Adaptive Clinical Trials

Short Course Presenters:
- Michael Krams, M.D., Wyeth Research
- Vladimir Dragalin, Ph.D., Wyeth Research
- Jeff Maca, Ph.D., Novartis Pharmaceuticals
- Keaven Anderson, Ph.D., Merck Research Laboratories
- Paul Gallo, Ph.D., Novartis Pharmaceuticals

Short Course Outline

<table>
<thead>
<tr>
<th>Topics</th>
<th>Time</th>
<th>Presenter</th>
</tr>
</thead>
<tbody>
<tr>
<td>General Introduction</td>
<td>10 min</td>
<td>Mike Krams</td>
</tr>
<tr>
<td>Terminology and Classification</td>
<td>40 min</td>
<td>Vlad Dragalin</td>
</tr>
<tr>
<td>Adaptive Seamless Design</td>
<td>40 min</td>
<td>Jeff Maca</td>
</tr>
</tbody>
</table>

Break 20 min

<table>
<thead>
<tr>
<th>Topics</th>
<th>Time</th>
<th>Presenter</th>
</tr>
</thead>
<tbody>
<tr>
<td>Sample Size Reestimation and Group Sequential Design</td>
<td>40 min</td>
<td>Keaven Anderson</td>
</tr>
<tr>
<td>Logistic, Operational and Regulatory Issues</td>
<td>40 min</td>
<td>Paul Gallo</td>
</tr>
<tr>
<td>General Discussion</td>
<td>20 min</td>
<td></td>
</tr>
</tbody>
</table>
Adaptive Designs
General Introduction

Michael Krams

Wyeth Research

On behalf of the PhRMA Adaptive Designs Working Group

PhRMA Adaptive Designs Working Group

- Co-Chairs:
  Michael Krams
  Brenda Gaydos

- Authors:
  Keaven Anderson
  Suman Bhattacharya
  Alun Bedding
  Don Berry
  Frank Bretz
  Christy Chuang-Stein
  Vlad Dragalin
  Paul Gallo
  Brenda Gaydos
  Michael Krams
  Qing Liu
  Jeff Maca
  Inna Perevozskaya
  Jose Pinheiro
  Judith Quinlan

- Members:
  Carl-Fredrik Burman
  David DeBrota
  Jonathan Denne
  Greg Enas
  Richard Entsuah
  Andy Grieve
  David Henry
  Tony Ho
  Telba Irony
  Larry Lesko
  Gary Littman
  Cyrus Mehta
  Allan Pallay
  Michael Poole
  Rick Sax
  Jerry Schindler
  Michael D Smith
  Marc Walton
  Sue-Jane Wang
  Gernot Wassmer
  Pauline Williams
Vision

- To establish a dialogue between statisticians, clinicians, regulators and other lines within the Pharmaceutical Industry, Health Authorities and Academia,
- with a goal to contribute to developing a consensus position on when and how to consider the use of adaptive designs in clinical drug development.

Mission

- To facilitate the implementation adaptive designs, but only where appropriate
- To contribute to standardizing the terminology and classification in the rapidly evolving field of adaptive designs
- To contribute to educational and information sharing efforts on adaptive designs
- To interact with experts within Health Authorities (FDA, EMEA, and others) and Academia to sharpen our thinking on defining the scope of adaptive designs
- To support our colleagues in health authorities in their work towards the formulation of regulatory draft guidance documents on the topic of adaptive designs.
Executive Summary of White Paper


ADAPTIVE DESIGNS IN CLINICAL DRUG
DEVELOPMENT—AN EXECUTIVE SUMMARY
OF THE PHARMA WORKING GROUP

Paul Gallo
Novartis Pharmaceuticals, East Hanover, New Jersey, USA

Christy Chuang-Stein
Pfizer Inc., Ann Arbor, Michigan, USA

Vladimir Dragalin
GlaxoSmithKline Pharmaceuticals, Collegeville, Pennsylvania, USA

Brenda Gaydos
Eli Lilly and Company, Indianapolis, Indiana, USA

Michael Krams
Pfizer Inc., Groton, Connecticut, USA

José Pinheiro
Novartis Pharmaceuticals, East Hanover, New Jersey, USA


ADAPTIVE DESIGN
PhRMA Working Group on Adaptive Designs Full
White Paper • PAUL GALLEO, MICHAEL KRAMS • 421

Adaptive Designs: Terminology and Classification • VLADIMIR DRAGALIN • 425

Implementing Adaptive Designs: Logical and
Operational Considerations • JUDITH A. QUINLAN,
MICHAEL KRAMS • 437

