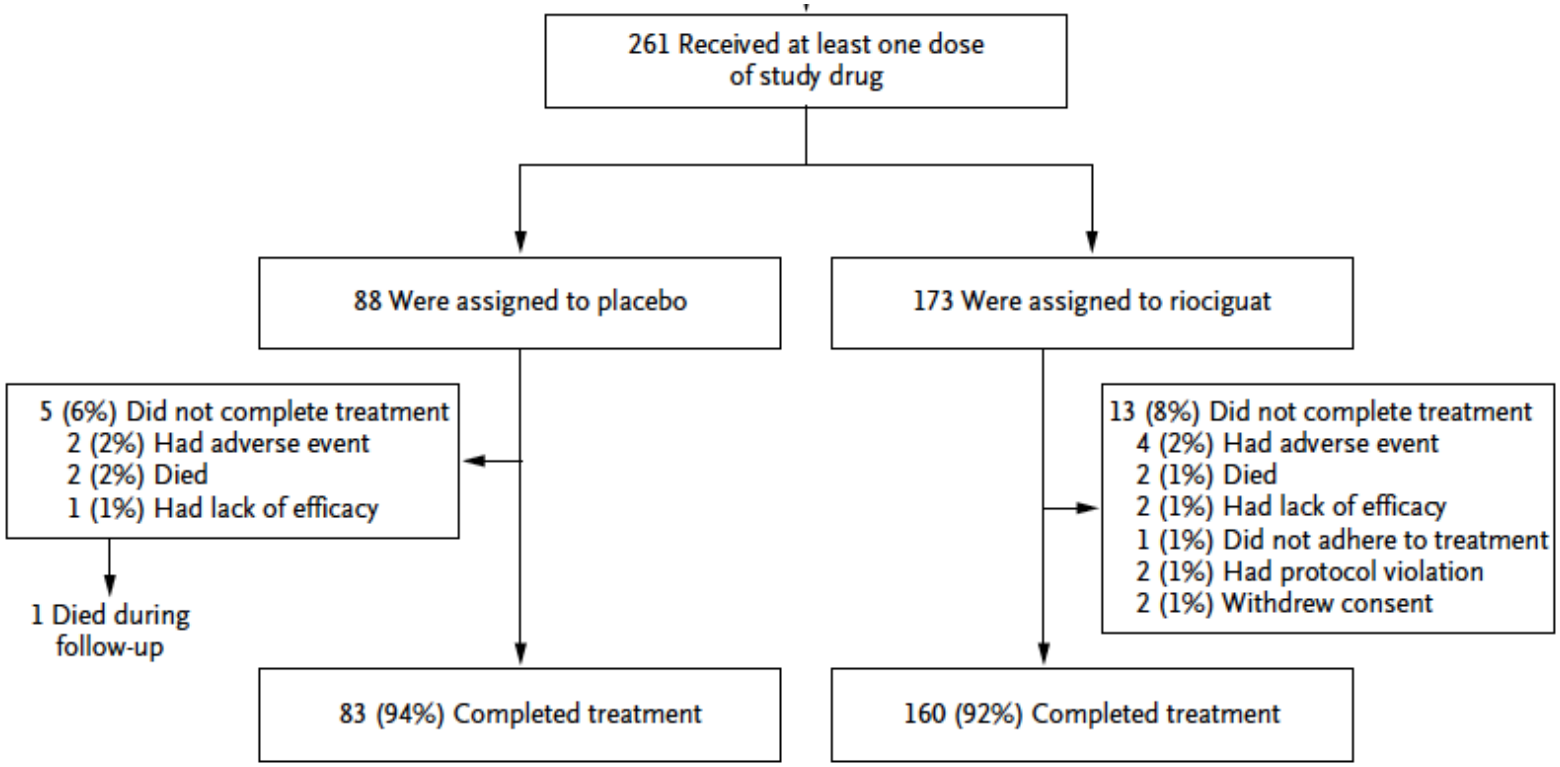
Ex: CHEST trial

- Example: CHEST trial: Ghofrani,*et.al.* NEJM (2013); 369: 319-29: Riociguat for the Treatment of Chronic Thromboembolic Pulmonary Hypertension.
  - Trial: Randomized double-blind placebo controlled trial in patients with inoperable CTEPH.
  - Primary endpoint: 16-week change in 6-minute walk distance (6MWD)
  - Summary of outcome: mean change denoted by θ_1 (riociguat) and θ_0 (placebo)
  - Measure of treatment effect: $\theta = \theta_1 - \theta_0$.
  - Results: "...By week 16, the 6-minute walk distance (had increased by a mean of 39 m in the riociguat group, as compared with a mean decrease of 6 m in the placebo group (least-squares mean difference, 46 m; 95% confidence interval [CI], 25 to 67; P<0.001)."

Prevention and treatment of missing data

Ex: CHEST trial

- Missing data in CHEST trial:



Ex: CHEST trial

- Analysis based on modified intention-to-treat population, defined as all patients who underwent randomization and received at least one dose of the study medication
- Pre-specified imputation for missing data:
  - Patients who died or withdrew due to clinical worsening without terminal visit:
    - 6MWD at 16 weeks set to worst possible value: 0 meters
  - Patients who stopped study medication prematurely:
    - 6MWD at 16 weeks set to value at terminal visit or last visit post baseline.

Prevention and treatment of missing data

Ex: CHEST trial

- Pre-specified sensitivity analyses for missing data:

Table S1. Change in 6-Minute Walk Distance from Baseline: Sensitivity Analyses (Modified Intention-To-Treat population).

Analysis	Estimated Treatment Difference* (m)	95% Confidence Interval
Multivariate linear model at week 16	44.40	27.94 to 60.85
Multiple imputation: fixed penalty: riociguat –60 m and placebo –60 m	43.69	26.25 to 61.13
Multiple imputation: decreasing slope: riociguat –20 m and placebo –20 m per visit	41.81	24.05 to 59.58
Multiple imputation: fixed penalty: riociguat –60 m and placebo –0 m	40.07	22.94 to 57.21
Multiple imputation: decreasing slope: riociguat –20 m and placebo –0 m per visit	38.71	21.27 to 56.15

* Riociguat – placebo

Ex: CHEST trial

- Conclusion (from the paper):

  "At week 16, the 6-minute walk distance had increased from baseline by a mean of 39 m in the riociguat group, as compared with a mean decrease of 6 m in the placebo group (least-squares mean difference, 46 m; 95% confidence interval [CI], 25 to 67; P<0.001), on the basis of an analysis of the modified intention-to-treat population with missing values imputed (Table 2 and Fig. 2). In sensitivity analyses for missing data that used statistical methods for longitudinal data (see the Supplementary Appendix), the benefit of riociguat was similar to that observed in the main analysis (Table S1 in the Supplementary Appendix)."

Introduction to Clinical Trials - Day 2

Session 5 - Independent Data Monitoring Committees

Presented July 24, 2018

Susanne J. May
Department of Biostatistics
University of Washington

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

Purpose of an IDMC

Mechanisms for ensuring ethical treatment of study subjects

- Before starting the study:
  - Institutional review board (IRB)
- During conduct of the study:
  - Data safety monitoring board (DSMB)
- After studies completed:
  - Regulatory agencies (e.g., FDA)

Motivating Example

Trial 002 of the CPCRA

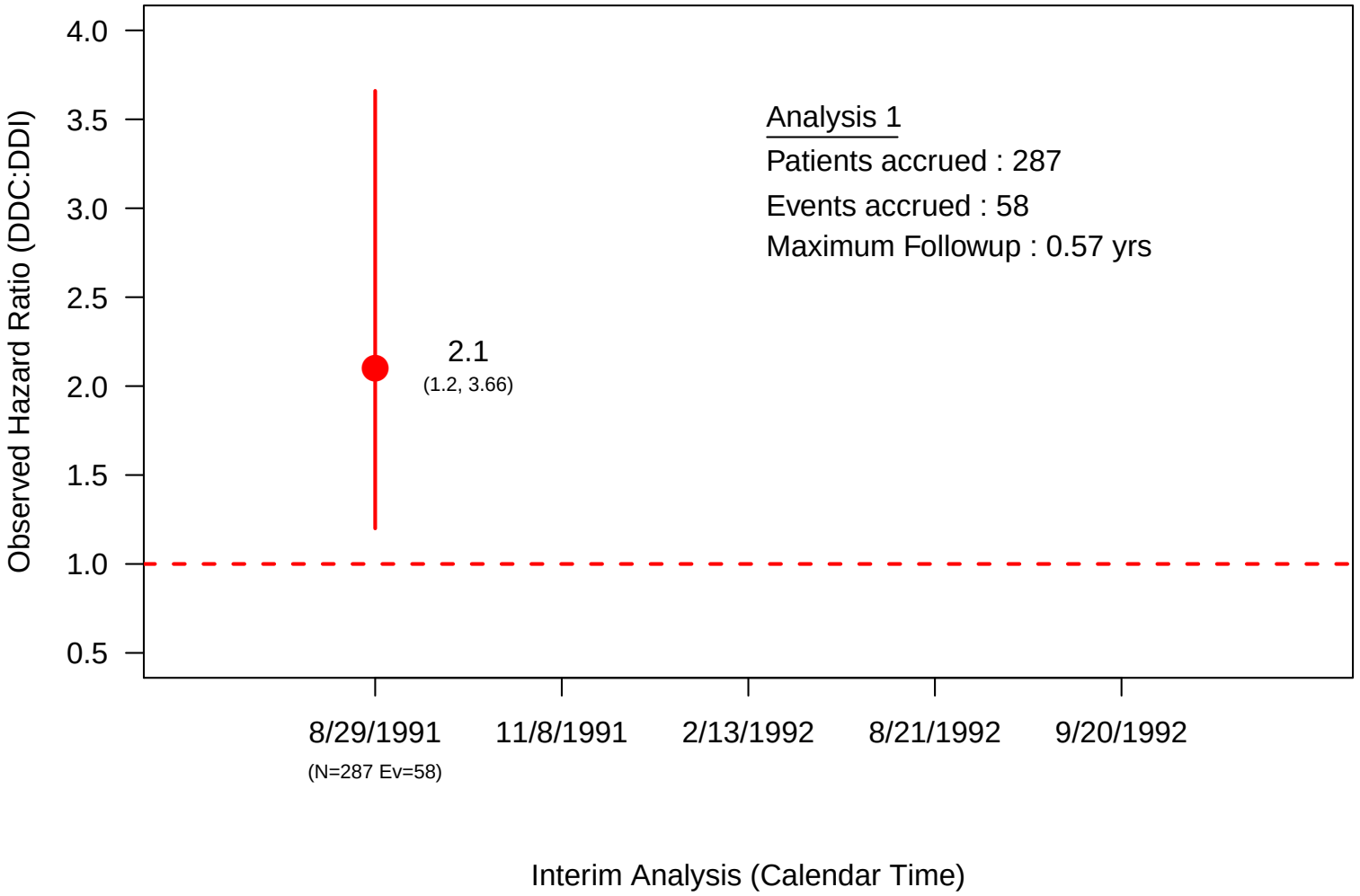
- Community Programs for Clinical Research in AIDS (CPCRA)
- Designed to compare the efficacy of two antiretroviral agents
  - Zalcitabine (DDC) - New experimental treatment
  - Didanosine (DDI) - Active control
- Patient population: Non-responders to zidovudine (AZT)
- Non-inferiority trial
  - DDI considered standard of care at the time

Motivating Example

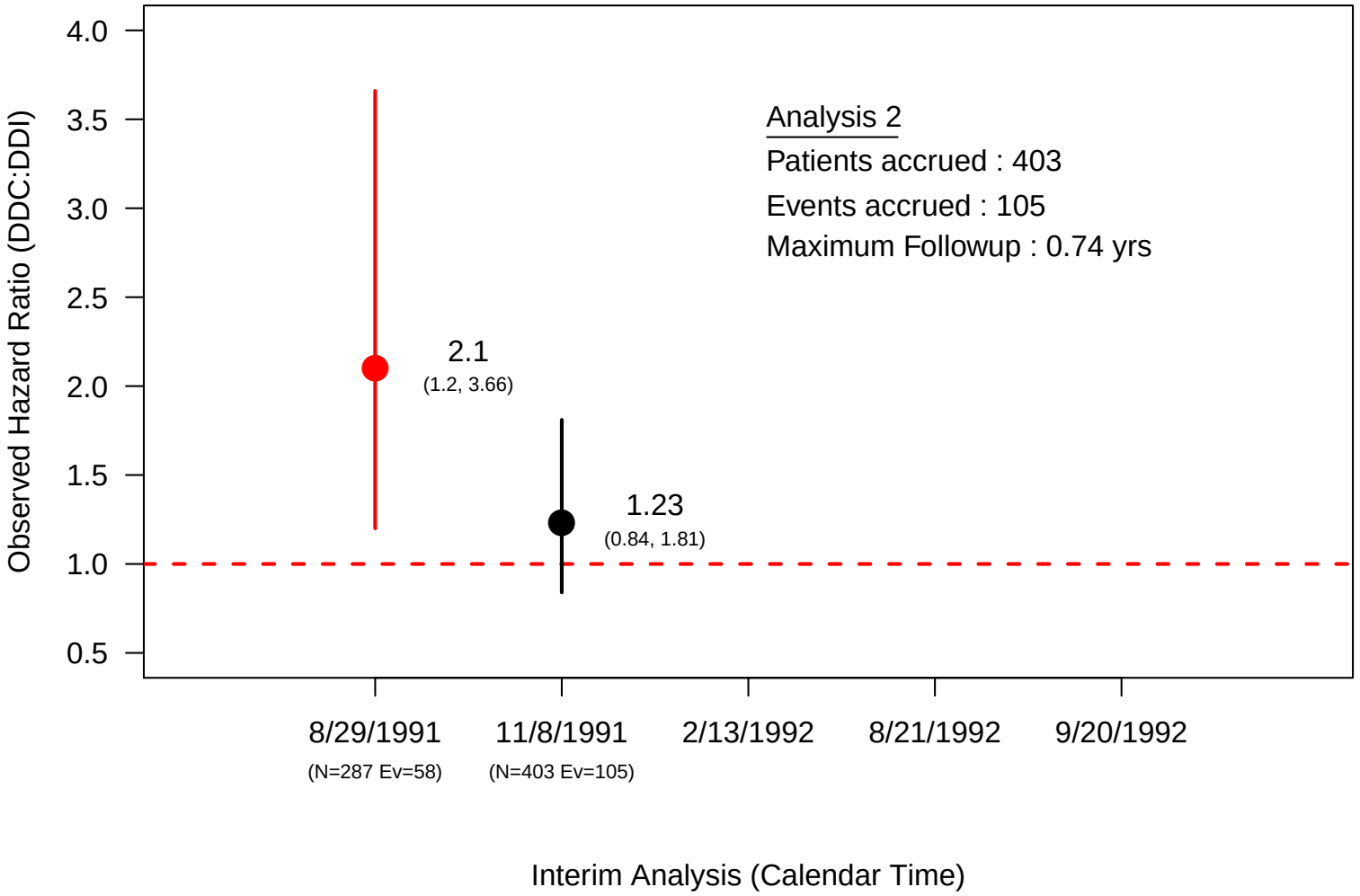
CPCRA Study Protocol

- Primary endpoint: Time to first of disease progression or death
- Sample size: 467 patients randomized
  - Powered for 243 events
  - Maximal duration expected to be 2 years
- Study initiated in December 1990
  - IDMC formed for monitoring approximately every 6 months

