Clinical Trials in Rare Diseases: Challenges in Design, Analysis, and Interpretation

Michael P. McDermott, Ph.D.
University of Rochester Medical Center

December 6, 2013

Overview

- Clinical trials in rare diseases present several challenges
  - Such trials are more prone to variability and may have power to detect only large treatment effects
  - Importance of study planning is magnified and planning requires more time
  - Critically important to forge collaboration between clinicians and statisticians

Institute of Medicine Report (2001)

- Several recommendations pertaining to the design, analysis, and interpretation of clinical trials that, for reasons that are unavoidable, are constrained to be small

http://www.nap.edu/catalog/10078.html

Institute of Medicine Report (2001)

- Recommendations
  - Define the research question
    - Need to help clinicians make therapeutic decisions
  - Tailor the design
  - Clarify methods of reporting of results in clinical trials
    - Research synthesis; clinical context
  - Perform corroborative statistical analyses
    - Uncertainty regarding analysis assumptions
  - Exercise caution in interpretation
    - Extrapolation of study results
  - More research on alternative designs is needed

Clinical Trials in Rare Diseases

- Limited availability of resources
  - Willing trial participants
  - Funding sources
- In this setting, feasibility constraints can lead to compromises in important principles of sound trial design

Some Important Principles of Sound Trial Design

- Precise formulation of a focused research question
  - Prioritization of outcome variables and analyses
- Tailoring of study design to best answer the research question posed
  - Minimization of bias
    - Randomization
    - Blinding
    - Appropriate control group
      - Context of existing treatment
      - Use of placebo/sham treatment

Some Important Principles of Sound Trial Design

- Tailoring of study design to best answer the research question posed
  - Appropriate eligibility criteria
    - Generalizability vs. efficiency
  - Appropriate outcome measures
    - Reliable, valid, responsive, applicable
    - Duration of follow-up
  - Appropriate and feasible sample size
  - Appropriate measures for participant recruitment and retention
    - Frequency and timing of assessments
    - KISS principle

Outcome Variables

- Continuous
  - Tend to be more responsive
  - Meaningful?
  - Normally distributed?
  - Replicate measures can increase precision
- Time-to-event
  - Example: Disease milestone
- Categorical
  - Tend to be less responsive

Outcome Variables

- Use of longitudinal data
  - Change from baseline to final visit
  - Use of data from all visits
    - Area under the response-time curve
    - Average of responses after a certain time point
    - Slope (rate of change)
  - Choice may depend on expected timing of onset/loss of maximal benefit
  - Choice also depends on the clinical question that is most relevant to address

Issues that Small Trials Are Better Equipped to Address

- Pharmacokinetics
  - Single- and multiple-dose studies
- Maximum tolerated dosage
  - άo Ь oέ ₴ℕJ Ŏ₴yűλyȡzI·f ℕJŜI·ăăŜăăϒŜyű ŘŜăλƎyã ó₴ℕJ ØI·ℕJλI·yűăú
- Short-term safety
- ά! ŎűűλØλűüѐέ ₴₮ űĸKŜ űℕJŜI·úϒŜyű
  - bŜŜŘ ₮₴ℕJ I·LJLJNJ₴LJNJλI·úŜ ϒI·ℕJ⅂ŜNJă ₴₮ άI·ŎűűλØλűüѐέ
- Preliminary efficacy or futility
  - Acceptance of higher error rates (false positive, false negative)
- Selection of a treatment

Selection Designs

- Goal is to select, out of potential treatments, the one with the best response
  - Randomized, parallel-group trial
  - Requires a much smaller sample size than a trial designed to formally test the null hypothesis of no treatment effects
    -

Use of Historical Controls

- Advantages
  - Approximately four-fold fewer subjects required compared to a two-arm trial with a concurrent control group
  - Recruitment
- Dangers of using historical controls
  - Changes in ancillary care over time
  - Differences in rater behavior
  - Differences in entry criteria
  - Differences in recruitment of subjects
  - Lack of blinding
  - CMT-1a example

Consequences of the Use of Invalid Historical Controls

- Biases tend to favor treatment under study
- Ability to conduct subsequent confirmatory trials can be compromised
  - ά t ž ă ł ú ł Ø Ŝ ε NJ Ŝ ă d z f ú ă ₹ N J ž Y ´ ú N J ł · f ă ú K · ú Ř ž ý ž ű dz ă Ŝ · randomized, concurrent control group
- Treatment can be worse than placebo
  - Recent examples: minocycline and lithium in ALS
- A rare disease is no excuse for a poorly designed study

Cross-Over Designs

- A cross-over trial is one in which subjects are given different treatments during different treatment periods, with the object of comparing the various treatments
- Treatments are given in a randomly determined sequence (e.g., A/B vs. B/A)