CONTINUING EDUCATION
Confidentiality and Trial Integrity Issues for
Adaptive Designs • PAUL GALLEO • 445

Adaptive Dose-Response Studies • BRENDA GAYDOS,
MICHAEL KRAMS, INGA PEREROZKAYA, FRANK BRETZ, CHENG
LII, PAUL GALLEO, DON BERRY, CHRISTY CHUANG-STEIN,
JOSE PINHEIRO, ALVIN BEEDING • 451

Adaptive Seamless Phase II/III Designs—
Background, Operational Aspects, and Examples •
JENNIFER S. BISHOP, SHIVA BHATTACHARYAYA, VLADIMIR DRAGALIN,
PAUL GALLEO, MICHAEL KRAMS • 463

Sample Size Reestimation: A Review and
Recommendations • CHRISTY CHUANG-STEIN, KEVEN
ANDERSON, PAUL GALLEO, SHIVA COLLINS • 475

PhRMA Adaptive Designs Working Group
PhRMA/FDA conference on Adaptive Design:
Opportunities, Challenges and Scope in Drug Development

- Nov 13/14th, 2006
  - Marriott Bethesda North Hotel & Conference Center
    North Bethesda, MD 20852

- Program committee

<p>| | | | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Dennis Erb</td>
<td>Merck</td>
<td>Greg Campbell</td>
<td>FDA CDRH</td>
</tr>
<tr>
<td>Brenda Gaydos</td>
<td>Eli Lilly</td>
<td>Shirley Murphy</td>
<td>FDA CDER</td>
</tr>
<tr>
<td>Michael Krams</td>
<td>Wyeth, Co-Chair</td>
<td>Robert O’Neill</td>
<td>FDA CDER</td>
</tr>
<tr>
<td>Walt Offen</td>
<td>Eli Lilly, Co-Chair</td>
<td>Robert Powell</td>
<td>FDA CDER</td>
</tr>
<tr>
<td>Frank Shen</td>
<td>BMS</td>
<td>Marc Walton</td>
<td>FDA CDER</td>
</tr>
<tr>
<td>Luc Truyen</td>
<td>J&amp;J</td>
<td>Sue Jane Wang</td>
<td>FDA CDER</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Adaptive Designs  
Terminology and Classification

Vlad Dragalin
Biomedical Data Sciences
GlaxoSmithKline
(currently with Wyeth Research)

Primary PhRMA references

- PhRMA White Paper section:
Outline

- Definition and general structure of adaptive designs
- Classification of adaptive designs in drug development
- Achieving the goals

What are Adaptive Designs?

- Flexible
- Multi-stage
- Dynamic
- Sequential
- Response-driven
- Self-designing
- Novel

- An adaptive design should be adaptive by "design" not an adhoc change of the trial conduct and analysis
- Adaptation is a prospective design feature, not a remedy for poor planning
What are Adaptive Designs?

**Adaptive Plan**

... not Adaptive Plane

---

**Definition**

**Adaptive Design**

- uses accumulating data to decide on how to modify aspects of the study

- without undermining the *validity* and *integrity* of the trial

---

*Validity* means
- providing correct statistical inference (such as adjusted p-values, unbiased estimates and adjusted confidence intervals, etc)
- assuring consistency between different stages of the study
- minimizing operational bias

*Integrity* means
- providing convincing results to a broader scientific community
- preplanning, as much as possible, based on intended adaptations
- maintaining confidentiality of data
General Structure

- An adaptive design requires the trial to be conducted in several stages with access to the accumulated data.
- An adaptive design may have one or more rules:
  - **Allocation Rule**: how subjects will be allocated to available arms
  - **Sampling Rule**: how many subjects will be sampled at next stage
  - **Stopping Rule**: when to stop the trial (for efficacy, harm, futility)
  - **Decision Rule**: the terminal decision rule and interim decisions pertaining to design change not covered by the previous three rules

- At any stage, the data may be analyzed and next stages redesigned taking into account all available data.