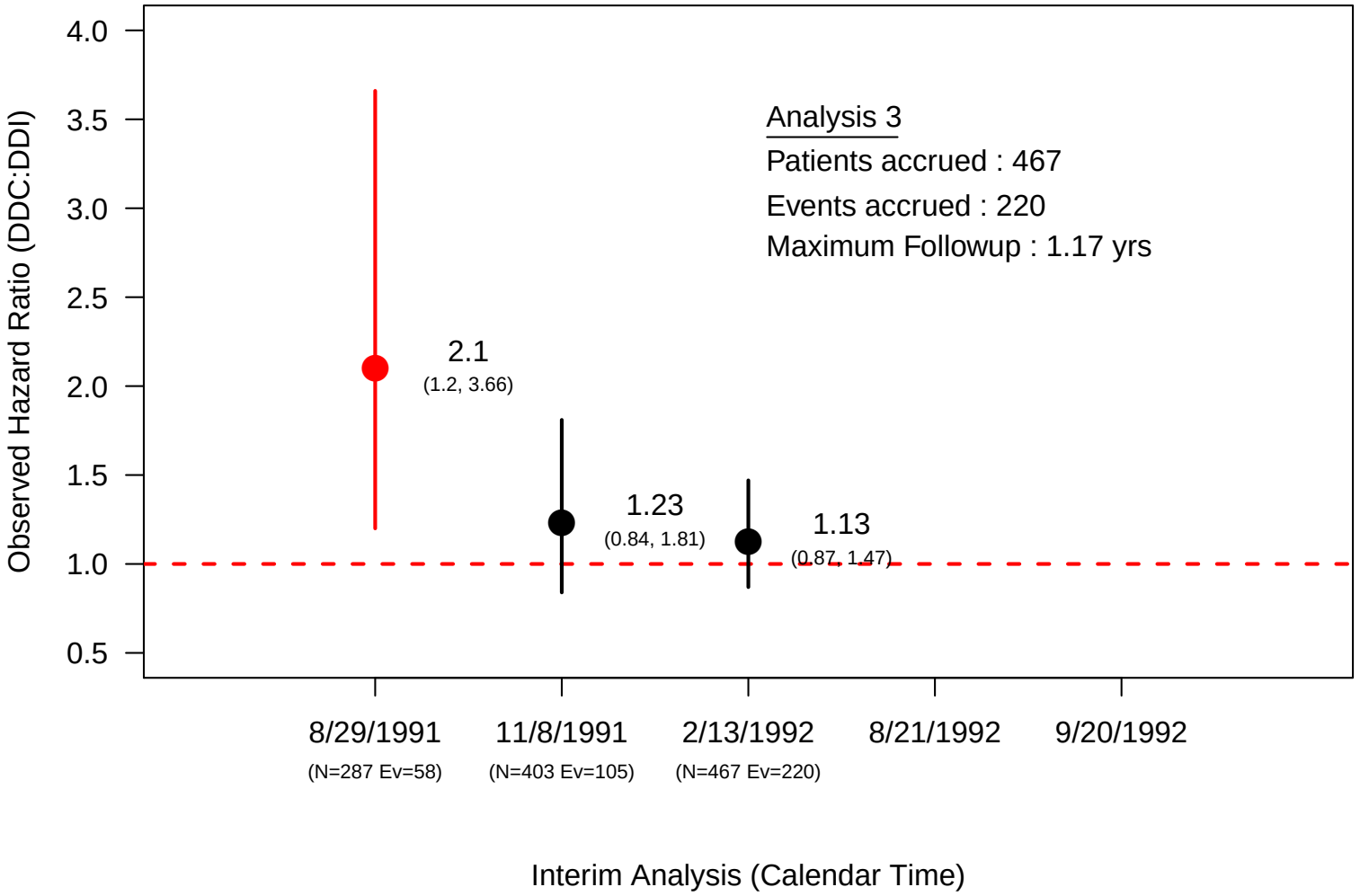
CPCRA Trial Results



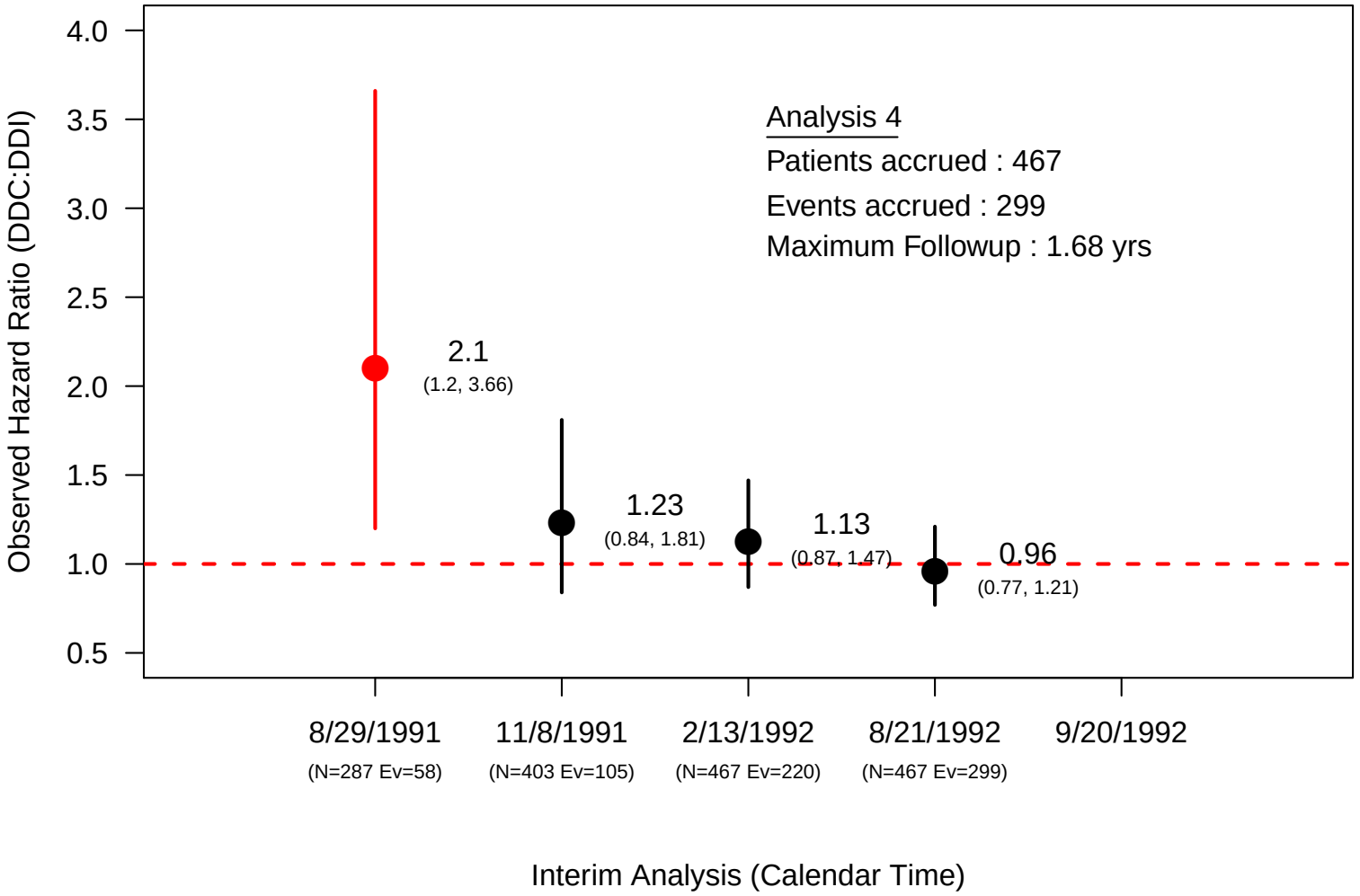
CPCRA Trial Results



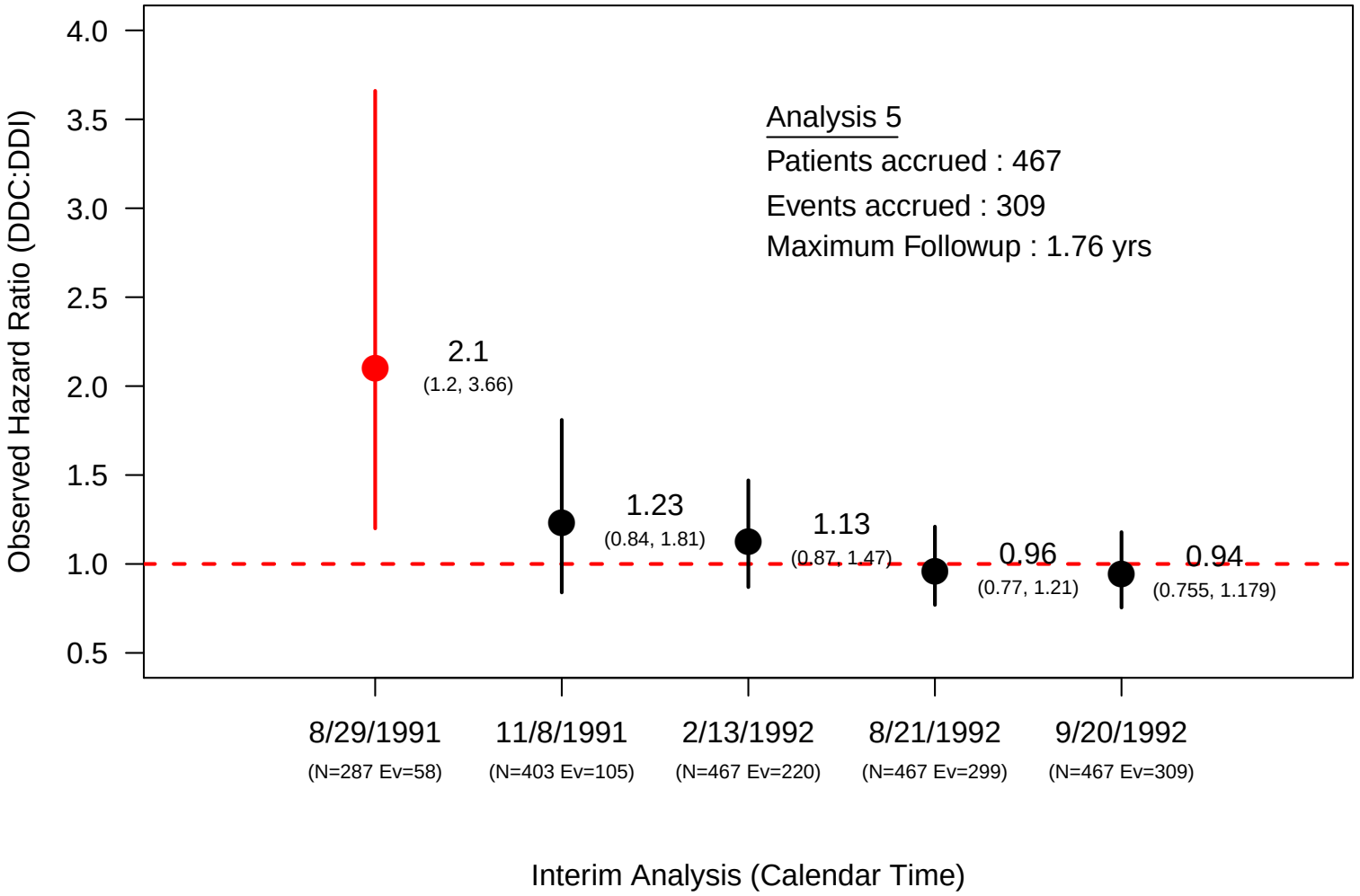
CPCRA Trial Results



CPCRA Trial Results



CPCRA Trial Results



Motivating Example

Comments on the CPCRA Study

- IDMC considered confidence intervals when making continuation decisions
- IDMC was experienced to understand the need for early conservatism under highly variable estimates
- IDMC was able to weigh risk vs benefit

Purpose of an IDMC

Reason for Study Monitoring

- To protect the interests of the study participants
- To preserve trial integrity and credibility in a manner that will enable the clinical trial
- To provide timely and reliable insights to the broader scientific community

Purpose of an IDMC

Requirements

- Achieving the objectives of trial monitoring requires one to confront multiple complex issues beyond the simple implementation of group sequential stopping boundaries (even well-defined boundaries!)
- Ultimately, monitoring requires solid judgement that must be
  - Well informed (clinically, ethically, scientifically, and statistically)
  - Independent and scientifically objective
- This motivates the fundamental principles for DMC membership and function

IDMC Membership

Fundamental principles

- Multidisciplinary representation
- Freedom from apparent significant conflicts of interest
  - Financial
  - Professional
  - Regulatory
- Sole access to interim results on safety of interventions *and* relative efficacy

IDMC Membership

Fundamental principles

- Multidisciplinary representation
- Freedom from apparent significant conflicts of interest
  - Financial
  - Professional
  - Regulatory
- Sole access to interim results on safety of interventions *and* relative efficacy

IDMC Membership

Example: Topical hemostatic agent

- Five members
  - 1 Statistician
  - 1 Hematologist
  - 2 Surgeons (1 soft tissue, 1 bone)
  - 1 Immunologist
- Facilitation of IDMC by independent statistician (not a member of the IDMC)
- Membership excludes
  - Industry
  - Regulatory agencies
  - Study investigators

IDMC Membership

Example: First-line Treatment of T-Cell Lymphoma

- Four members
  - 1 Statistician
  - 3 Clinical oncologists (USA, France, England)
- Three non-voting members
  - 1 Statistician
  - 2 Clinical oncologists (USA, England)
- Facilitation of IDMC by independent statistician (not a member of the IDMC)
- Membership excludes
  - Industry
  - Regulatory agencies
  - Study investigators

IDMC Communication

Formal meetings

- When monitoring a single study it is typical for and IDMC to have at least two meetings a year
  - One teleconference
  - Highly recommended to have at least one face-to-face
- When monitoring multiple trials, more frequent meetings are likely necessary
  - DSMB for CFCCC at UCI meets monthly

IDMC Communication

Formal meetings

- General structure of a meeting generally follows a open, closed, and optional open session format
- Participants in each:
  - Open : IDMC, (Sponsor, Program Investigators, Regulatory), Independent statistician
  - Closed : IDMC, Independent statistician
  - Open : IDMC, (Sponsor, Program Investigators, Regulatory), Independent statistician

IDMC Communication

Open statistical report : Typical outline

1. Executive summary of the study design with schema
2. Overview of salient points of the trial protocol
3. Statistical commentary explaining issues presented in the Open Report figures and tables
4. DMC monitoring plan and summary of past Open Report data presented at prior meetings, along with prior open session minutes
5. Major protocol changes
6. Information on patient screening

*Note: All Open Report data presented in the Open Report should be pooled by treatment arm

IDMC Communication

Open statistical report : Typical outline (cont'd)

7. Study accrual by month and by site (actual and anticipated)
8. Eligibility violations
9. Baseline characteristics
   - Demographics
   - Laboratory values and other measurements
   - Concomitant medications
10. Measure of how up-to-date data are (use benchmark visits)
11. Days between randomization and initiation of treatment

*Note: All Open Report data presented in the Open Report should be pooled by treatment arm

Open statistical report : Typical outline (cont'd)

12. Length of followup data available ("censoring distribution")
13. Participant treatment and study status along with CONSORT diagram
14. Attendance at scheduled visits
15. Compliance with treatment

*Note: All Open Report data presented in the Open Report should be pooled by treatment arm

IDMC Communication

Closed statistical report : Typical outline

1. Detailed statistical commentary explaining issues raised by Closed Report tables, listing, and figures
2. DMC monitoring plan and summary of Closed Report data presented at prior meetings
3. All of items in the Open Report separated by treatment arm
4. Kaplan-Meier estimates of time to treatment and study discontinuation
5. Analyses of primary and secondary efficacy endpoints
   - Important for weighing risk/benefit

Closed statistical report : Typical outline (cont'd)

6. Analyses of adverse events and overall safety data
   - Broken down by system organ class and preferred term
   - All grades
   - Serious adverse events only
   - Stratified by grade
   - “Treatment emergent" adverse events
   - Adverse events leading to treatment modification or discontinuation
7. Listings of adverse events

- Finally, it is a common task of the IDMC to periodically request new analyses as concerns or questions arise during the progression of a trial

IDMC Issues

Issues : Blinding

- Not controversial : An IDMC should always be free to unblind themselves at any time
- However, there are differing opinions on whether the IDMC should start out unblinded

IDMC Issues

Issues : Blinding

- Pros of blinding the IDMC:
  - Avoids leaks in trial results (data falling into wrong hands)
  - Avoids inadvertent leaks of study results by DMC members
  - Avoids overreaction to early variable results
- Cons of blinding the IDMC:
  - Need timely and informed integration of patterns for weighing risk/benefit
  - Can provide earlier detaching of something “real" using evidence that has been observed

IDMC Issues

Issues : Blinding

- Ex: The CAST Trial
  - DMC blinded through X/Y coding for encainide and flecainide vs. placebo
  - First DMC meeting : 13 vs 7 deaths
    - DMC recommended continuation
  - Emergency DMC meeting : 56 vs. 22 deaths
    - DMC recommended immediate termination
- Had the DMC been unblinded, would they have acted sooner?

Issues : Blinding

- In my opinion, if the DMC does choose to be blinded then:
  - They should be able to unblind at any time it is felt necessary
  - If one member becomes unblinded, then all members should be unblinded
  - It is essential for all DMC members to play the hypothetical
    - When looking at a potential imbalance in safety events, must ask whether knowing the actual treatment codes would lead to a different recommendation
- Even if the DMC is unblinded, the Closed Report should have dummy labels with actual treatment codes available through a separate form of communication
  - Avoid unintentional leaking of trial results

IDMC Issues

Issues : Conflict of interest and sponsor/DMC relationship

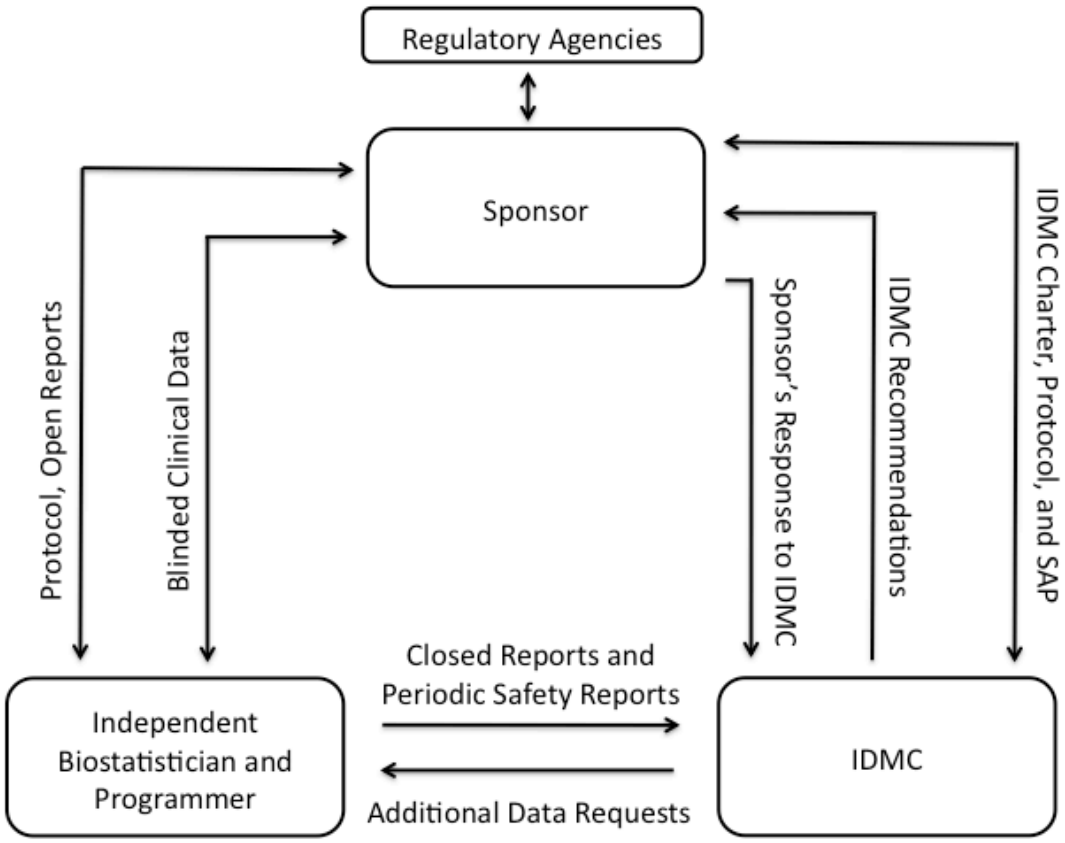
- Different strategies are taken in industry sponsored trials

1. No interim analyses
2. Strictly in-house monitoring
3. Independent DMC with in-house analyses
   - Loosely controlled in-house blinding, or
   - Only study statistician(s) unblinded
4. Independent DMC and independent statistician, with data collection in-house
5. Completely hands-off

IDMC Issues

Issues : Conflict of interest and sponsor/DMC relationship

- (4) and (5) are good approaches
  - Helps to keep sponsor above suspicion of “intention-to-cheat"



IDMC Issues

Issues : Conflict of interest and sponsor/DMC relationship

- Certainly the DMC members should be free of potential conflicts of interest:
  - Financial, scientific, or regulatory in nature
  - Shouldn't own (significant?) stock in company
  - No conflicts with competing products
- Conflicts should be updated as they arise

IDMC Issues

Issues : Indemnification of the IDMC

- DMCs or members can subpoenaed and become defendants in litigation
- DMCs must be indemnified by the sponsor or through some other defined process
- Indemnification language should be part of the DMC Charter as well as contracts
- Indemnification should be provided in order to keep DMC member free to use best judgement when issuing trial recommendations without fear of litigation

Introduction to Clinical Trials - Day 2

Session 6 - Group Sequential Monitoring

Presented July 24, 2018

Susanne J. May
Department of Biostatistics
University of Washington

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

Trial monitoring

Elements and motivation for trial monitoring

- Motivation: Many trials have been stopped early:
  - Physician health study showed that aspirin reduces the risk of cardiovascular death.
  - A phase III study of tamoxifen for prevention of breast cancer among women at risk for breast cancer showed a reduction in breast cancer incidence.
  - A phase III study of anti-arrhythmia drugs for prevention of death in people with cardiac arrhythmia stopped due to excess deaths with the anti-arrhythmia drugs.
  - A phase III study of folic acid supplements for prevention of neural tube defects.
  - Women's Health Initiative: Hormones cause heart disease.