Two-Period Cross-Over Design

Sequence	Period 1	Washout	Period 2
A/B	A	———	B
B/A	B	———	A

Cross-Over Designs

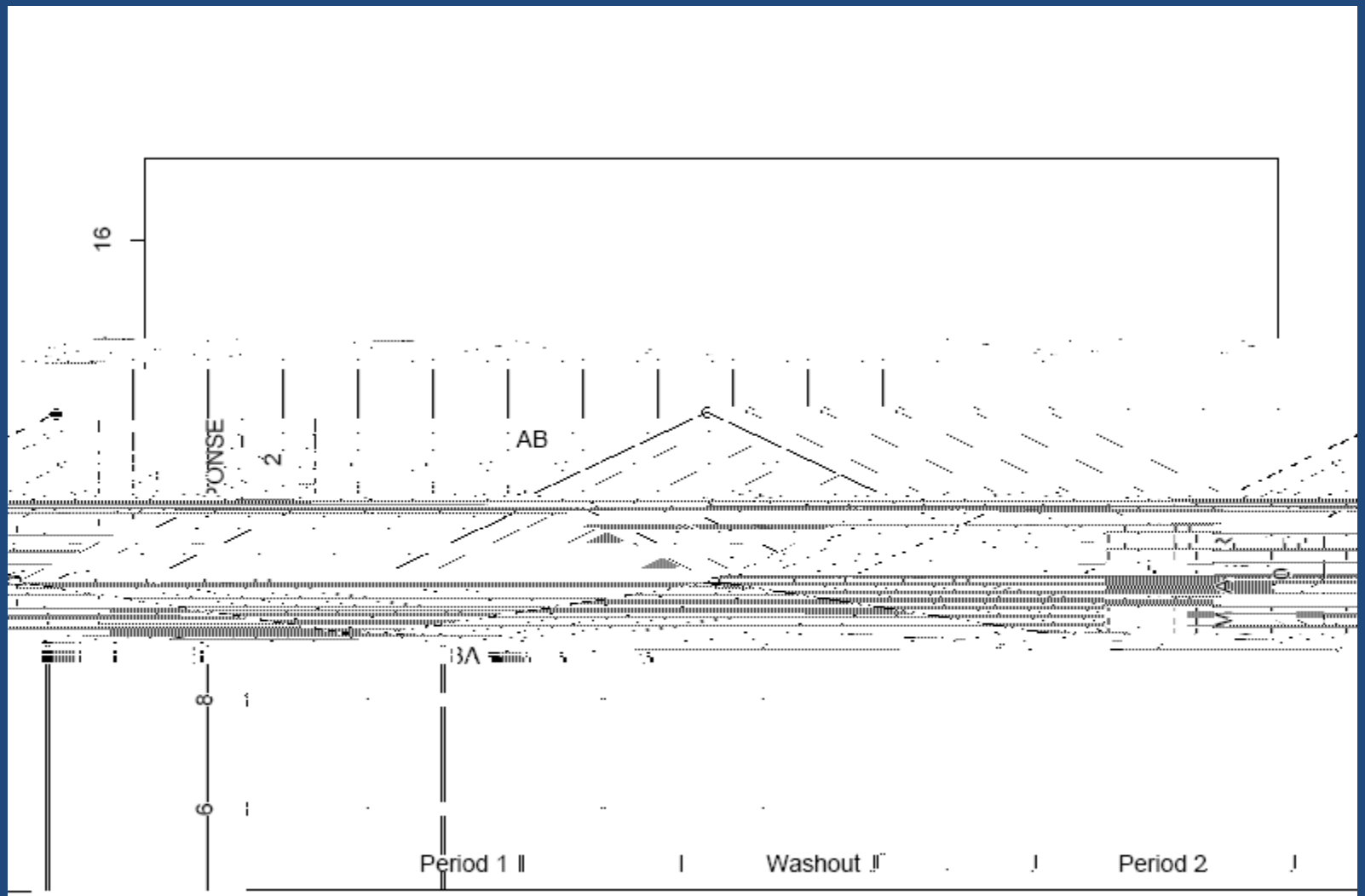
- Appropriate for treatments that may offer short-term relief of signs or symptoms, not a cure for the condition
  - Asthma, hypertension, epilepsy, pain, other chronic conditions
- It is assumed that the symptom or condition will return ahETc4[c60non)33 Tm[ )]TJi‡