Examples

- Group Sequential Designs: only **Stopping Rule**
- Response Adaptive Allocation: only **Allocation Rule**
- Sample Size Re-assessment: only **Sampling Rule**
- Flexible Designs:
  - Adaptive AR: changing the randomization ratio
  - Adaptive SaR: the timing of the next IA
  - Stopping Rule
  - Adaptive DR: changing the target treatment difference; changing the primary endpoint; varying the form of the primary analysis; modifying the patient population; etc
Allocation Rules

- Fixed (static) AR:
  - Randomization used to achieve balance in all prognostic factors at baseline
  - Complete randomization uses equal allocation probabilities
  - Stratification improves the randomization

- Adaptive (dynamic) AR:
  - Response-adaptive randomization uses interim data to unbalance the allocation probabilities in favor of the “better” treatment(s): urn models, RPW, doubly adaptive biased coin design
  - Bayesian AR alters the allocation probabilities based on posterior probabilities of each treatment arm being the “best”

Sampling Rules

- Sample size re-estimation (SSR)
  - Restricted sampling rule
  - Blinded SSR or Unblinded SSR based on estimate of nuisance parameter

- Traditional Group Sequential Designs
  - Fixed sample sizes per stage

- Error Spending Approach
  - Variable sample sizes per stage (but do not depend on observations)

- Sequentially Planned Decision Procedures
  - Future stage sample size depends on the current value of test statistic

- Flexible SSR uses also the estimated treatment effect
Stoopping Rules

- Early Stopping based on Boundary Crossing
  - Superiority
  - Harm
  - Futility
- Stochastic Curtailment
  - Conditional power
  - Predictive power
- Bayesian Stopping Rules
  - Based on posterior probabilities of hypotheses
  - Complemented by making predictions of the possible consequences of continuing

Decision Rules

- Changing the test statistics
  - Adaptive scores in trend test or under non proportional hazards
  - Adaptive weight in location-scale test
  - Including a covariate that shows variance reduction
- Redesigning multiple endpoints
  - Changing their pre-assigned hierarchical order in multiple testing
  - Updating their correlation in reverse multiplicity situation
- Switching from superiority to non-inferiority
- Changing the hierarchical order of hypotheses
- Changing the patient population
  - going forward either with the full population or with a pre-specified subpopulation
Classification

<table>
<thead>
<tr>
<th>Disease selection</th>
<th>Target Family selection</th>
<th>Compound Progression Stages</th>
<th>Lifecycle Management</th>
</tr>
</thead>
<tbody>
<tr>
<td>FTIM to Commit to PoC/Phase II</td>
<td>Phase II to Commit to Phase III</td>
<td>Phase III to launch</td>
<td></td>
</tr>
</tbody>
</table>

**SINGLE ARM TRIALS**
- Two-stage Designs
- Screening Designs

**TWO-ARM TRIALS**
- Group Sequential Designs
- Information Based Designs
- Adaptive GSD (Flexible Designs)

**MULTI-ARM TRIALS**
- Bayesian Designs
- Group Sequential Designs
- Flexible Designs

**DOSE-FINDING STUDIES**
- Dose-escalation designs
- Dose-finding designs (Flexible Designs)
- Adaptive Model-based Dose-finding

**SEAMLESS DESIGNS**
- Dose-escalation based on efficacy/toxicity
- Learning/Confirming in Phase II/III

---

**Two-Stage Designs**

- **Objective**: single-arm studies using short-term endpoints; hypothesis testing about some minimal acceptable probability of response
- Gehan design: early stopping for futility; sample size of the 2nd stage gives a specified precision for response rate
- Adaptive two-stage design: Banerjee&Tsiatis (2006)
- Bayesian designs: Thall&Simon (1994)
### Screening Designs

- **Objective**: adaptive design for the entire screening program
  - Minimize the shortest time to identify the “promising” compound
  - Subject to the given constraints on type I and type II risks for the entire screening program
    - type I risk = $Pr$(screening procedures stops with a FP compound)
    - type II risk = $Pr$(any of the rejected compounds is a FN compound)
- Two-stage design (Yao & Venkatraman, 1998)
- Adaptive screening designs (Stout and Hardwick, 2002)
- Bayesian screening designs (Berry, 2001)

---

### Classification

<table>
<thead>
<tr>
<th>Disease selection</th>
<th>Target to tractable hit to candidate</th>
<th>Candidate selection to FTIM</th>
<th>FTIM to Commit to PoC/Phase II</th>
<th>Phase II to Commit to Phase III</th>
<th>Phase III to launch</th>
<th>Lifecycle Management</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>SINGLE ARM TRIALS</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Two-stage Designs</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Screening Designs</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>TWO-ARM TRIALS</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Group Sequential Designs</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Information Based Designs</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Adaptive GSD (Flexible Designs)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>MULTI-ARM TRIALS</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Bayesian Designs</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Group Sequential Designs</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Flexible Designs</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>DOSE-FINDING STUDIES</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Dose-escalation designs</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Dose-finding designs (Flexible Designs)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Adaptive Model-based Dose-finding</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>SEAMLESS DESIGNS</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Dose-escalation based on efficacy/toxicity</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Learning/Confirming in Phase I/II</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