Trial monitoring

Elements and motivation for trial monitoring

- What is trial monitoring?
  - Monitoring for quality control; for example,
    - Patient accrual.
    - Data quality/completeness.
    - Unanticipated adverse events.
  - Monitoring study endpoints(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.

Trial monitoring

Elements and motivation for trial monitoring

- Reasons to monitor study 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).

Trial monitoring

Elements and motivation for trial monitoring

- If not done properly, monitoring of endpoints can lead to biased results:
  - Data driven analyses cause bias:
    - Analyzing study results because they look good leads to an overestimate of treatment benefits.
  - Publication or 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. Decisions made 'in the heat of the moment' are 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 and motivation for trial monitoring

- 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 and motivation for trial monitoring

Key elements of monitoring

- How are trials monitored?
  - Investigator knowledge of interim results can lead to biased results:
    - Negative results may lead to loss of enthusiasm.
    - Positive interim results may lead to inappropriate early publication.
    - Either result may cause changes in the types of subjects who are recruited into the trial.
  - “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:

      1. Assure the trial is safe to continue.
      2. Make decisions about early termination based on the statistical monitoring plan (“group-sequential clinical trial design").

Elements and motivation for trial monitoring

Key elements of monitoring

The trial monitoring plan is typically pre-specified in two 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

Elements and motivation for trial monitoring

Key elements of monitoring

- Typical content for DSMB charter:
  - Trial synopsis; for example:
    - Summary of design
    - Eligibility/exclusions
    - Statistical design and sample size
  - DSMB organization
    - Composition and selection of members
  - Responsibilities of DSMB
    - What will be monitored (accrual, QC, safety, endpoints?)
  - Responsibilities of sponsor
    - Providing open/closed reports; data summaries
  - Committee meetings:
    - Open session; closed session; executive session
  - Communication
    - Open report; closed report to be provided to DSMB
    - Responsibility for meeting minutes (open and closed minutes)
    - Process for DSMB recommendations

Elements and motivation for trial monitoring

Key elements of monitoring

- Typical content for ISAP:
  - Safety monitoring plan (if there are formal safety interim analyses)
    - Decision rules for formal safety analyses
    - Evaluation of decision rules (power, expected sample size, stopping probability)
    - Methods for modifying rules (changes in timing of analyses)
    - Methods for inference (bias adjusted inference)
  - Monitoring plan for primary endpoint(s)
    - Decision rules and reasons for early termination (e.g., efficacy, futility, equivalence, harm)
    - Evaluation of decision rules (power, expected sample size, stopping probability)
    - Methods for modifying rules (changes in timing of analyses)
    - Methods for inference (bias adjusted inference)
  - Data handling and responsibilities for analysis

Statistical framework for trial monitoring: Statistical design of the fixed-sample trial

- The interim statistical analysis plan is based on the fixed sample design
  - Primary endpoint
  - Probability model
  - Functional
  - Contrast
  - Statistical hypotheses
  - Statistical standards for decisions (interval estimate)

Statistical framework for trial monitoring: Statistical design of the fixed-sample trial

- The statistical decision criteria are referenced to the trial's design hypotheses. For example:
  - One-sided superiority test (assume small θ favors new treatment):

$$\begin{aligned} \text{Null:} \quad & \theta \geq \theta_{\emptyset} \\ \text{Alternative:} \quad & \theta \leq \theta_{+} \end{aligned}$$

with $\theta_{+} < \theta_{\emptyset}$, and θ_{+} is chosen to represent the smallest difference that is clinically important.

  - Two-sided (equivalence) test:

$$\begin{aligned} \text{Null:} \quad & \theta = \theta_{\emptyset} \\ \text{Lower Alternative:} \quad & \theta \leq \theta_{-} \\ \text{Upper Alternative:} \quad & \theta \geq \theta_{+} \end{aligned}$$

with $\theta_{-} < \theta_{\emptyset} < \theta_{+}$. θ_{-} and θ_{+} denote the smallest important differences.

Statistical framework for trial monitoring: Selecting decision criteria

- A decision to stop needs to consider what has or has not been ruled out. For example
  - One-sided superiority test (assume small θ favors new treatment):
    - Stop for superiority when any harm ($\theta \geq \theta_{\emptyset}$) has been ruled out.
    - Stop for futility when important benefits ($\theta \leq \theta_{+}$) have been ruled out.
  - Two-sided (equivalence) test:
    - Stop for treatment *A* better than treatment *B* when inferiority of *A* ($\theta \leq \theta_{\emptyset}$) has been ruled out.
    - Stop for treatment *B* better than treatment *A* when inferiority of *B* ($\theta \geq \theta_{\emptyset}$) has been ruled out.
    - Stop for equivalence when important differences (either $\theta \geq \theta_{+}$ or $\theta \leq \theta_{-}$) have been ruled out.
- The hypotheses that have been ruled in/out are given by the interval estimate.

Overview of group sequential designs

Statistical framework for trial monitoring: Group sequential designs (superiority trial)

- Suppose that the trial is planned for $j = 1, ..., J$ interim analyses.
- Let $\hat{\theta}_j$ denote the estimated treatment effect at the jth analysis.
- Consider stopping criteria $a_j < d_j$ with:

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

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

- How should we choose these critical values?

Statistical framework for trial monitoring

Inadequacy of Fixed Sample Methods

- Suppose we simply ignore the fact that we are repeatedly testing our hypothesis
- We can quickly see the impact of this via simulation
  - Let $X_i \sim_{\text{iid}} \mathcal{N}(\theta, \sigma^2)$
  - $j = 1, ..., 4$ equally spaced analyses at 25, 50, 75, and 100 observations
  - Test statistic after n_j observations have been accrued

$$\bar{X}_{n_j} = \frac{1}{n_j} \sum_{i=1}^{n_j} X_i$$

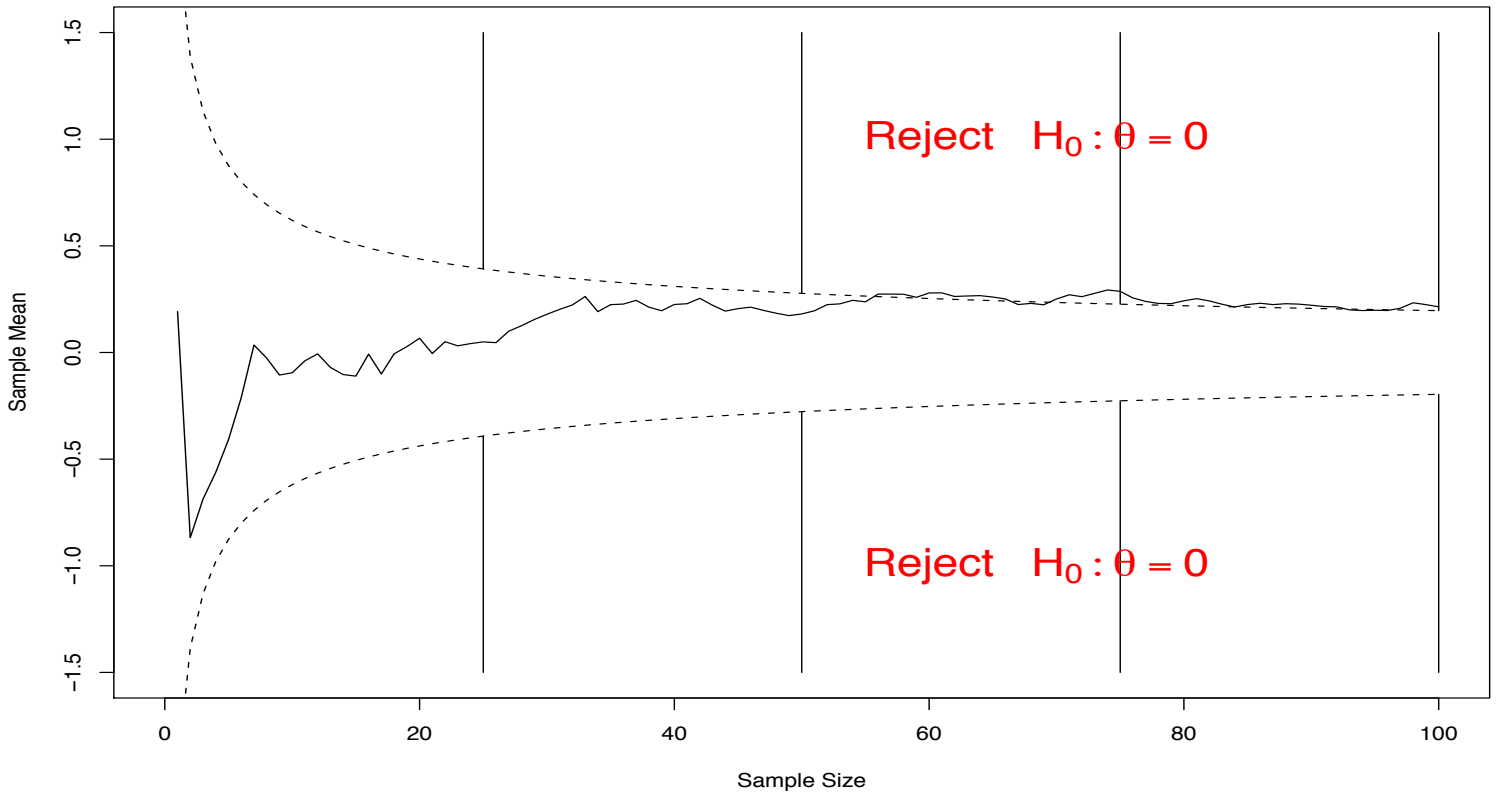
  - Test $H_0 : \theta = 0$ with level $\alpha = .05$
- Fixed sample methods (2-sided test): Reject H_0 first time

$$|\bar{X}_{n_j}| > z_{1-\alpha/2} \frac{\sigma}{\sqrt{n_j}}, \qquad j = 1, 2, 3, 4$$

Statistical framework for trial monitoring

Inadequacy of Fixed Sample Methods : Simulation

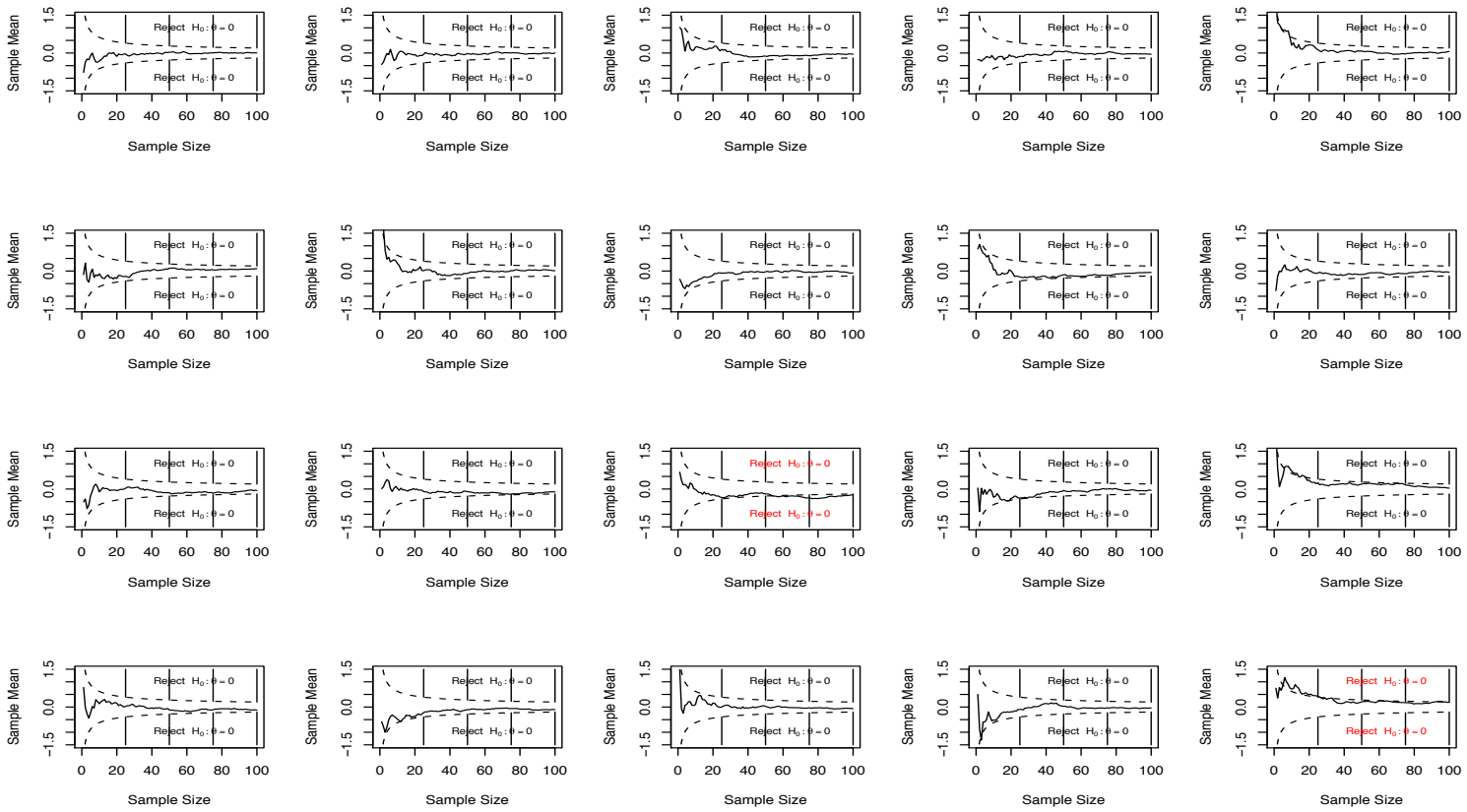
- Consider the sample path of the statistic for a single simulated trial



Statistical framework for trial monitoring

Inadequacy of Fixed Sample Methods : Simulation

- Consider the sample path of the statistic for 20 randomly sampled trials



Statistical framework for trial monitoring

Inadequacy of Fixed Sample Methods : Simulation

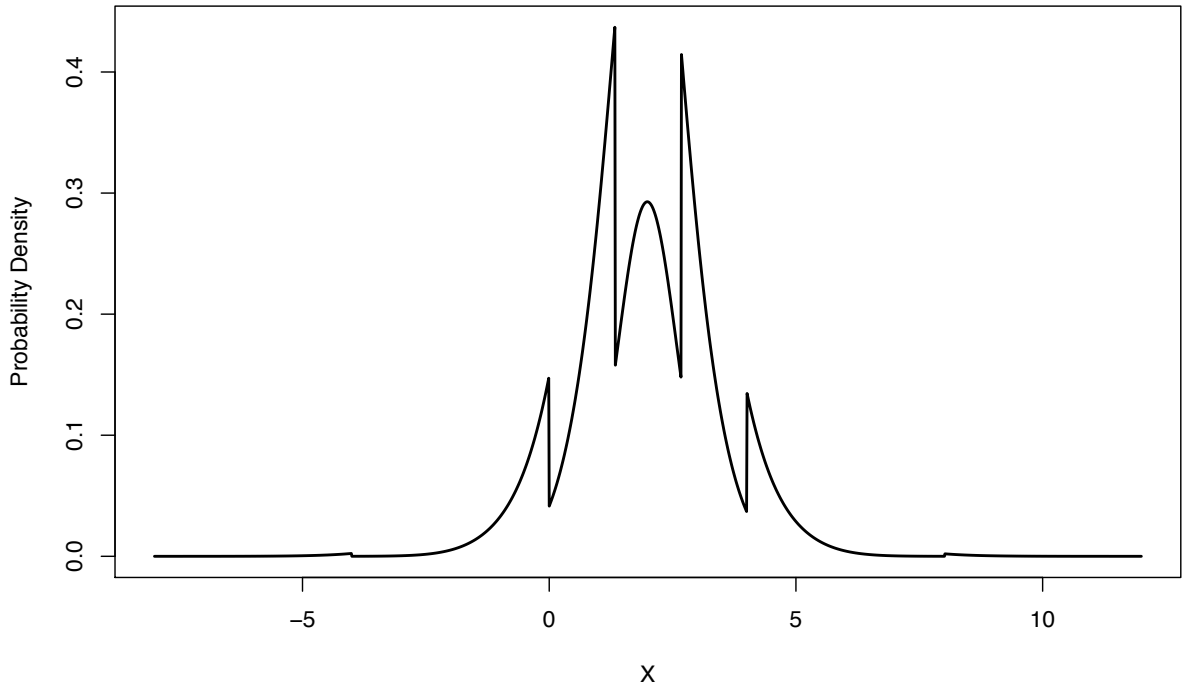
- Simulated type I error rate using fixed sample methods
- Based on 100,000 simulations