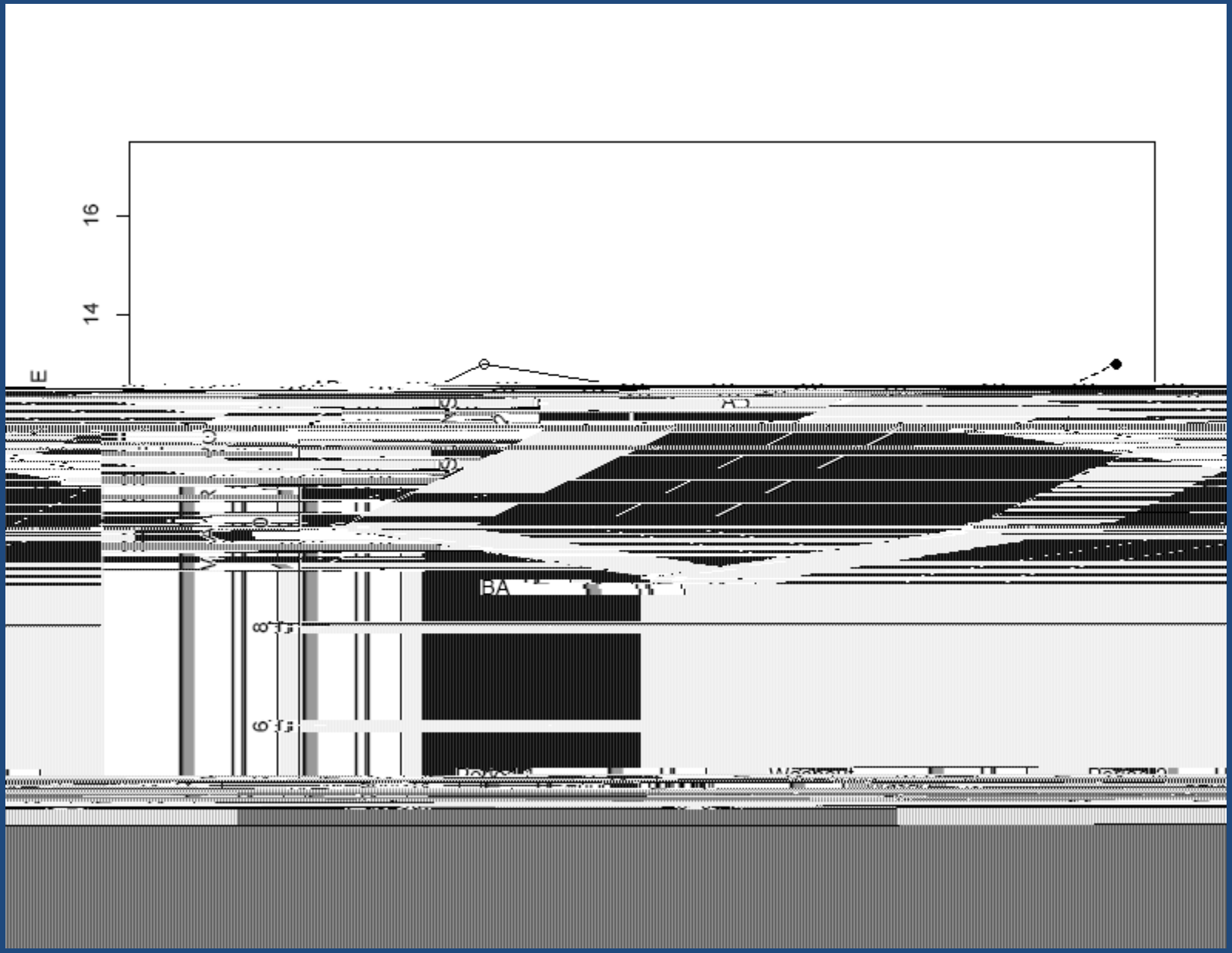
Advantages of Cross-Over Designs

- Profound savings in sample size
  - Within-subject comparisons
- Participants gain access to all treatments under study
  - May enhance recruitment/retention

Disadvantages of Cross-Over Designs

- Not suitable for all conditions
  - JNCL?
- Impact of subject withdrawal
- Importance of blinding is magnified
- Inconvenience to participants
  - Multiple treatment/washout periods
  - Total duration of follow-up


16
AB
Period 1
Washout
Period 2


16
14
E
BA

N-of-1 Trials

- Performed in multiple pairs of treatment/placebo periods
  - Example: AB BA BA AB . . .
  - Feasibility of multiple treatment periods
  - Same limitations as those for cross-over trials discussed earlier
- Require rapid onset/washout of the treatment and its effects
- Inference for individual patients is limited without having many periods
- A series of N-of-1 trials in different patients can be much more powerful
  - Random effects models can be used to combine information across patients