---

Adaptive Designs Working Group
Fully Sequential Designs

- **Objective:** testing two hypotheses with given significance level and power at the prespecified alternative

- **AR:** fixed randomization
- **SaR:** after each observation
- **StR:** boundary crossing (e.g. SPRT, repeated significance test, triangular test)
- **DR:** final decision - to accept or reject the null hypothesis

- **References:** Siegmund (1985); Jennison&Turnbull (2000)

Group Sequential Designs

- **Objective:** testing two hypotheses with given significance level and power at the specified alternative, prefixed maximum sample size

- **AR:** fixed randomization
- **SaR:** after a fixed number (a group) of observations,
  - or using error-spending function,
  - or using “Christmas-tree” adjustment
- **StR:** boundary crossing
  - Haybittle, Pocock, O’Brien-Fleming type
  - linear boundaries
  - error-spending families
  - conditional power, stochastic curtailment
- **DR:** final decision - to accept or reject the null hypothesis

- **References:** Jennison&Turnbull (2000); Whitehead (1997)
Information Based Designs

- **Objective**: testing two hypotheses with given significance level and power at the specified alternative, prefixed maximum information

  - **AR**: fixed randomization
  - **SaR**: after fixed increments of information
  - **StR**: boundary crossing as for Group Sequential Designs
  - **DR**: adjust maximum sample size based on interim information about nuisance parameters

- **References**: Mehta&Tsiatis (2001); East (2005)

Adaptive GSD (Flexible Designs)

- **Objective**: testing two hypotheses with given significance level and power at the specified alternative or adaptively changing the alternative at which a specified power is to be attained

  - **AR**: fixed or adaptive randomization
  - **SaR**: sample size of the next stage depends on results at the time of interim analysis
  - **StR**: p-value combination, conditional error, variance-spending
  - **DR**: adapting alternative hypothesis, primary endpoint, test statistics, inserting or skipping IAs

- **References**: Bauer; Brannath et al; Müller&Schäfer; Fisher
**Classification**

<table>
<thead>
<tr>
<th>Disease selection</th>
<th>Target Family selection</th>
<th>Candidate selection to FTIM</th>
<th>FTIM to Commit to PoC/Phase II</th>
<th>Phase II to Commit to Phase III</th>
<th>Phase III to launch</th>
<th>Lifecycle Management</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>SINGLE ARM TRIALS</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Two-stage Designs</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Screening Designs</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>TWO-ARM TRIALS</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Group Sequential Designs</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Information Based Designs</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Adaptive GSD (Flexible Designs)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>MULTI-ARM TRIALS</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Bayesian Designs</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Group Sequential Designs</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Flexible Designs</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>DOSE-FINDING STUDIES</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Dose-escalation designs</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Dose-finding designs (Flexible Designs)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Adaptive Model-based Dose-finding</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>SEAMLESS DESIGNS</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Dose-escalation based on efficacy/toxicity</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Learning/Confirming in Phase III</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

**Objective:** to use the posterior probabilities of hypotheses of interest as a basis for interim decisions (*Proper Bayesian*) or to explicitly assess the losses associated with consequences of stopping or continuing the study (*Decision-theoretic Bayesian*).

- **AR:** equal randomization or *play-the-winner* (next patient is allocated to the currently superior treatment) or *bandit designs* (minimizing the number of patients allocated to the inferior treatment)
- **SaR:** not specified
- **StR:** not formally pre-specified stopping criterion, or using a *skeptical prior* for stopping for efficacy and an *enthusiastic prior* for stopping for futility, or using *backwards induction*
- **DR:** update the posterior distribution; formal incorporation of external evidence; inference not affected by the number and timing of IAs

**References:** Berry (2001, 2004); Berry et al. (2001); Spiegelhalter et al. (2004).
Pairwise comparisons with GSD

- **Objective:** compare multiple treatments with a control; focus on type I error rate rather than power

  - A simple Bonferroni approximation is only slightly conservative
  - Treatments may be dropped in the course of the trial if they are significantly inferior to others
  - “Step-down” procedures allow critical values for remaining comparisons to be reduced after some treatments have been discarded

- **References:** Follmann et al (1994)

---

p-value combination tests

- **Objective:** compare multiple treatments with a control in a two-stage design allowing integration of data from both stages in a confirmatory trial

- **Focus:** control of multiple (familywise) Type I error level

- **Great flexibility:**
  - General distributional assumptions for the endpoints
  - General stopping rules and selection criteria
  - Early termination of the trial
  - Early elimination of treatments due to lack of efficacy or to safety issues or for ethical/economic reasons