Significant at	Proportion Significant	Number Significant	Proportion Significant
Analysis 1	0.05075	Exactly 1	0.07753
Analysis 2	0.04978	Exactly 2	0.02975
Analysis 3	0.05029	Exactly 3	0.01439
Analysis 4	0.05154	All 4	0.00554
		Any	0.12721

Interim analyses require special methods

Sampling density for sequentially-monitored test statistic

- The filtering due to interim analyses creates non-standard sampling densities as the basis for inference.
- Sampling density depends on the stopping rule.
- In order to correct the type 1 error rate, we must be able to compute the density of the statistic that accounts for the possibility of stopping at interim analyses



Sampling density for sequentially sampled test statistic

- Let C_j denote the continuation set at the jth interim analysis.
- Let (M, S) denote the bivariate statistic where M denotes the stopping time ($1 \leq M \leq J$) and $S = S_M$ denotes the value of the partial sum statistic at the stopping time.
- The sampling density for the observation $(M = m, S = s)$ is:

$$p(m, s; \theta) = \begin{cases} f(m, s; \theta) & s \notin C_m \\ 0 & else \end{cases}$$

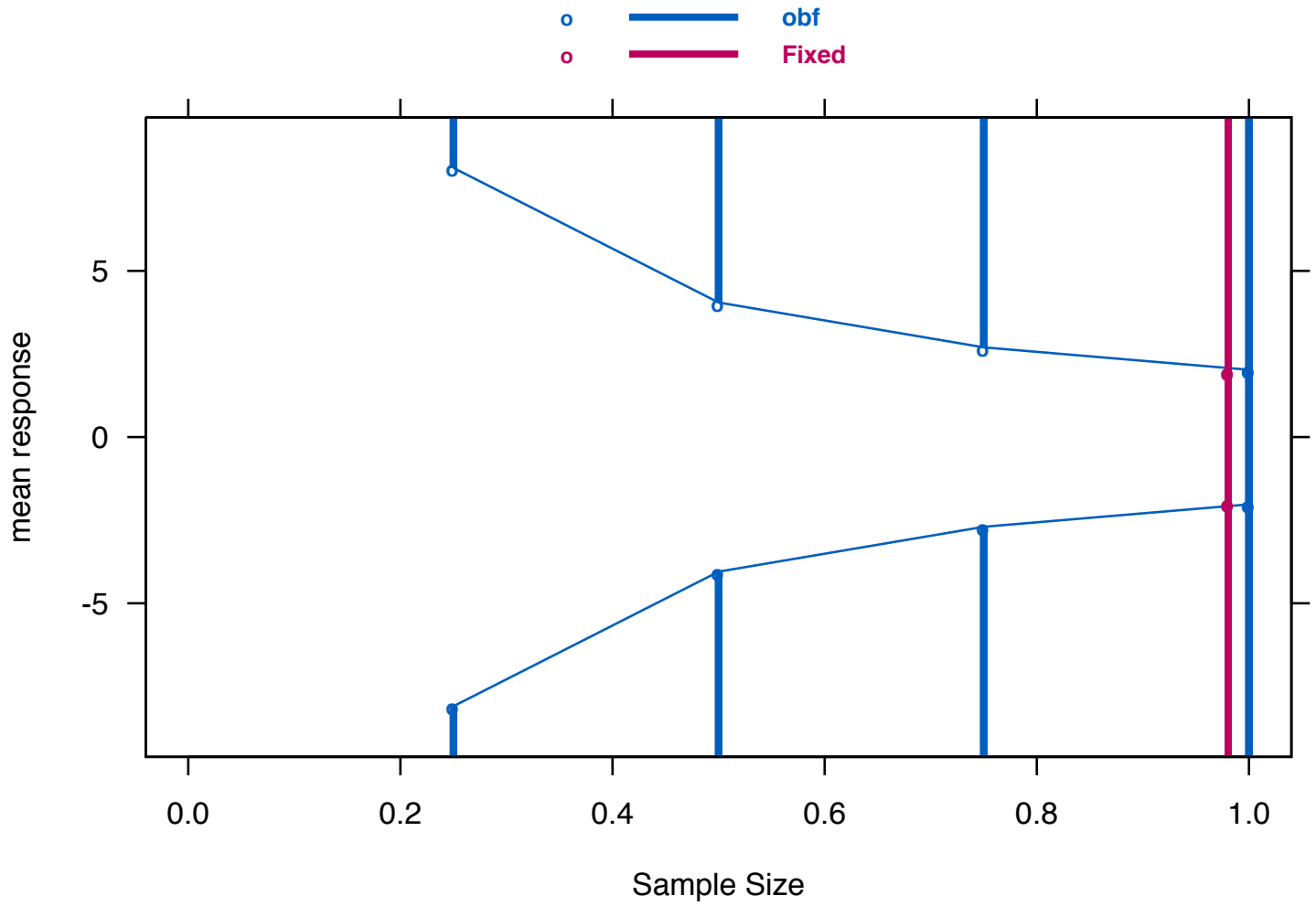
where the (sub)density function $f(j, s; \theta)$ is recursively defined as

$$\begin{aligned} f(1, s; \theta) &= \frac{1}{\sqrt{n_1 V}} \phi\left(\frac{s - n_1\theta}{\sqrt{n_1 V}}\right) \\ f(j, s; \theta) &= \int_{C_{(j-1)}} \frac{1}{\sqrt{n_j V}} \phi\left(\frac{s - u - n_j\theta}{\sqrt{n_j V}}\right) f(j-1, u; \theta)\, du, \\ & \quad j = 2, \ldots, m \end{aligned}$$

Types of group sequential designs

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

- Using the correct sampling density, we can choose boundary values that maintain experiment wise Type I error



Types of group sequential designs

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

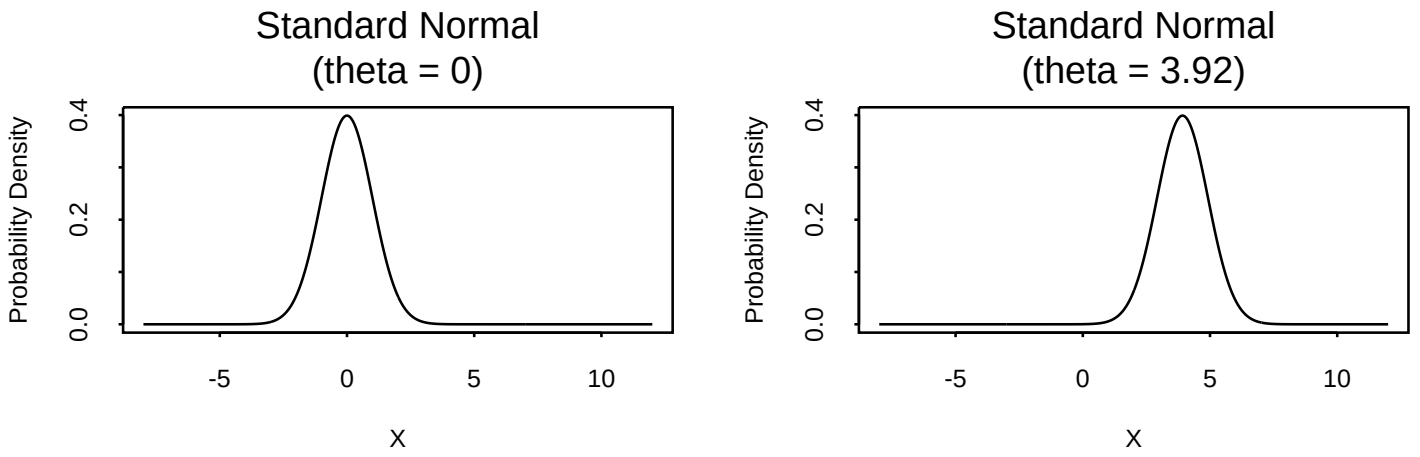
- Simulated type I error rate using fixed sample methods
- Based on 100,000 simulations

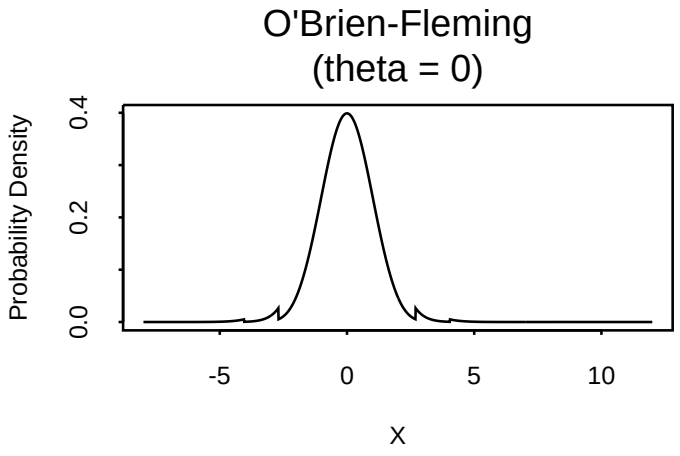
Significant at	Proportion Significant	Number Significant	Proportion Significant
Analysis 1	0.00006	Exactly 1	0.03610
Analysis 2	0.00409	Exactly 2	0.01198
Analysis 3	0.01910	Exactly 3	0.00210
Analysis 4	0.04315	All 4	0.00001
		Any	0.05019

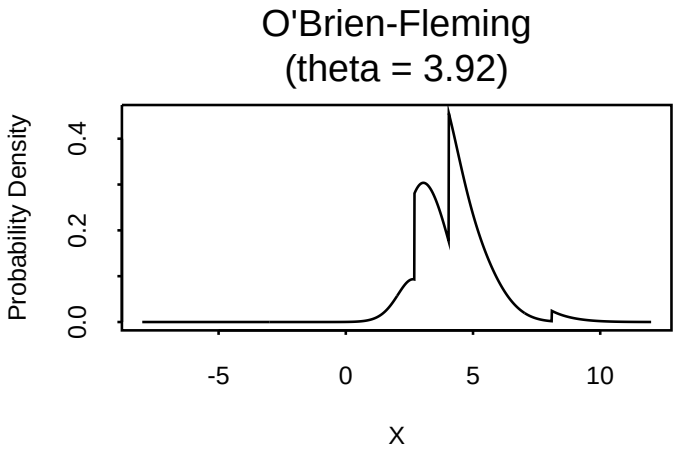
Types of group sequential designs

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

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







Types of group sequential designs

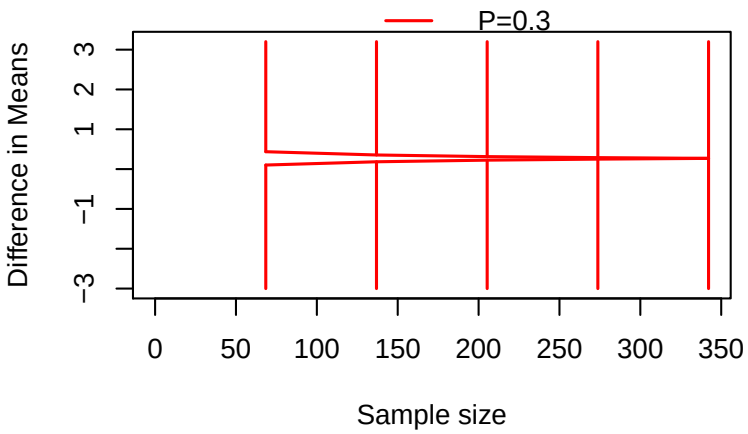
Boundary shape functions

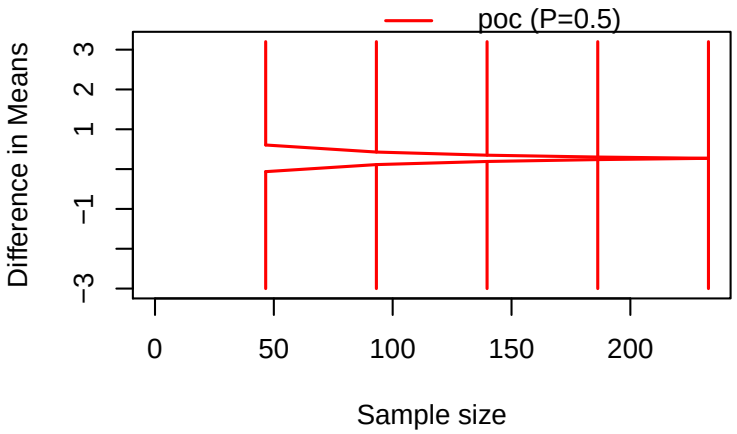
- There are an infinite number of stopping boundaries to choose from that will maintain a given family-wise error
  - They will differ in required sample size and power
- Kittelson and Emerson (1999) described a “unified family" of designs that are parameterized by three parameters (A, R, and P)
- Parameterization of boundary shape function includes many previously described approaches
  - Wang & Tsiatis boundary shape functions:
    - $A = 0$, $R = 0$, and $P > 0$
    - $P = 0.5$: Pocock (1977)
    - $P = 1.0$: O’Brien-Fleming (1979)
  - Triangular Test boundary shape functions (Whitehead):
    - $A = 1$, $R = 0$, and $P = 1$
  - Sequential Conditional Probability Ratio Test (Xiong):
    - $R = 0.5$, and $P = 0.5$