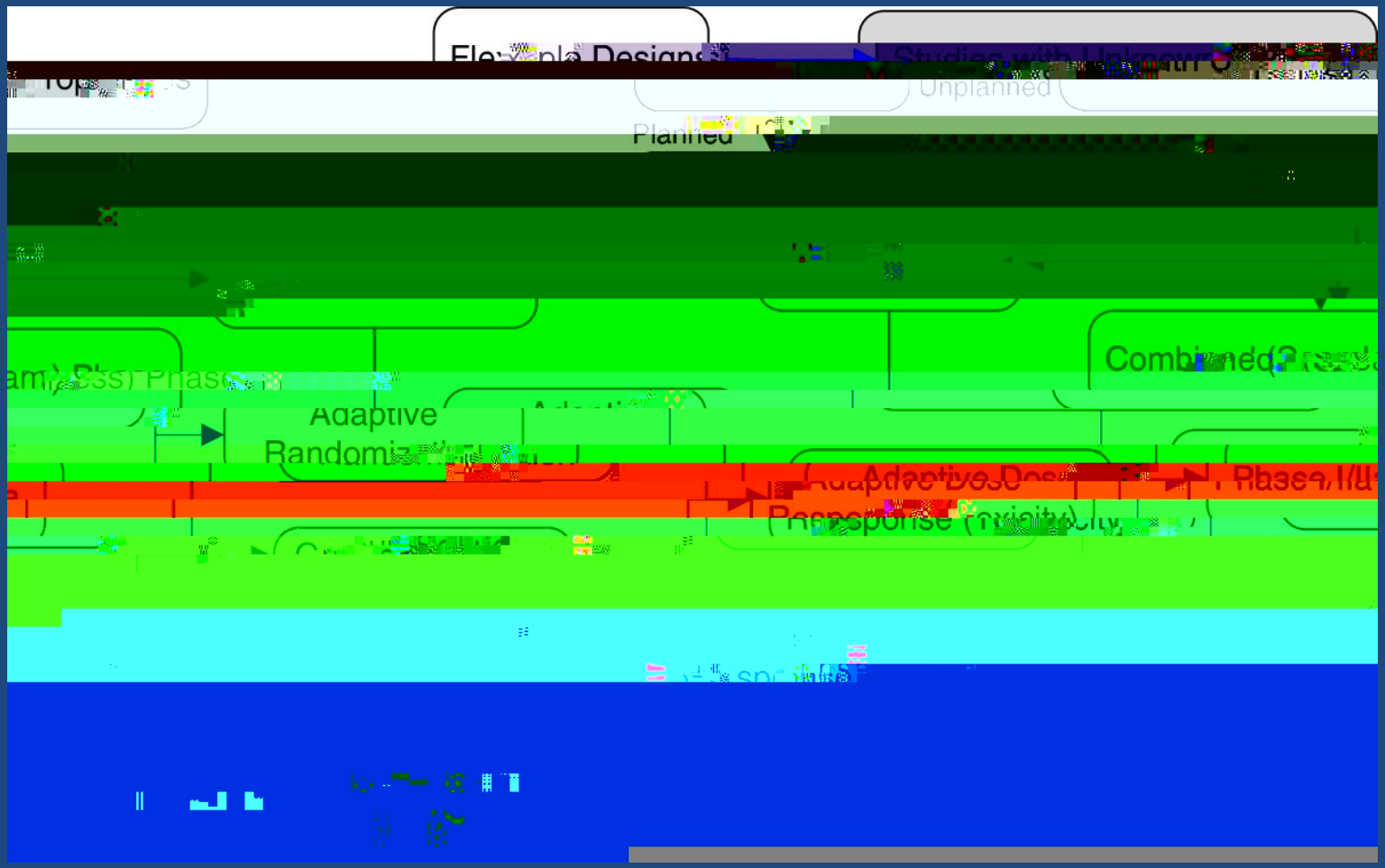
Adaptive Designs

- Pharmaceutical Research and Manufacturers of America (PhRMA) Working Group (2006):
  - ά.ѐ , we refer to a clinical study design that uses accumulating data to decide how to modify aspects of the study as it continues, without undermining the validity Ι·ƴŘ ƛƴǘŜƎǊƛǘѐ ƻŦ ǘƘŜ ǘǊƛŀƭ.ϵέ
  - άLƴ ŀŘŘƛǘƛƻƴ ǘƘŀƛ ŎƘŀƴƎŜǎ ŀǊŜ ƴƻǘ ƳŀŘŜ ƻƴ ŀƴ ad hoc basis; therefore, adaptation is a aimed to enhance the trial, not a remedy for inadequate ǇƭŀƴƴƛƴƎ.ϵέ

Adaptive Designs

- Validity
  - Correct statistical inference
    - Control of Type I and Type II errors
    - Minimization of bias
  - Consistency between stages of the trial
  - Low operational bias
- Integrity
  - Results are convincing to a broader scientific community
  - Pre-planned adaptations
  - Maintenance of the blind to interim analysis results