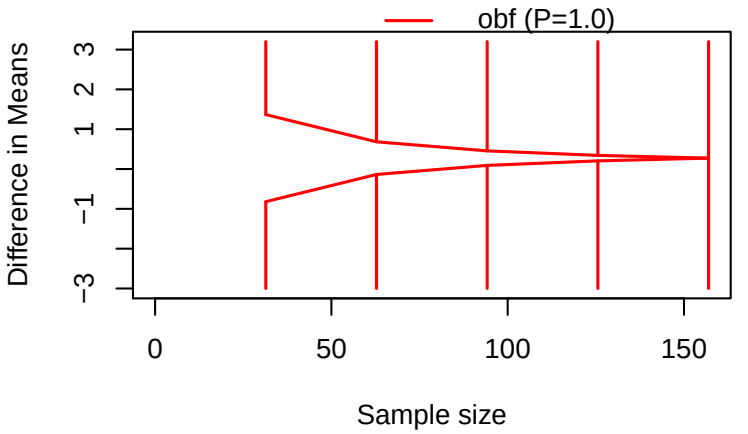
Types of group sequential designs

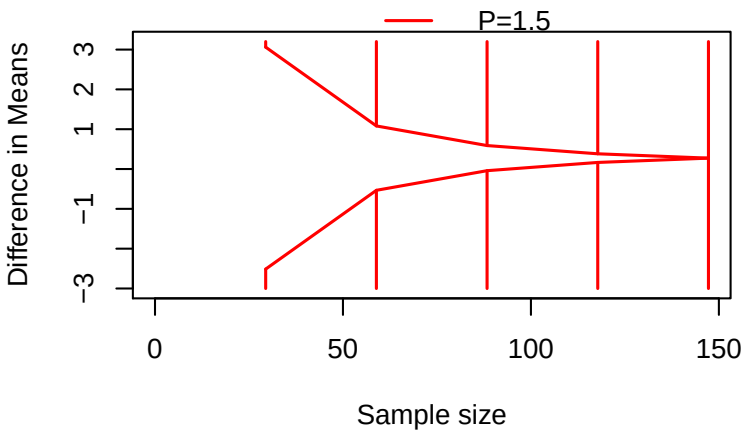
Boundary shape functions

- Consider differing choices of P

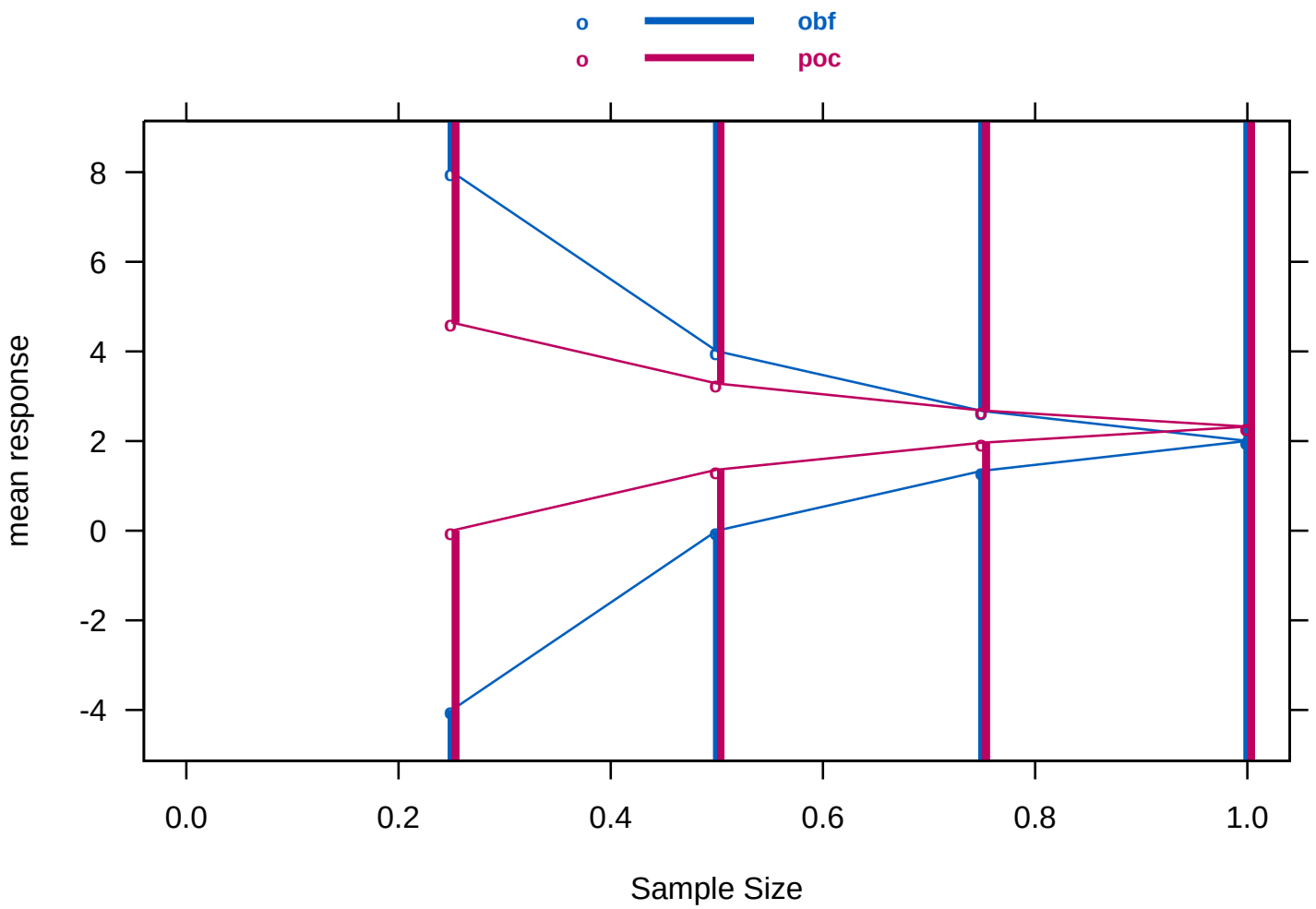








Example: OBF (P=1) versus Pocock (P=0.5) 1-sided designs

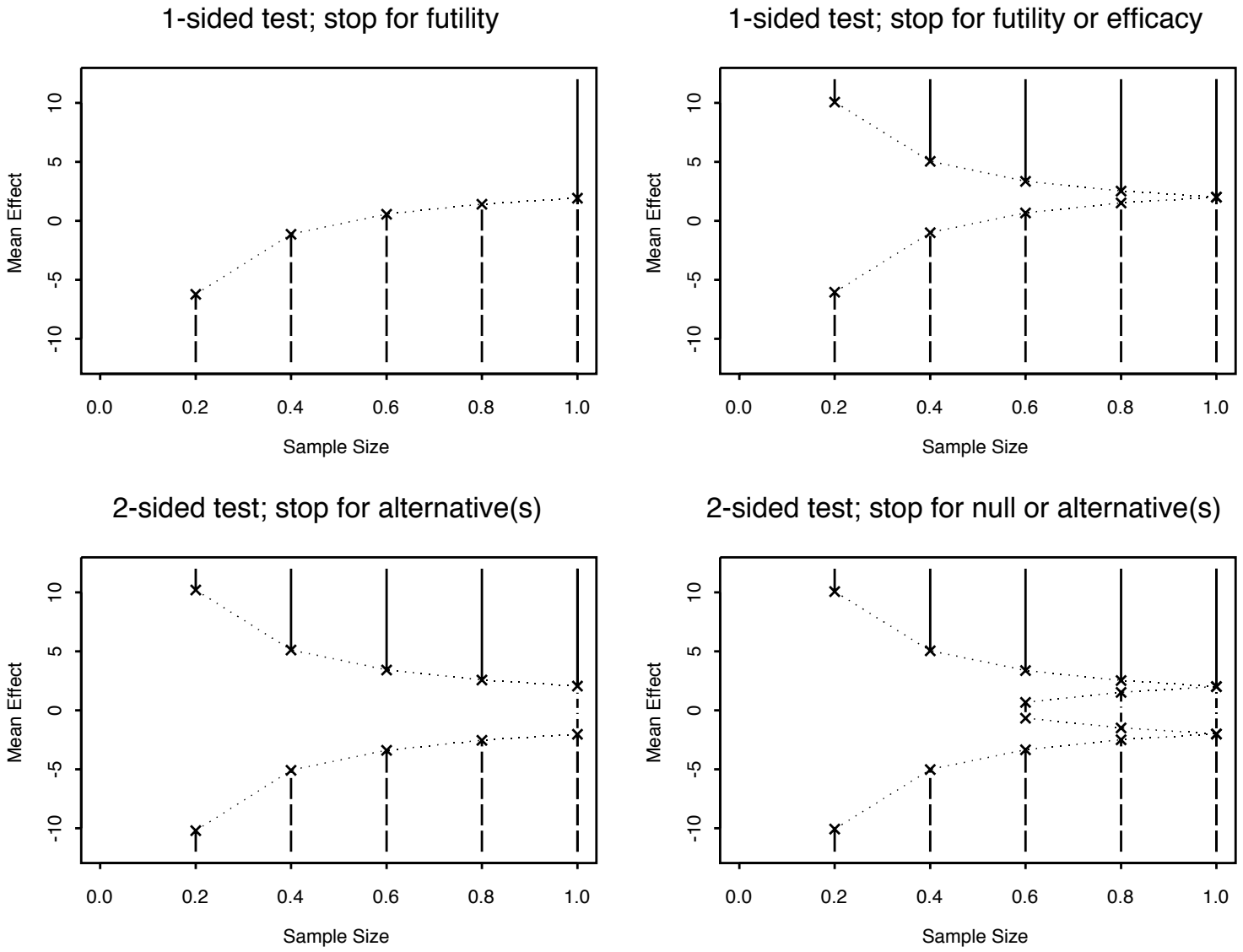


Types of group sequential designs

Group sequential designs can be formulated for various hypotheses

- Four design categories:
  - One-sided test; One-sided stopping
    (allow stopping for efficacy *or* futility, but not both)
  - One-sided test; Two-sided stopping
    (allow stopping for either efficacy or futility)
  - Two-sided test; One-sided stopping
    (allow stopping only for the alternative(s))
  - Two-sided test; Two-sided stopping
    (allow stopping for either the null or the alternative)

Four general design categories



Types of group sequential designs

So how should we choose a stoping rule?

- Consider appropriate type of hypothesis to test
- Maintain statistical design criteria of the fixed sample trial:
  - Type I error rate of $\alpha = 0.025$ (one-sided test) or $\alpha = 0.05$ (two-sided test).
  - Maintain maximal sample size (with potential loss of power)
  - Maintain power (with larger maximal sample size)
- Other considerations when selecting critical values:
  - Number of interim analyses
  - Timing of interim analyses
  - Degree of early conservatism
  - Characteristics of the sample size distribution:
    - Expected sample size (Average Sample Number; ASN)
    - Quantiles of the sample size distribution
    - Maximal sample size
    - Stopping probabilities at each of the interim analyses

Interim analyses require special methods

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

Case Study : Sepsis Trial

Background

- Critically ill patients often get overwhelming bacterial infection (sepsis), after which mortality is high
- Gram negative sepsis is often characterized by production of endotoxin, which is thought to be the cause of much of the ill effects of gram negative sepsis
- Hypothesis: Administering antibody to endotoxin may decrease morbidity and mortality
- Two previous randomized clinical trials showed a slight benefit
- There were no safety concerns at the inception of the trial

Case Study : Sepsis Trial

Definition of Treatment

- Single administration of antibody to endotoxin within 24 hours of diagnosis of sepsis
- Reductions in dose not applicable
- Ancillary treatments unrestricted

Case Study : Sepsis Trial

Defining the target population

- Patients in ICU with newly diagnosed sepsis
- Infected with gram negative organisms
  - culture proven
  - gram stain

Case Study : Sepsis Trial

Defining the Comparison Group

- Need to ensure scientific credibility for regulatory approval
- Crossover designs impossible
- Ultimate decision:
  - Single comparison group treated with placebo
    - Not interested in studying dose response
    - No similar current therapy (still ethical to use placebo)
  - Randomized
    - Allow for causal inference
    - No blocking

Case Study : Sepsis Trial

Defining the Outcomes of Interest

- **Goals**:
  - Primary: Increase survival
    - Long term (always best)
    - Short term (many other processes may intervene)
  - Secondary: Decrease morbidity
- **Refinement of the primary endpoint**
  - Possible primary endpoints
    - Time to death
    - Mortality rate at a fixed point in time
    - Time alive out of ICU during fixed period of time

Case Study : Sepsis Trial

Refinement of the primary endpoint

Option 1: Time to death (censored continuous data)

- Trial is likely to have early censoring due to logistical constraints of the trauma centers
- Such early censoring might place emphasis on clinically meaningless improvements in very short term survival
  - eg. We may be detecting differences in 1 day survival even though there is no difference in survival at 10 days

Case Study : Sepsis Trial

Refinement of the primary endpoint

Option 2: Mortality rate at a fixed point in time (binary data)

- Allows for choice of a *scientifically* relevant time frame
  - Treatment is a single administration; short half-life
- Allows for choice of a *clinically* relevant time frame
  - Avoids sensitivity to improvements lasting only short periods of time

Case Study : Sepsis Trial

Refinement of the primary endpoint

Option 3: Time alive out of the ICU during a fixed period of time (continuous data)

- Incorporates morbidity endpoints
- Addresses patient quality of life
- May be sensitive to clinically meaningless improvements depending upon the time frame chosen

Case Study : Sepsis Trial

Refinement of the primary endpoint

Final Choice: Mortality rate at a fixed point in time (binary data)

- Sponsor proposed 14 day mortality
- FDA countered with a suggestion of 28 day mortality

Case Study : Sepsis Trial

Method of analysis

- Test for differences in binomial proportions
  - Ease of interpretation
  - 28 day mortality not a rare event
  - 1:1 correspondence with tests of odds ratio (for known baseline event rates)
- No adjustment for covariates
- Statistical information dictated by mean variance relationship of Bernoulli random variables:
  - Let Y_{ki} denote binary response (mortality at 28 days) for i-th subject in group k, $k = 0, 1$
  - $Y_{ki} \sim \mathcal{B}(1, \theta_k)$
  - $\mathrm{E}[Y_{ki}] = \theta_k$ and $\mathrm{Var}[Y_{ki}] = \theta_k(1 - \theta_k)$

Case Study : Sepsis Trial

Definition of statistical hypotheses

Null hypothesis

- No difference in mortality between groups
- Estimated baseline rate
  - 28 day mortality: 30%
  - (needed in this case to estimate variability)

Alternative hypothesis

- One-sided test for decreased mortality
- Targeted 28 day mortality rate in antibody arm: 25%
  - 5% absolute difference in mortality

Case Study : Sepsis Trial

Criteria for statistical evidence

- Type I error: Probability of falsely rejecting the null hypothesis Standards:
  - Two-sided hypothesis tests: 0.050
  - One-sided hypothesis test: 0.025
- Power: Probability of correctly rejecting the null hypothesis (1-type II error)
- Popular choice: 80% power

Case Study : Sepsis Trial

Determination of sample size

- Sample size chosen to provide desired operating characteristics
  - Type I error : 0.025 when no difference in mortality
  - Power : 0.80 when 5% absolute difference in mortality
  - Statistical variability based on mortality rate of 30% in placebo arm

Case Study : Sepsis Trial

Determination of sample size

- General sample size formula:
  - δ = standardized alternative
  - Δ = difference between null and alternative treatment effects
  - V = variability of a single sampling unit
  - n = number of sampling units

$$n = \frac{\delta^2 V}{\Delta^2}$$

Case Study : Sepsis Trial

Determination of sample size

- Parameter values in the present case:
  - $\delta = (z_{1-\alpha} + z_{\beta})$ with $\alpha = 0.025$ and $\beta = 0.80$
  - $\Delta = \theta_{1,H_1} - \theta_{0,H_1} = -0.05$
  - $V = \theta_{1,H_1}(1 - \theta_{1,H_1}) + \theta_{0,H_1}(1 - \theta_{0,H_1}) =$ $.25 \times .75 + .3 \times .7 = .3975$
  - n = sample size per arm

$$n = \frac{\delta^2 V}{\Delta^2} = \frac{(1.96 + .841)^2 \times .3975}{(-.05)^2} = 1247.97 \rightarrow 1248$$

Case Study : Sepsis Trial

Resulting Fixed sample design

- Problem: Sponsor was concerned that 2496 (2×1248) patients would be logistically infeasible and wanted to consider a design with 1700 patients
- Operating characteristics with N=1700:
  - Critical value : -0.0424
  - 64% power for alternative of 5% absolute difference; 90% power for alternative of 7% absolute difference; Corresponding p-value : 0.025
  - 95% confidence interval : (-0.085, 0)
  - Interpretation: Smallest magnitude of (observed) effect which would result in a significant result is a 4.24% decrease in mortality on the treatment arm with corresponding CI (-0.085, 0).

Example: Sepsis Trial

Addition of interim analyses

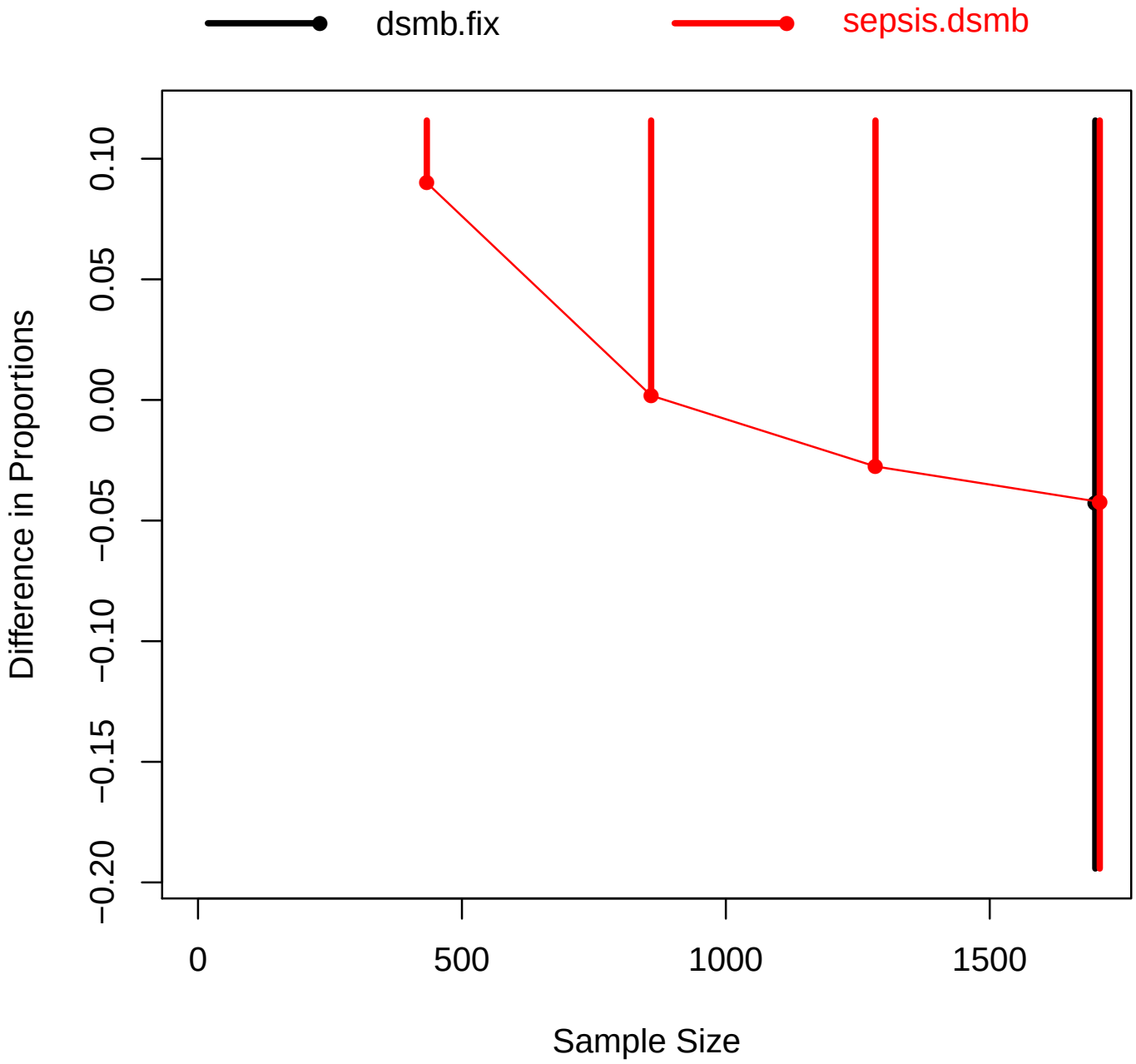
- FDA requires an interim safety analysis
- DSMB considers 4 interim analyses to stop for harm or futility using an O'Brien-Fleming stopping rule

```
PROBABILITY MODEL and HYPOTHESES:
   Theta is difference in probabilities (Treatment - Comparison)
   One-sided hypothesis test of a lesser alternative:
          Null hypothesis : Theta >=  0.00    (size  = 0.0250)
    Alternative hypothesis : Theta <= -0.07    (power = 0.9021)

STOPPING BOUNDARIES: Sample Mean scale
                     Efficacy Futility
    Time 1 (N=  425)     -Inf   0.0883
    Time 2 (N=  850)     -Inf   0.0019
    Time 3 (N= 1275)     -Inf  -0.0269
    Time 4 (N= 1700)  -0.0413  -0.0413
```

Example: Sepsis Trial

- Stopping boundaries



Example: Sepsis Trial

Addition of interim analyses

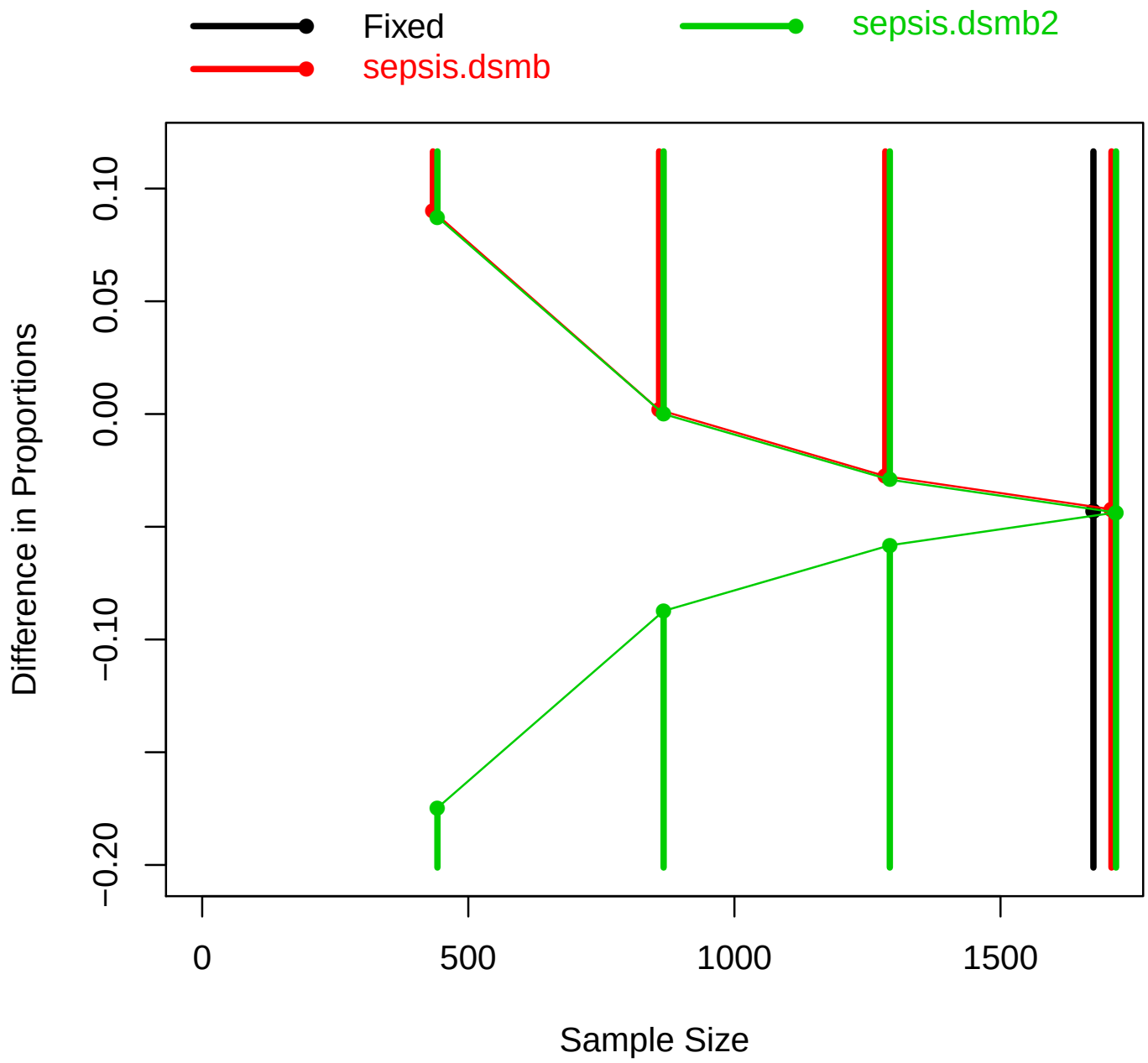
- Sponsor and DSMB would also like to consider stopping for efficacy
- Consider an O'Brien-Fleming boundary for both efficacy and futility

```
PROBABILITY MODEL and HYPOTHESES:
   Theta is difference in probabilities (Treatment - Comparison)
   One-sided hypothesis test of a lesser alternative:
           Null hypothesis : Theta >=  0.00    (size  = 0.0250)
    Alternative hypothesis : Theta <= -0.07    (power = 0.8947)
   (Emerson & Fleming (1989) symmetric test)

STOPPING BOUNDARIES: Sample Mean scale
                     Efficacy Futility
    Time 1 (N=  425)  -0.1710   0.0855
    Time 2 (N=  850)  -0.0855   0.0000
    Time 3 (N= 1275)  -0.0570  -0.0285
    Time 4 (N= 1700)  -0.0427  -0.0427
```

Example: Sepsis Trial

- Stopping boundaries



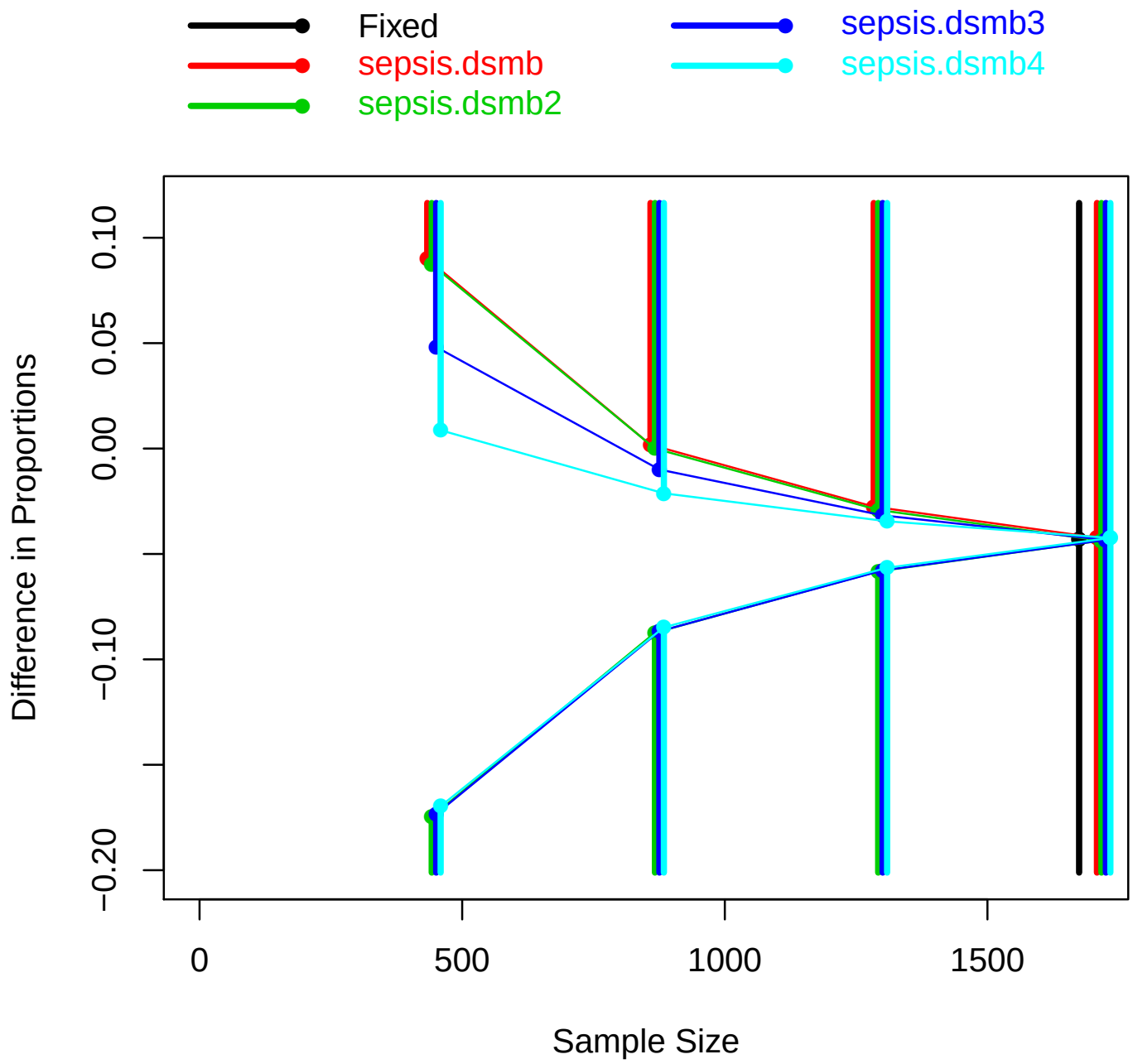
Example: Sepsis Trial

Addition of interim analyses

- DSMB sought a design with less early conservatism for futility
- Sponsor considered a Pocock futility bound and something between an O'Brien-Fleming and Pocock design

Example: Sepsis Trial

- Stopping boundaries



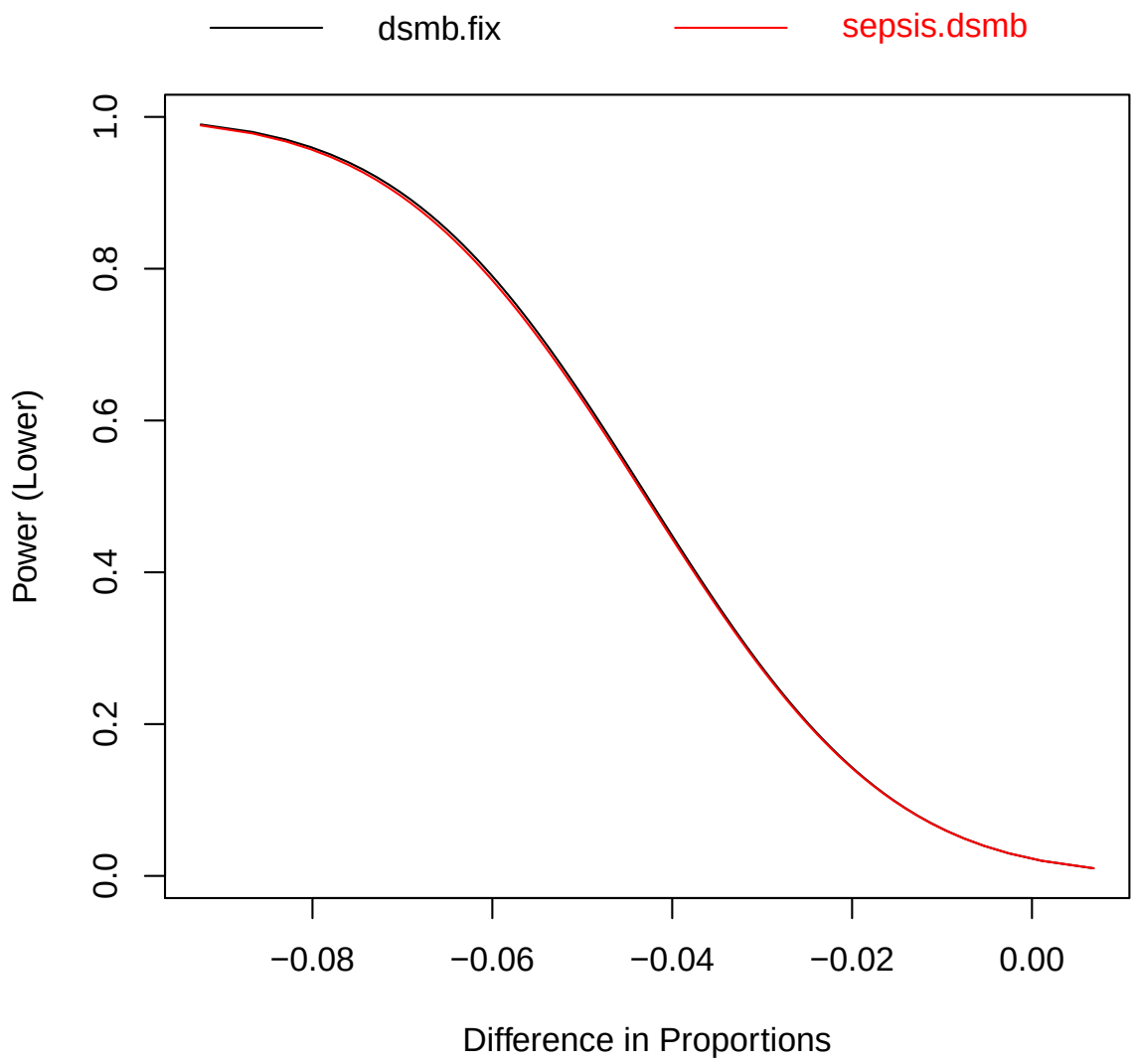
Example: Sepsis Trial

Choosing a boundary

- In order to choose between the considered designs, need to consider operating characteristics
  - Point estimates of treatment effect at boundary decisions
  - Confidence intervals resulting from decisions on the boundary
  - Statistical power as a function of treatment effect
  - Sample size distribution as a function of treatment effect

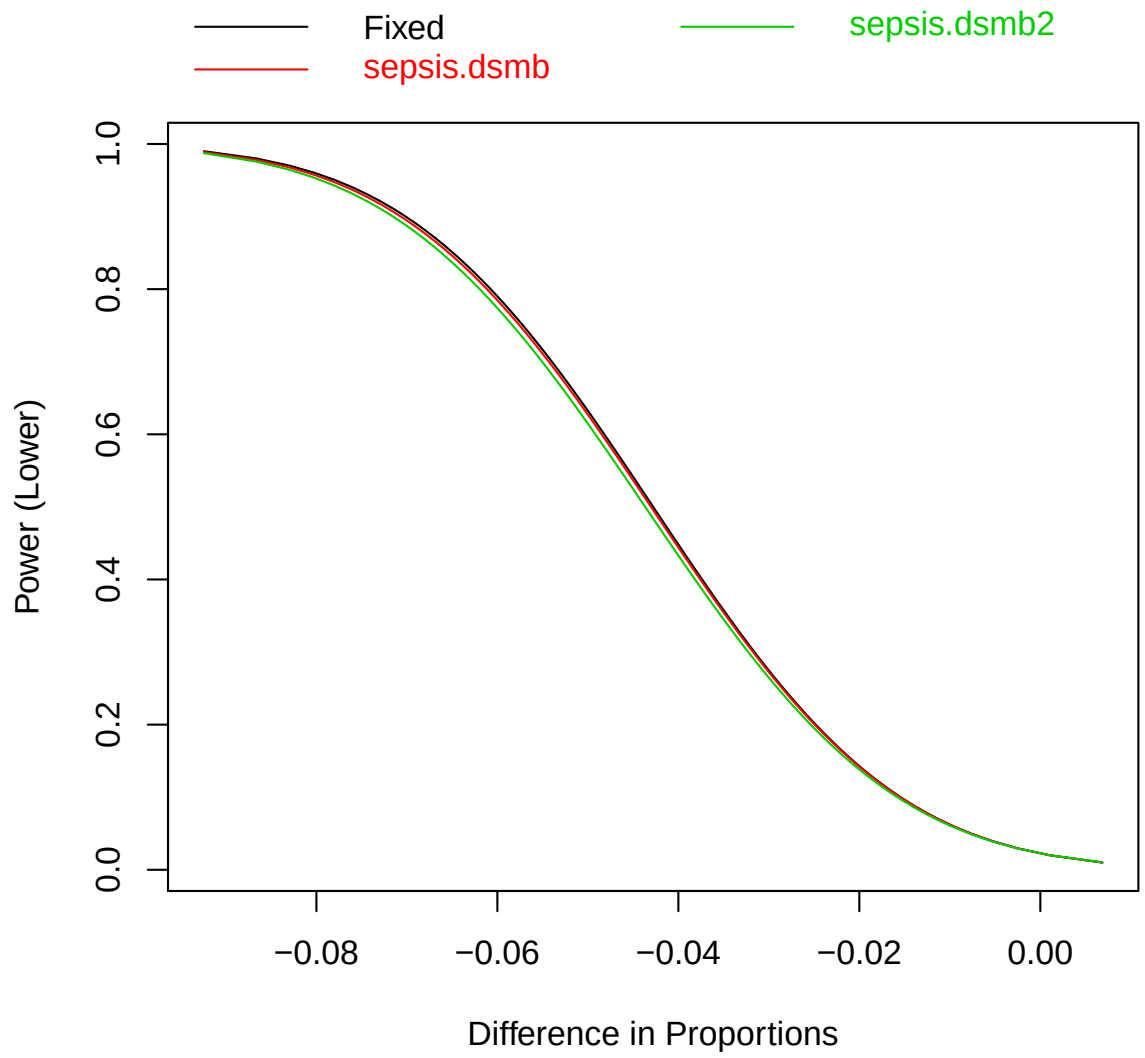
Example: Sepsis Trial

- Comparing power (adding futility-only stopping):