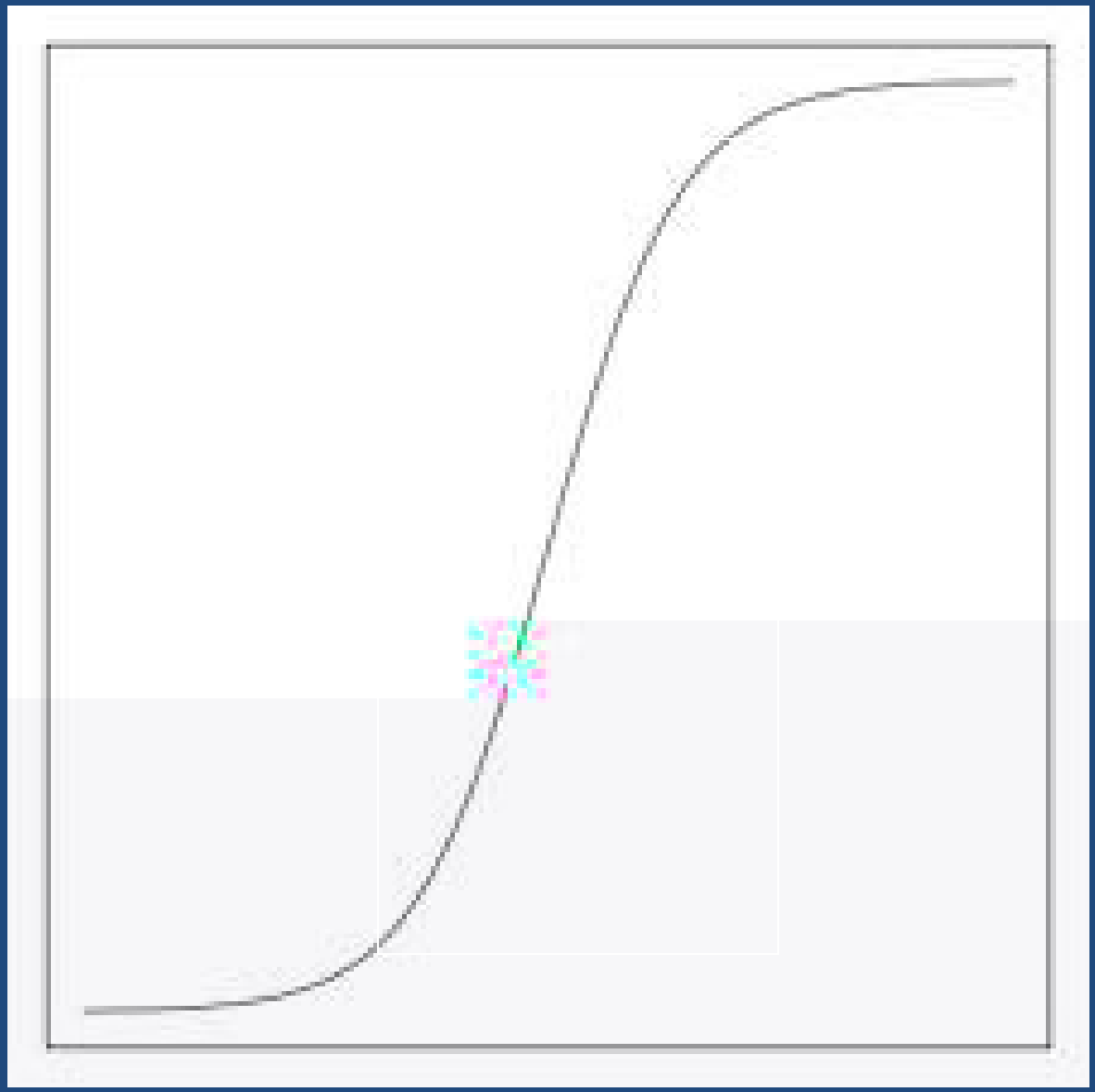
Some Types of Adaptive Designs

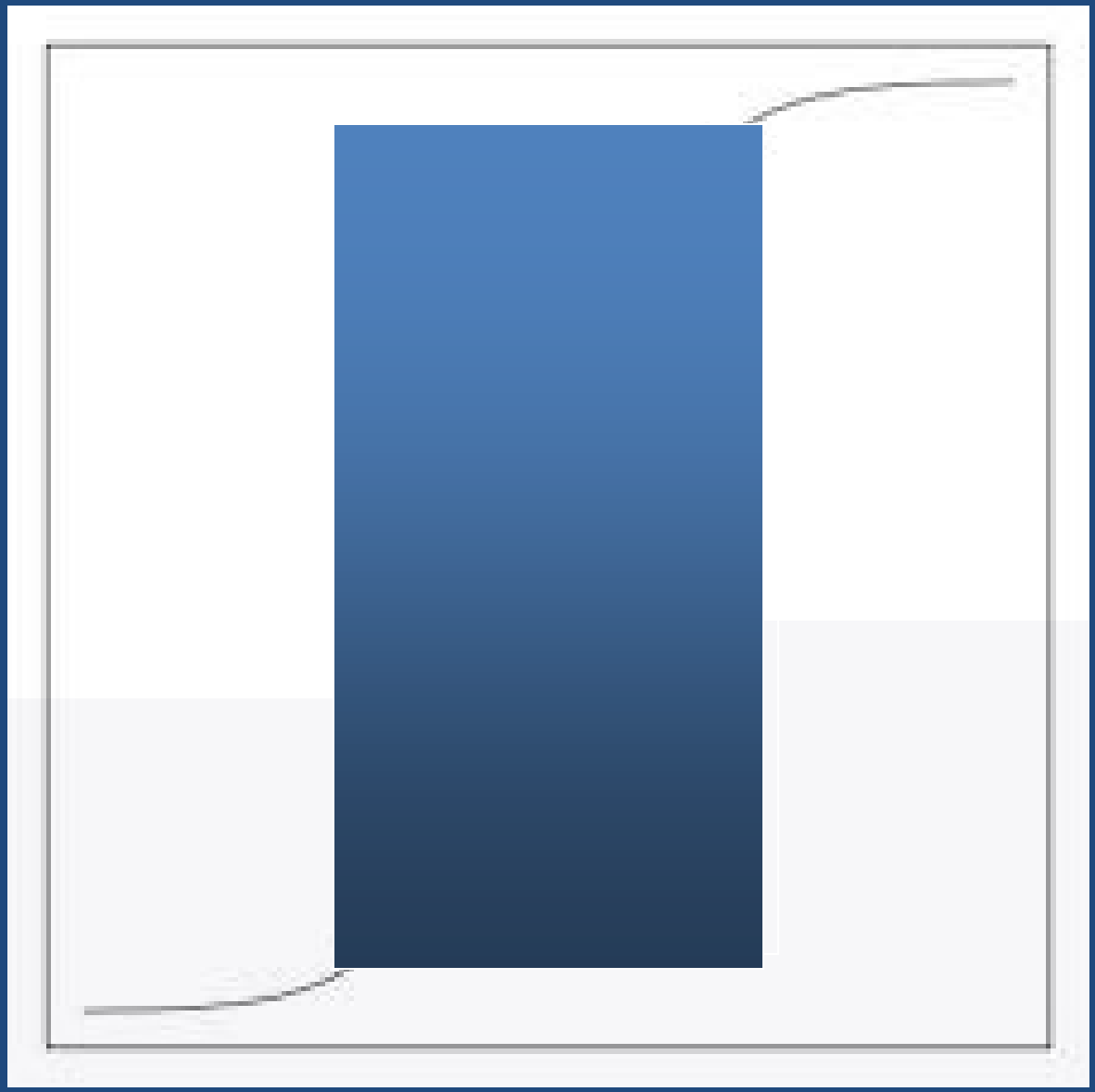


Kairalla et al. Trials 2012; 13:145

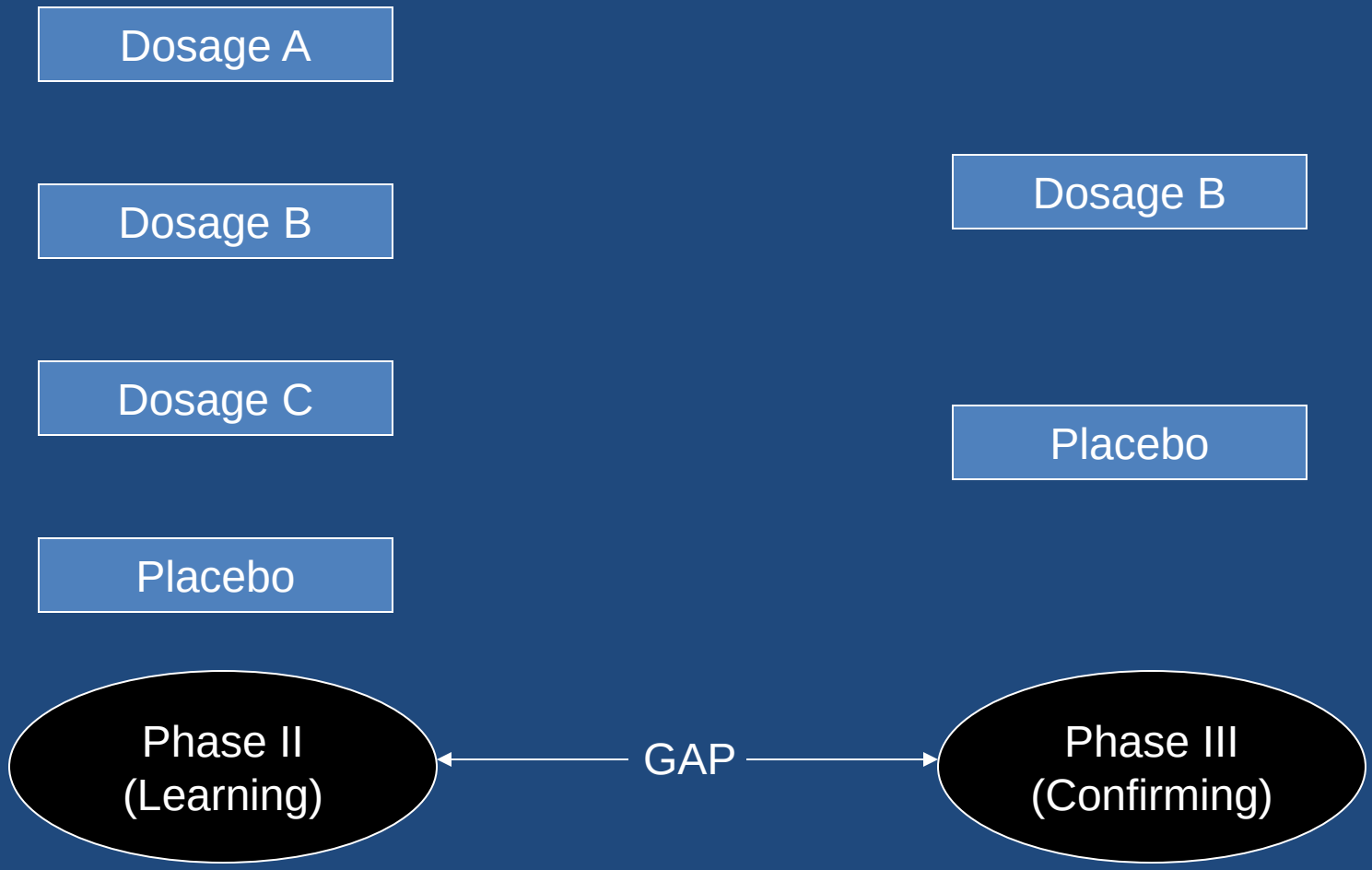
Adaptive Dose Finding

- Traditional approach in Phase II
  - Randomization to a relatively small number of fixed dosages (3-4) and placebo
  - Disadvantages
    - [I·ℕ∃Ŝ άŘƛăűI·yŎŜέ ōŜűøŜŜy I·Řƨ I·ŎŜyű Ř₴ăI·∃Ŝă
      - Optimal dosage may not be studied
    - Some of the studied dosages may not be useful
      - This may become apparent relatively quickly
    - Accumulating evidence may suggest early stopping for futility or identification of a sufficient dosage to study further

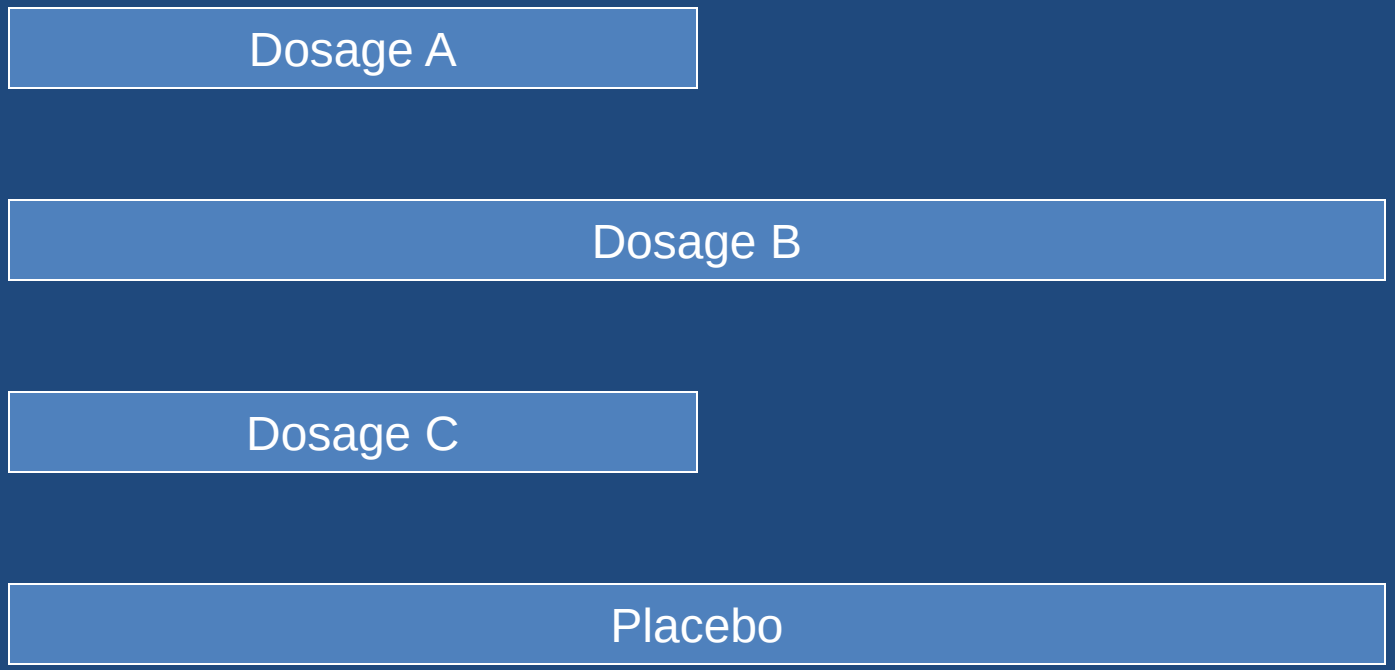

Response
Dosage


Response
Dosage

Seamless Phase II/III Designs

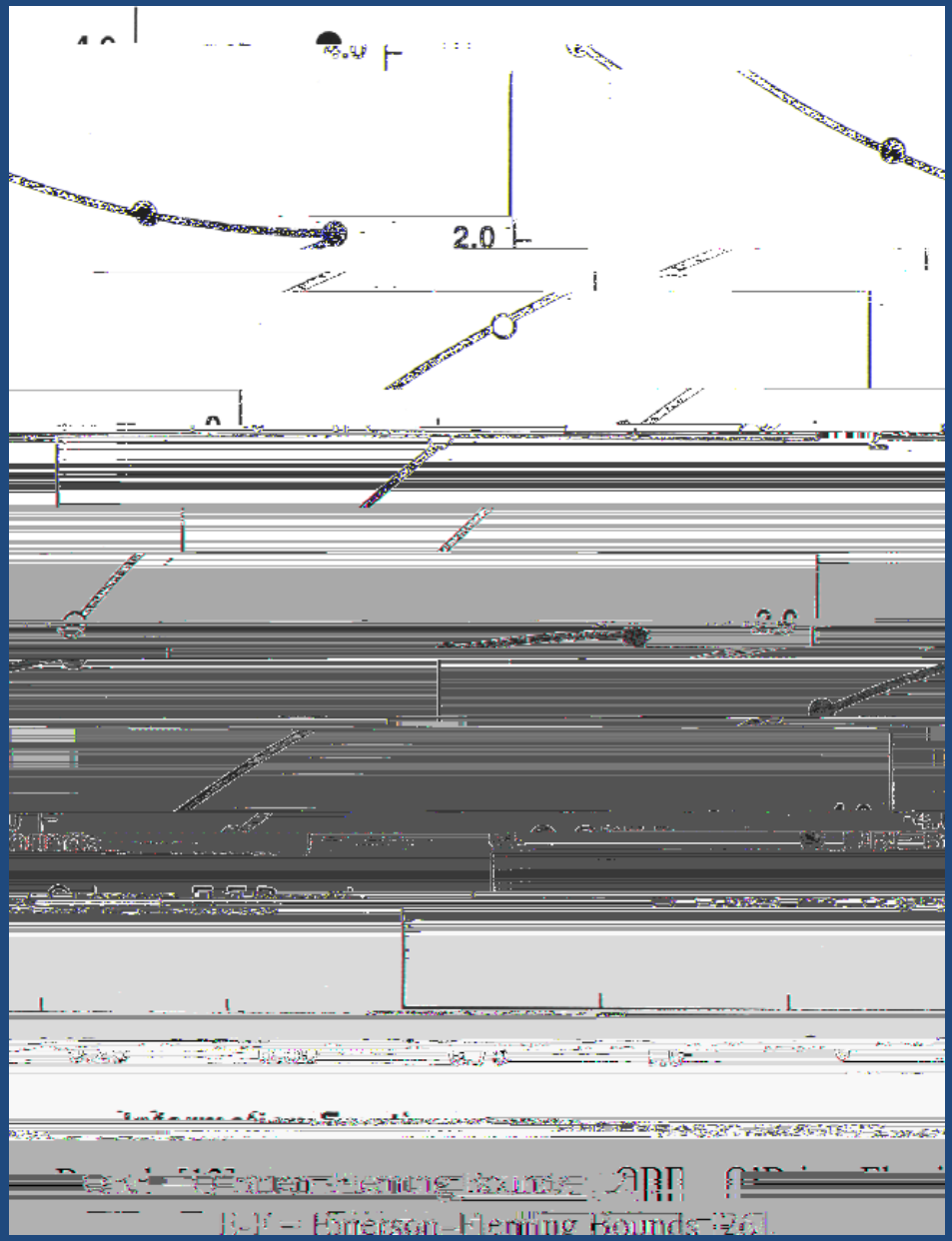


Seamless Phase II/III Designs



Group Sequential Designs

- Interim analyses of accumulating data
  - Ethical issues
  - Efficiency/cost
  - Consideration of safety, efficacy, and futility
  - Problem of repeated significance testing
    - LyǑNJŜI·ăŜŘ LJNJ2ōI·ōλfλűề 2Ŧ άŦI·făŜ-LJ2ăλűλØŜέ óŜŦŦλǑI·Ǒèú
    - LyǑNJŜI·ăŜŘ LJNJ2ōI·ōλfλűề 2Ŧ άŦI·făŜ-yŶŜƎI·űλØŜέ óŦdzűλfλűèú
    - - and -spending functions
      - ¦ ăŜŘ ű2 ŘŜŦλyŶŜ άăű2LJLJλyƎ ō2dzyŘI·NJλŜăέ