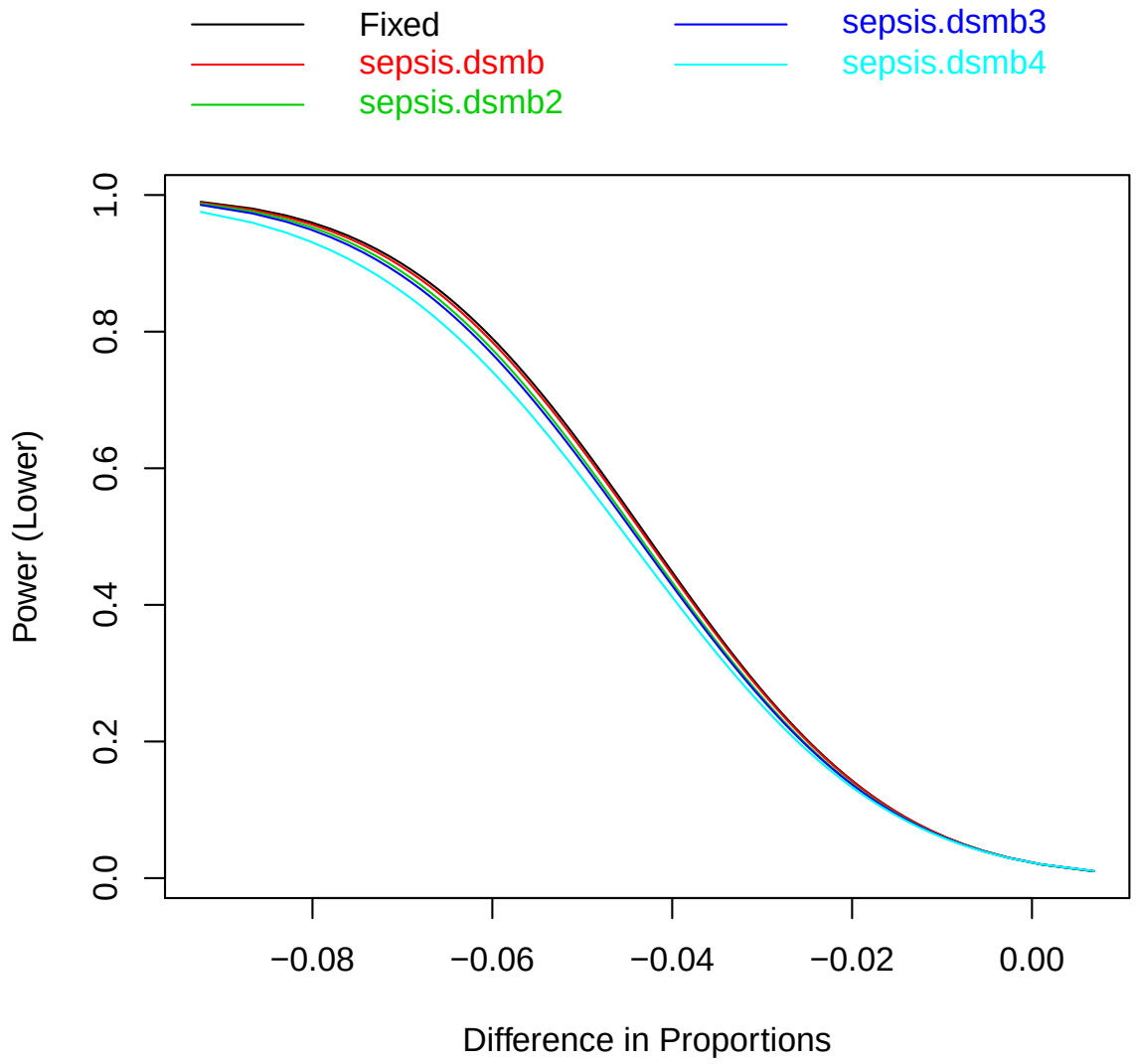
Example: Sepsis Trial

- Comparing power (adding futility and efficacy stopping):



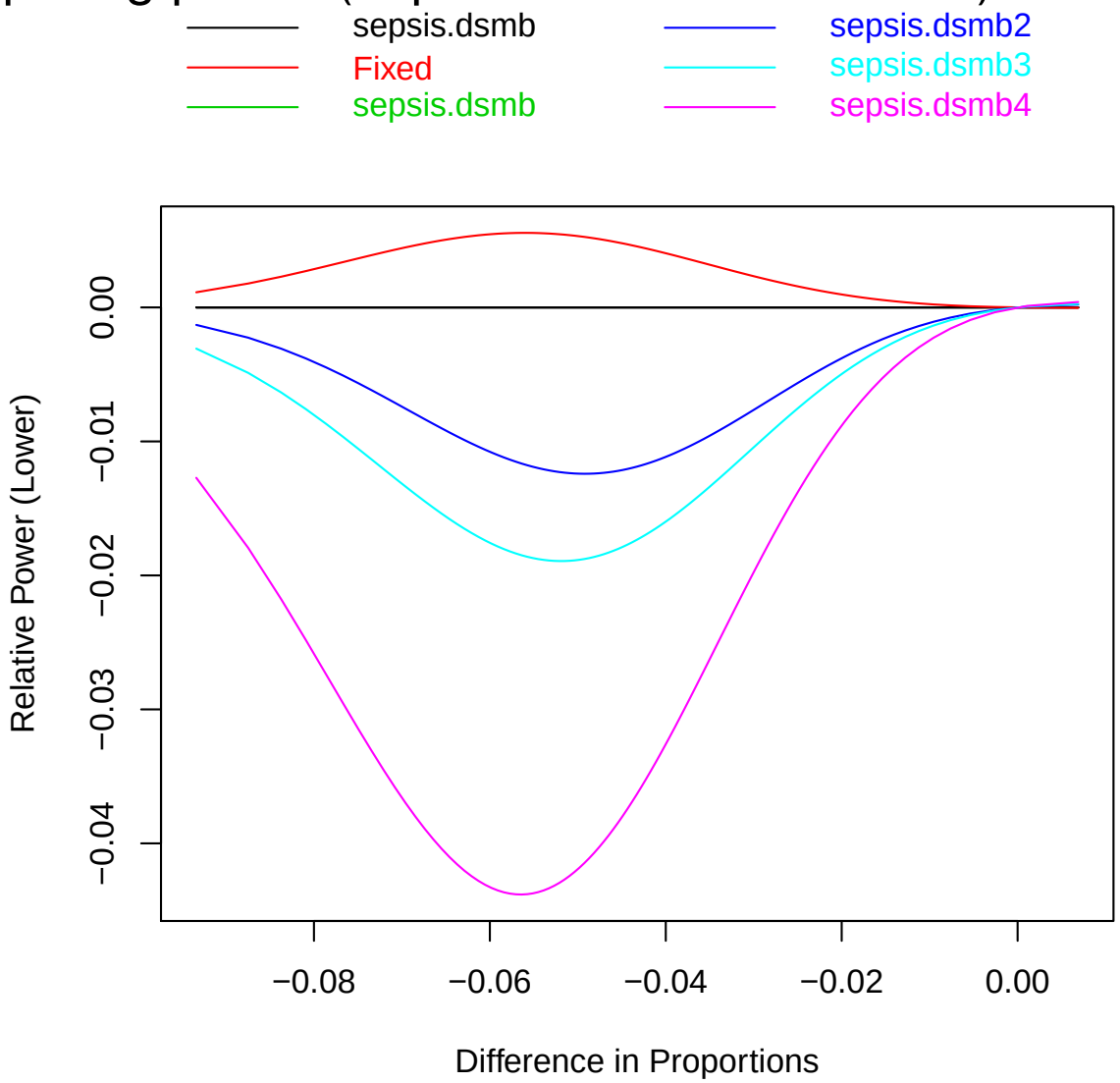
Example: Sepsis Trial

- Comparing power (effect of conservatism):



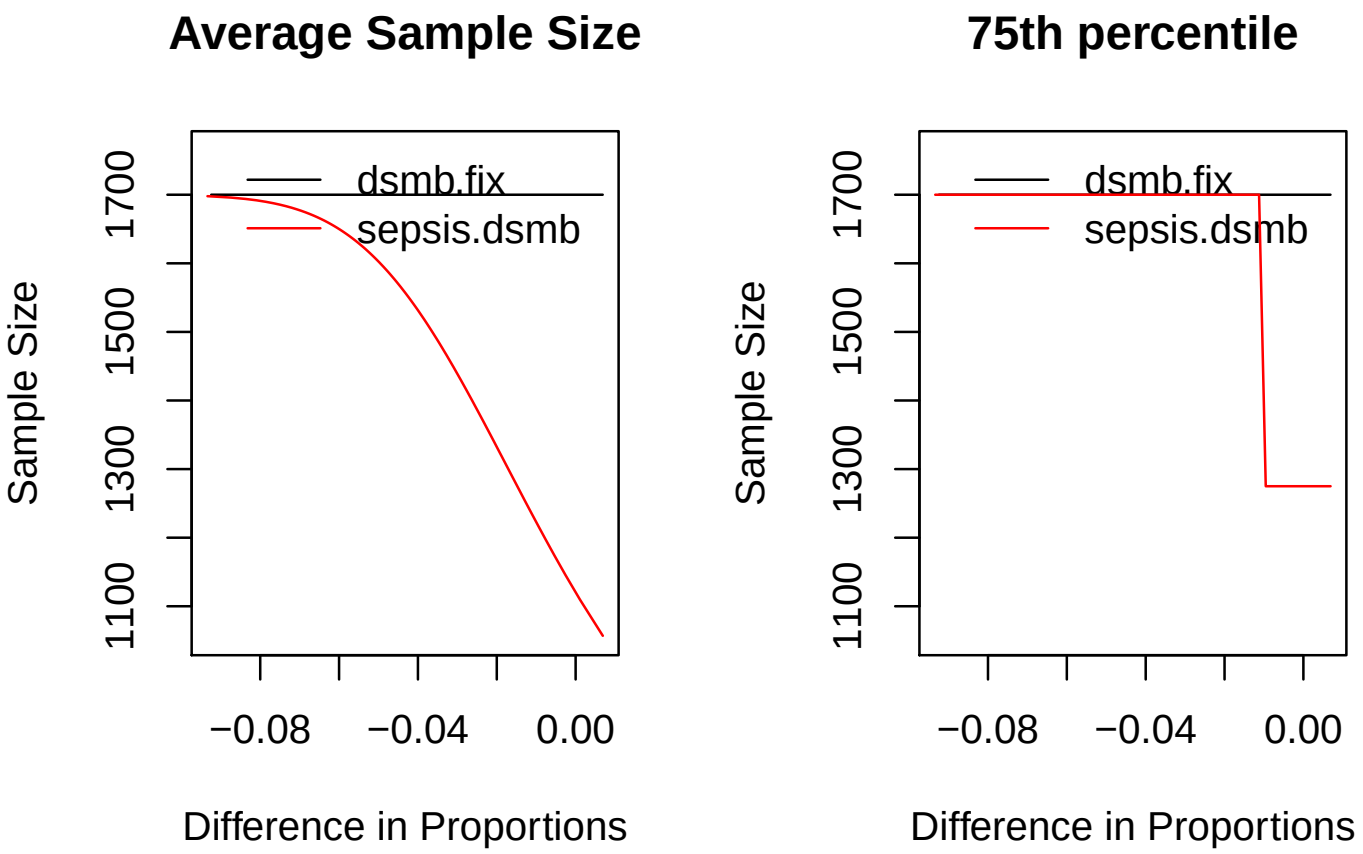
Example: Sepsis Trial

- Comparing power (sepsis.dsmb as reference):



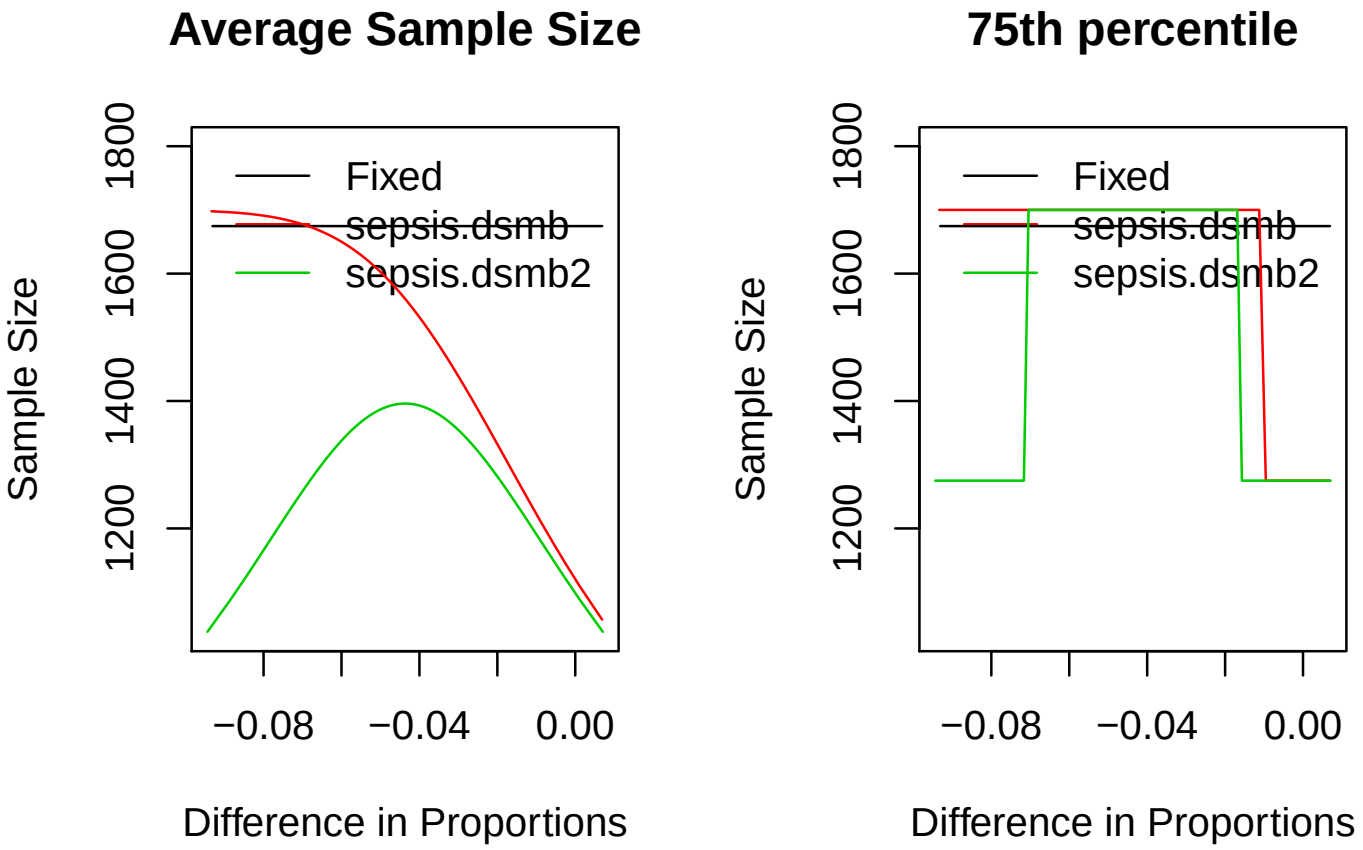
Example: Sepsis Trial

- Comparing expected sample size (ASN): adding futility-only stopping:



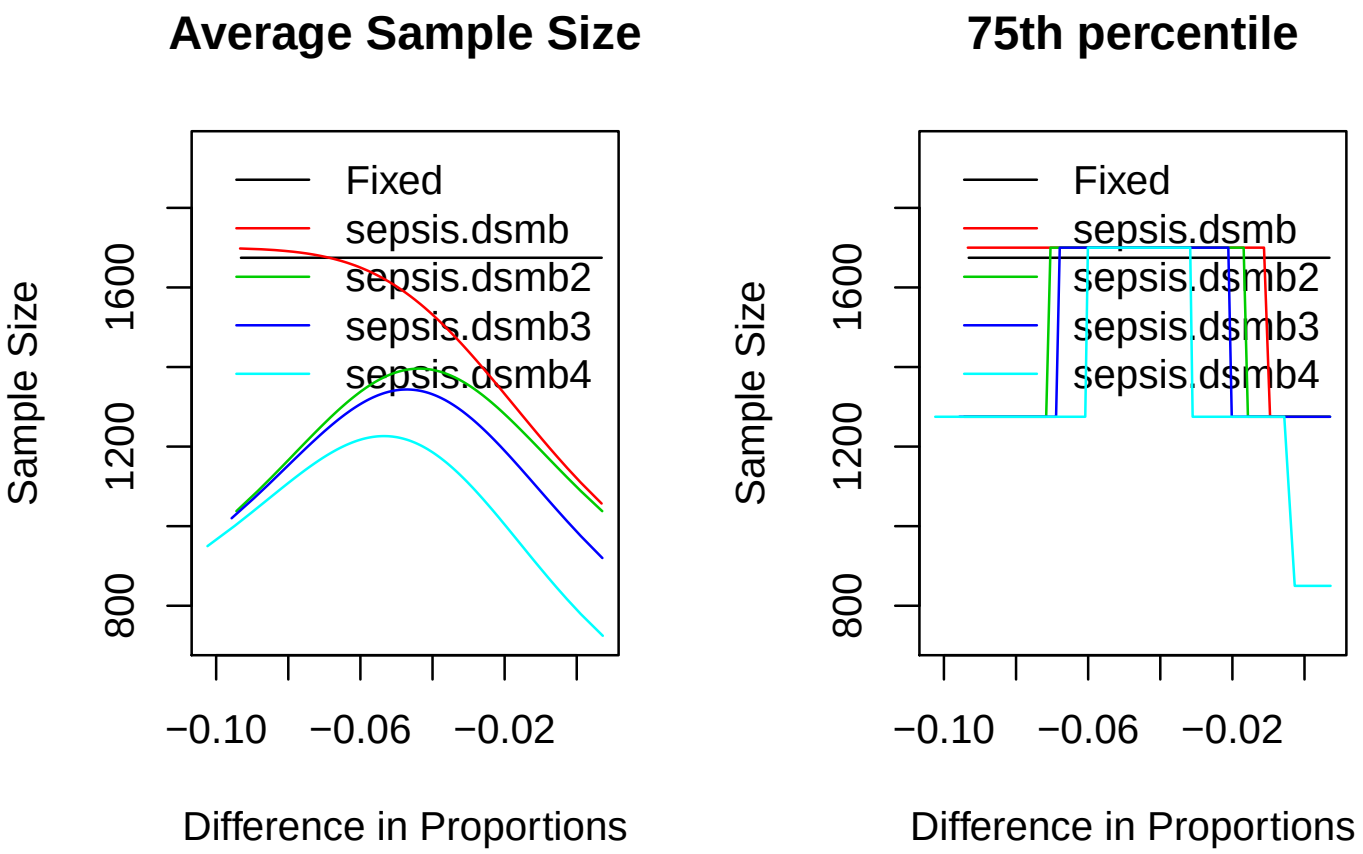
Example: Sepsis Trial

- Comparing expected sample size (ASN): futility and efficacy stopping:



Example: Sepsis Trial

- Comparing expected sample size (ASN): early conservatism:



Example: Sepsis Trial

General behavior of interim analyses

- Decreasing early conservatism gave smaller ASN for unimportant benefits.
- Decreasing early conservatism also reduces power for efficacy.

Example: Sepsis Trial

General behavior of interim analyses

- For any given sample size, adding interim analyses reduces power.
- For any given power, adding interim analyses increases the sample size.
- Having fewer interim analyses:
  - Leads to properties (maximal sample size, power, etc) that are closer to those of a fixed sample study.
  - However, ASN may be larger and stopping probabilities lower.
- Having more early conservatism:
  - Leads to properties (maximal sample size, power, etc) that are closer to those of a fixed sample study.
  - However, ASN may be larger and stopping probabilities lower.

Introduction to Clinical Trials - Day 2

Session 7 - Data Management in Clinical Trials

Presented July 26, 2016

Susanne J. May
Department of Biostatistics
University of Washington

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

Planning for Data Collection

Paul Dickson, in The Official Rules, Delacorte Press, 1978, gives this:

Stamp's Statistical Probability "The government [is] extremely fond of amassing great quantities of statistics. These are raised to the nth degree, the cube roots are extracted, and the results are arranged into elaborate and impressive displays. What must be kept ever in mind, however, is that in every case, the figures are first put down by a village watchman, and he puts down anything he damn well pleases."

(Attributed to Sir Josiah Stamp, 1840-1941, H.M. collector of inland revenue.)

Planning for Data Collection

Ultimate goal of an RCT

- The goal of a RCT is to find effective treatment indications
- At the conclusion, this will require reporting the experiment

1. Overall goal
2. Specific aims
3. Materials and Methods
   - Patients, dosing, adherence to monitoring
4. Results
   - Disposition, compliance, adverse events, outcomes
5. Conclusions

Planning for Data Collection

Role of Data : Overall goal and specific aims

- Goal / aims ideally determined prior to start of study
- BUT, the question actually answered is specific to
  - the subjects actually sampled
  - the methods actually used
  - the data actually gathered
  - the analysis actually performed
- Generalization of results depends on all of the above

Role of Data : Materials

- Eligibility criteria are usually broad
- Need to describe the population actually sampled
- Need to describe how the sample might differ from the ultimate target population

Role of Data : Materials

Conceptual framework

- Population of patients with disease
  - Definition of disease by cause vs signs / symptoms
- Subpopulation with disease targeted by intervention
  - “Disease" truly defined by treatment?
- Subpopulation eligible for study accrual
  - Restricted due to general clinical trial setting
- Eligible patients from which sampled
  - Restricted due to specific clinical trial (location, time)
- Study sample
  - Restricted due to willingness to participate
- Analysis sample
  - Data collection

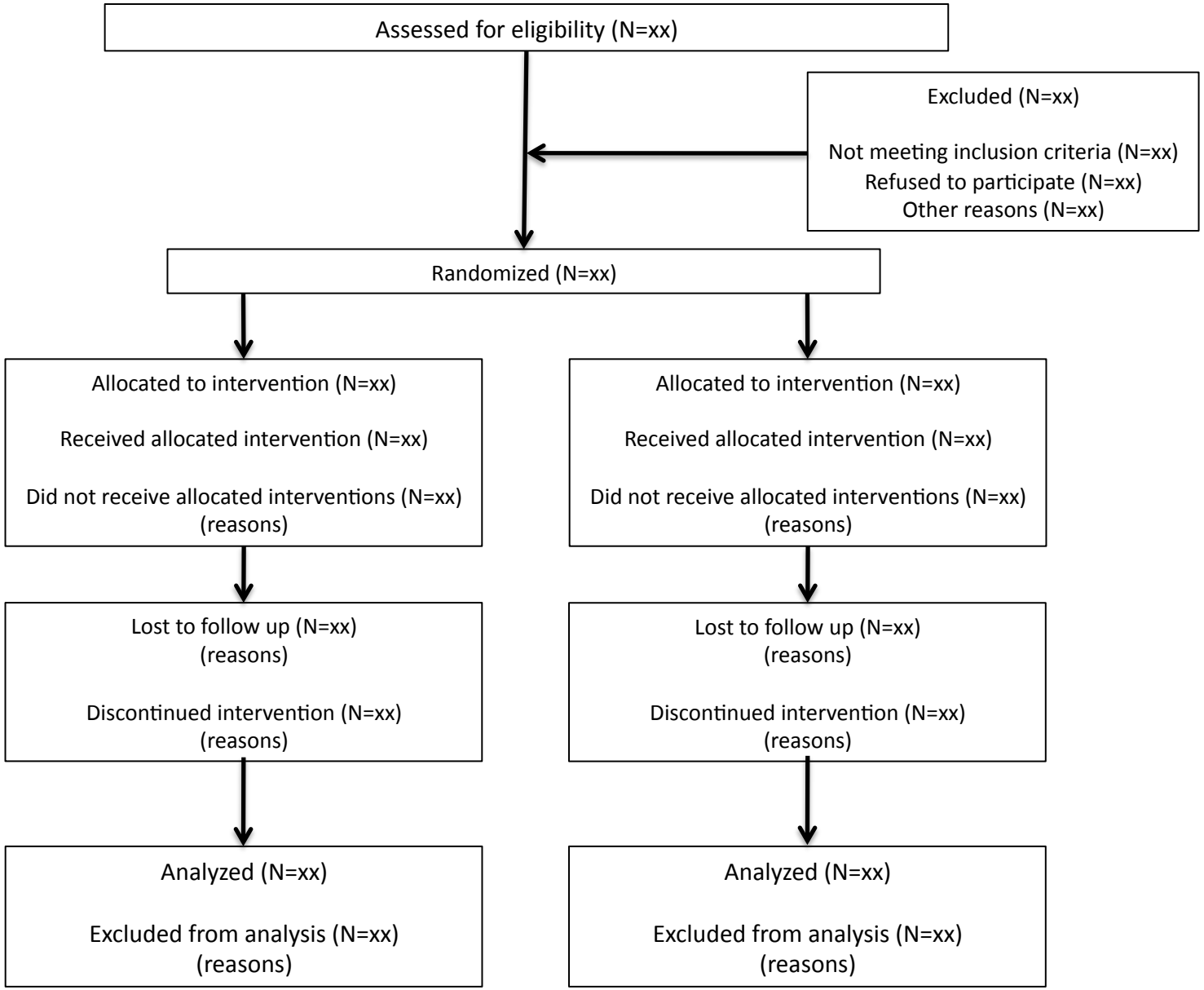
Role of Data : Materials

Generalizability

- CONSORT: Consolidated Standards of Reporting Trials
- Evidence based, minimum standards
- Report flow of patients from screening to collection of primary outcomes
  - Screened
  - Enrolled
  - Randomized
  - Completed

Role of Data : Materials

CONSORT Diagram



Role of Data : Materials

Initial Screening Data

- Source of screened patients
- Number screened
- Characteristics (may require consent)
  - Demographics
  - Disease characteristics
- Reasons for ineligibility
  - Inclusion criteria
  - Exclusion criteria
  - No participation
    - Unable to contact
    - Refused participation

Role of Data : Materials

Screening Visit(s) Data

- Consent for screening
- Contact information: Name, address, alternative contacts...
- Demographics: Sex, age, race, ethnicity...
- Disease characteristics: Duration, severity,...
- Prior and ongoing treatments
- Eligibility data
  - Inclusion criteria
  - Exclusion criteria
- Consent for randomization

Role of Data : Materials

Baseline Visit(s) Data

- Characterize patients
  - Severity of disease, concomitant disease...
- Baseline measures of outcomes
  - Concomitant medications
  - Adverse events
  - Efficacy outcomes (eg. initial tumor size for progression)
- Note differing detail needed for screening vs baseline

Role of Data : Materials

Run-in Data (if applicable)

- Placebo: All patients take placebo
  - Washout vs assessing compliance
  - Patients may be blinded to existence of run-in
- Active: All patients take experimental therapy
  - Allows randomized comparison of efficacy in patients actually taking drug
    - Randomized withdrawal of drug (among “responders"?)
    - Usually patients aware of run-in
  - Assess tolerability for AEs
  - Assess compliance

Role of Data : Materials

Randomization Data

- Documentation of eligibility
- Informed consent
- Stratification variable
- Variables needed for determination of dosing
  - Weight, BSA, renal function, severity of disease...
- Time, date of randomization
- Documentation of assigned group (blinded)
  - Cluster?
- Receipt of first treatment: time, date

Role of Data : Results

Treatment Data : Why

- Intention to treat analysis is the standard for efficacy
- Patients are analyzed in assigned group irrespective of their compliance
- Compliance data is an outcome
  - Assess possible AEs
  - Assess possible mechanism for lack of effect
  - Describe realized exposure to treatment
  - Exploratory analyses for dose / response?
- Safety analyses are typically analyzed according to drug exposure
  - AEs / SAEs occurring within 28 days (?) of last dose

Role of Data : Results

Treatment Data : What

- Initial assignment
  - Dose, administration, frequency, duration, ancillary treatments
- Protocol specified modifications
  - Dose reduction / escalation / holidays
  - Date, time, reasons for change (eg. AE, efficacy or lack of efficacy)
- Patient compliance
  - Dose, frequency, duration
  - ntermittent vs permanent change

Role of Data : Results

Treatment Data : How

- Protocol specified modifications
  - Regularly scheduled visits
  - Interim visits
- Patient compliance
  - Patient diaries
  - Pill counts
  - Clocks on container lids
  - Biochemical measures: blood, biopsies

Role of Data : Results

Patient Monitoring Data : Safety

- Protocol defined safety endpoints
  - Clinical events, subclinical laboratory measurements
- Adverse events
  - Review of interim AEs at regular visits
  - Undesirable clinical events that occur during the study
    - Treatment emergent: new or exacerbated
    - Classification (e.g. MEDRA), grade of severity
  - Treatment relatedness (but do not necessarily believe)
- Serious adverse events
  - Fatal, life-threatening, hospitalization or prolongation, birth defects
  - Expedited reporting if unexpected

Role of Data : Results

Patient Monitoring Data : Efficacy

- Protocol defined efficacy endpoints
- Clinical events
  - Create patient symptoms
- Quality of life
- Subclinical events
  - Signs thought to be indicative of clinical risk
  - Protocol specified monitoring schedules of
    - Patient performance (FEV, 6 minute walk, etc.)
    - Blood
    - Tissue biopsies
    - Radiology

Role of Data : Results

Missing Data : Efficacy

- Lack of training: Patients, investigators
  - Off study drug
  - Decline invasive procedures
  - Withdraw consent
- Poor endpoint definition
  - All randomized patients must have defined outcome
    - E.g, Quality of life after death, GFR in dialysis, symptom relief with noncompliance
- Sloppy conduct of RCT
  - Excessive loss of follow-up

Patient Monitoring Data: Compliance

- Patient adherence to measurement of outcomes
  - Clinic visits
    - Timing relative to window
  - Outcome assessments : Efficacy
    - Blood, tissue samples; radiology, special exams
    - Withdrawn consent for invasive procedures?
  - Outcome assessments : Adverse events
    - Periodic reports per protocol
    - Capture of interim SAEs

Role of Data : Results

Patient Monitoring Data: Logistics

- Patient change of address
  - (sometimes schedule phone visits to maintain contact)
- Site compliance with timeliness completeness

Role of Data : Results

End of Study

- Reason for stopping study
  - Completion per protocol
    - May be off study drug but still followed
    - Death
    - Withdrawn consent (Reasons)
- Permission for further follow-up
  - Change of address
- Conjectured treatment assignment?

Data Collection

Sources of Data

- Subject self report
- Proxy for subject
- Clinic staff and study records
  - Standard medical care
  - Protocol specified procedures
- Medical records
- Laboratory, radiology, pathology
  - Local vs central labs
- Adjudication panels
- Public health records
  - Registries
  - National Death Index?

Data Collection Methods

- Forms
  - Abstracted from medical records
    - Indication bias
  - Completed by subject
  - Completed by proxy
  - Administered by study personnel
  - Completed by clinic staff, study personnel
  - Completed by adjudication panels
- Data files
  - E.g., laboratory, Medicare, National Death Index

Data Collection

- Development of forms
  - Administrative information
    - For follow-up, etc.
    - Often text
  - Scientific information
    - Needs to be appropriate for statistical analysis
    - Free text is difficult to analyze
    - Coding of response by person closest to the source

Data Collection

- Development of forms (cont.)
  - Format of forms should facilitate
    - Completion of form
    - Brief as possible
    - Make sure no portions overlooked
    - “skip patterns", two columns, back of page
    - Cover all cases (explicit “does not apply")
  - Data entry
    - Coding on form

Data Collection

- Issues in form development
  - Number of distinct forms
  - Guidance to the subject, clinic staff on form
    - Study specific definitions
    - Indications for study procedures, other forms
    - Convenience versus increased length
  - Manual and training for form completion
  - Forms for subject vs proxy vs administered
  - Translations
  - Pretesting: subject, staff, investigators, statistician
  - Mapping between different versions over time

Planning for Data Management

- Data to be collected: What? Why?
- Methods of collection: Who? Where? When? How?
  - Forms development
- Methods for data storage
  - Development of database
    - Administrative data: often dynamic
    - Scientific data: usually static
- Methods for data entry
  - Distributed versus central

Data Management

Handling of Data

- Collection
- Data entry
- Storage of forms, primary records
- Data verification
- Checking for errors
- Data reporting
- Data analysis
- Final database

Data Entry and Storage

- Data Entry
  - Transcription of data from forms into computerized data base
  - Personnel often clerical staff
    - Little scientific knowledge
  - Minimize data entry errors
    - Screen for impossible values
    - Screen for inconsistencies within form
    - Double entry
- Storage of forms, primary records
  - Subject confidentiality is a major concern
  - Must ensure limited access to confidential information

Data Verification and Error Checking

- Data entry errors
- Data collection errors
  - Audit clinics
  - Compare study data to medical records
- Maintaining an audit trail
  - Changing database versus making corrections in separate files

Data Reporting

- Administrative analyses
  - Accrual rates
  - Timeliness of data collection
  - Completeness of data collection
- Baseline characteristics
- Event rates (combined treatment groups only)

Data Analysis

- The ultimate purpose of collecting the data
- MUCH easier, more generalizable if all the previous stages conducted properly
- Complete record of all analyses should be maintained
  - date of analysis
  - version of data base and software

Introduction to Clinical Trials - Day 2

Session 8 - Documentation for a Clinical Trial

Presented July 24, 2018

Susanne J. May
Department of Biostatistics
University of Washington

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

Documenting the study

Motivation, need, and processes

- Problem:
  - Trial design is pre-specified in order to assure a carefully designed experiment
  - Changes will be necessary during trial implementation:
    - Unanticipated design elements (hopefully minimal)
    - Results on safety or tertiary endpoints that are discovered at interim analyses
    - New results from other trials of similar agents
    - Changes in study-related procedures
  - These changes must be implemented in a manner that maintains the integrity of the original design:
    - Avoid data-driven changes to the design
    - Pre-specify the process
    - Provide framework for documentation

Documenting the study

Key elements of trial oversight and documentation

- Key elements:
  - Trial oversight
    - Trial steering committee
    - Institutional Review Boards (IRB's)
    - FDA
    - Trial sponsor (NIH or pharmaceutical company)
  - Trial documentation
    - Trial protocol: complete documentation of the experiment: $\approx$ 80 pages
    - Statistical analysis plan (SAP): Complete pre-specification of all statistical analysis: $\approx$ 25 pages (plus tables)
    - Interim statistical analysis plan (ISAP): Complete documentation of the interim analysis plan: $\approx$ 20 pages
    - ClinicalTrials.gov: central repository for all trials
  - DSMB documents:
    - DSMB charter
    - DSMB open-report template
    - DSMB closed-report template

Documenting the study

Trial Protocol

- Purpose:

  Complete documentation to assure reproducibility

- Key elements:
  - Background
  - Objectives
  - Study design
  - Materials and methods
  - Human subjects

- Note: the protocol is supplemented by the manual of procedures (MOP):
  - Documentation of specific trial procedures (e.g., measurement methods)
  - Documents refinements to procedures (changes or details that are specified in the midst of a trial)
  - Documents nuance of eligibility/exclusions
  - MOP is updated as needed (incorporating mid-trial refinements)

Documenting the study

Statistical Analysis Plan

- Purpose:
  - Prespecification of all analyses
  - Prespecification of interpretation of multiple analyses (how will results be synthesized to answer trial questions)
- Key elements:
  - Summarize design (from protocol)
  - Preliminary data checking process
  - Primary analysis
  - Secondary analyses
  - Tertiary/exploratory analyses
  - Data-driven (post-hoc) analyses (keep a running record)
  - Draft shells for result tables

Documenting the study

Interim Statistical Analysis Plan

- Purpose: prespecify interim decision plans (related to trial outcomes)

- Key elements:
  - Summarize trial design and SAP
  - Define endpoint(s) for interim analyses
  - Specify interim decision criteria
  - Evaluate properties of interim decision criteria (power, ASN, inference at boundary, etc)
  - Specify process for implementing the monitoring plan:
    - Error-spending vs constrained boundary approaches
    - How revised decision rules are calculated:
      Boundary shape function
      Linear interpolation
    - Method for bias-adjusted inference upon completion
      BAM, RB-adjusted, MUE,
      analysis time ordering, sample mean ordering

Documenting the study

Key resources

- ICH guidelines (www.ich.org):
  - Part E8: General Considerations
  - Part E9: Statistical Principals
  - Part E10: Choice of Control Group
- CONSORT Statement (www.consort-statement.org):
  - Standards for reporting results (25 parts):
    - Title
    - Introduction
    - Methods
    - Results
    - Discussion
    - Other information