DeMets. Clin Trials 2006; 3:522-529

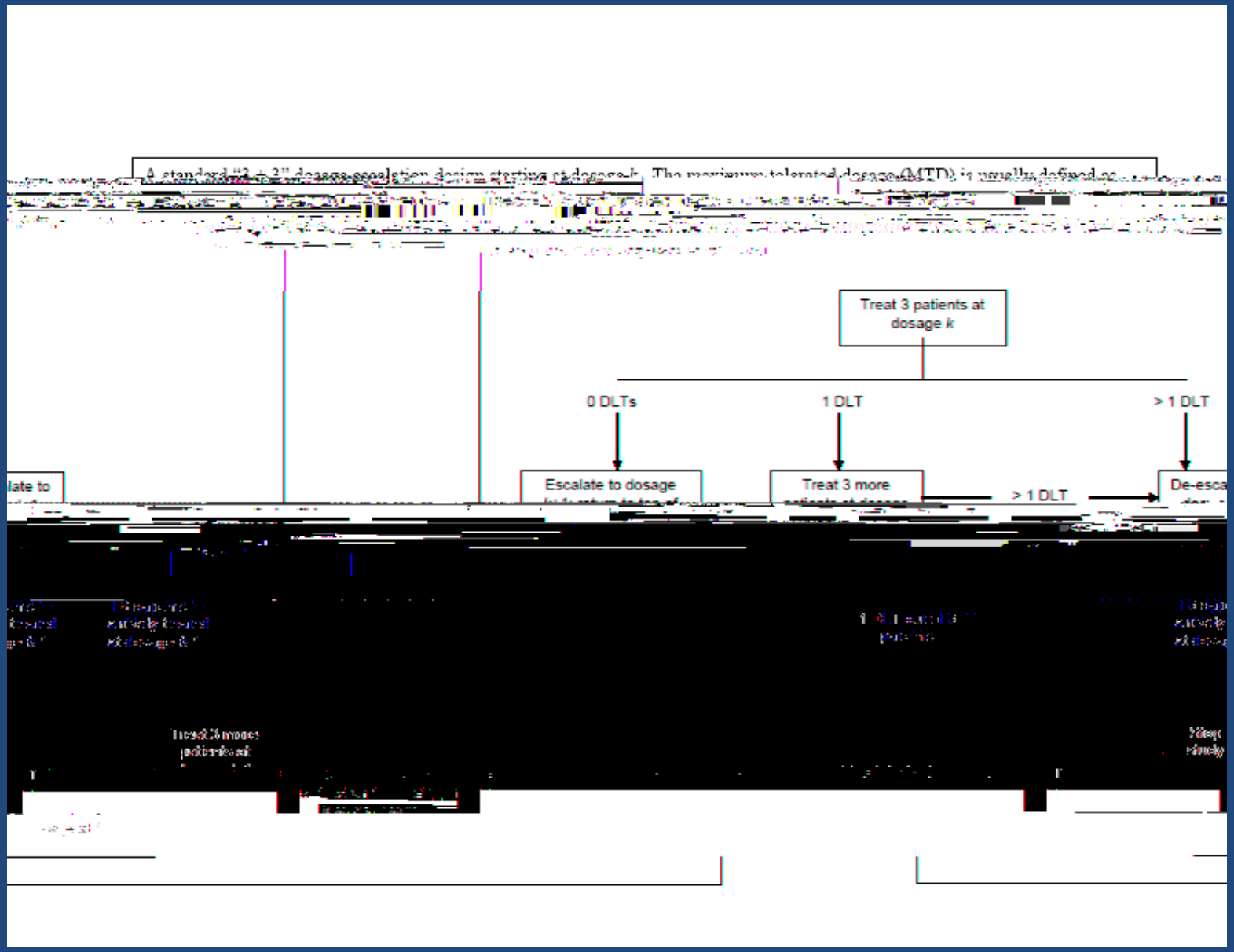
Adaptive Designs

- There are many logistical and procedural issues that are introduced by the possibility of adaptation
  - Careful planning; evaluation of feasibility; infrastructure
- Trial integrity should be preserved by minimizing access to information on interim analyses and their results
  - Control of operational bias

Small is Not Big

- In a very rare disease, sacrifices in important areas may have to be considered
  - Early/middle development
  - Error rates (significance level; power)iddle devy74(dl)4

EXTRA SLIDES


Treat 3 patients at dosage k
0 DLTs
1 DLT
> 1 DLT
Escalate to dosage
Treat 3 more
> 1 DLT

Problems with Many Preliminary Studies

- Often, preliminary studies, particularly in rare diseases:
  - Are very small
  - Are uncontrolled
  - Do not address a focused question
  - Do little to enhance decision-making for further study of the intervention and, as a consequence, slow research progress

Potential Adverse Consequences of Small Trials

- Discarding of potentially effective treatments due to lack of statistically significant benefits
  - ά bŜ∃l·űƛǾŜέ Ǿăɸ άĿƴŎƻƴŎŦƌăƛǾŜέ ăűƌŘƛŜă
  - P-values vs. confidence intervals
- Inappropriate emphasis on informally defined όƻǋ ƌƴŘŜŦƛƴŜŘύ άűǋŜƴŘă ƛƴ űKŜ ǋƛ∃Kű ŘƛǋŜŎűƛƻƴέ (or lack thereof)
- Illusion of safety

Role of Confidence Intervals in Trial Interpretation

% Difference in Rate of Progression	95% Confidence Interval	P-value	Evidence for Treatment Effect
30%	(-20%, 80%)	0.30	Inconclusive
30%	(20%, 40%)	0.003	Positive
2%	(-4%, 8%)	0.30	Negative
2%	(1%, 3%)	0.003	Positive, but not clinically important
2%	(-30%, 34%)	0.93	???

Role of Confidence Intervals in Trial Interpretation

% Difference in Rate of Progression	95% Confidence Interval	P-value	Evidence for Treatment Effect
30%	(-20%, 80%)	0.30	Inconclusive
30%	(20%, 40%)	0.003	Positive
2%	(-4%, 8%)	0.30	Negative
2%	(1%, 3%)	0.003	Positive, but not clinically important
2%	(-30%, 34%)	0.93	Inconclusive