- **References:** Bauer&Kieser (1994); Liu&Pledger (2005)
**Objective:** target the MTD (Phase I) or the best safe dose (Phase I/II) or find the therapeutic window

- **AR:** non-parametric (3+3 rule, up-and-down)
  - or model-based (Continual Reassessment Methods)
  - or Escalation With Overdose Control (EWOC)
  - or Bayesian Decision Design
  - or Bayesian Optimal Design
  - or Penalized Adaptive D-optimal Design
- **SaR:** cohorts of fixed size or in two stages (Storer design)
- **StR:** no early stopping or stopping by design (e.g. 3+3 rule)
- **DR:** update model parameters (for model-based AR)

**References:** O'Quigley et al.; Babb et al.; Edler; O'Quigley
Adaptive Model-based Dose-finding

- **Objective**: find the optimal dose; working model for the dose-response; dose sequence identified in advance
  - AR: *Bayesian* (based on predictive probabilities: smallest average posterior variance) or *frequentist* (based on optimal experimental design: maximum information per cost)
  - SaR: cohorts of fixed size or after each observation
  - StR: stopping for futility or when the optimal dose for confirmatory stage is sufficiently well known (estimation!)
  - DR: update model parameters, Bayesian predictions of long-term endpoint using a longitudinal model

- **References**: Berry et al. (2001); Dragalin & Fedorov; Fedorov & Leonov

Adaptive Dose-finding (Flexible Designs)

- **Objective**: establishing a dose-response relationship or combining Phase II/III using p-value combination tests
  - AR: drop or add doses
  - SaR: sample size reassessment for the next stage
  - StR: early stopping for futility or early termination of some inferior doses
  - DR: adapting hypotheses, primary endpoint, test statistics, inserting or skipping IAs

- **References**: Bauer & Kohne; Lehmacher et al
Seamless Designs

- Two-stage adaptive designs
  - 1st Stage: treatment (dose) selection – “learning”
  - 2nd Stage: comparison with control – “confirming”
- Treatment selection may be based on a short-term endpoint (surrogate), while confirmation stage uses a long-term (clinical) endpoint
- 2nd Stage data and the relevant groups from 1st Stage data are combined in a way that
  - Guarantees the Type I error rate for the comparison with control
  - Produces efficient unbiased estimates and confidence intervals with correct coverage probability
Selection and testing

- **Objective:** to select the “best” treatment in the 1st stage and proceed to the 2nd stage to compare with control

- **Focus:**
  - overall type I error rate is maintained (TSE)
  - trial power is also achieved (ST)
  - selection is based on surrogate (or short-term) endpoint (TS)

- **Method includes:**
  - early termination of the whole trial
  - early elimination of inferior treatments

- **References:** Thall, Simon & Ellenberg; Stallard & Todd; Todd & Stallard

Bayesian model-based designs

- **Objective:** adaptive dose ranging within a confirmatory trial

- **Focus:** efficient learning, effective treatment of patients in the trial

- **Method includes:**
  - **AR:** to maximize information about dose response
  - **SaR:** Frequent analysis of the data as it accumulates
  - Seamless switch to confirmatory stage without stopping enrollment in a double-blind fashion
  - Use of longitudinal model for prediction of the clinical endpoint

- **References:** Berry et al; Inoue et al
Classification

<table>
<thead>
<tr>
<th>Classification</th>
<th>Compound Progression Stages</th>
</tr>
</thead>
<tbody>
<tr>
<td>Disease selection</td>
<td>Target to tractable hit to candidate</td>
</tr>
<tr>
<td>Target Family selection</td>
<td>FTIM to Commit to PoC/Phase II</td>
</tr>
<tr>
<td>Candidate to Phase III</td>
<td>Phase III to launch</td>
</tr>
<tr>
<td>Lifecycle Management</td>
<td></td>
</tr>
</tbody>
</table>

SINGLE ARM TRIALS
- Two-stage Designs
- Screening Designs

TWO-ARM TRIALS
- Group Sequential Designs
- Information Based Designs
- Adaptive GSD (Flexible Designs)

MULTI-ARM TRIALS
- Bayesian Designs
- Group Sequential Designs
- Flexible Designs

DOSE-FINDING STUDIES
- Dose-escalation designs
- Dose-finding designs (Flexible Designs)
- Adaptive Model-based Dose-finding

SEAMLESS DESIGNS
- Dose-escalation based on efficacy/toxicity
- Learning/Confirming in Phase II/III

The objective of a clinical trial may be either:
- to target the MTD or MED or to find the therapeutic range
- or to determine the OSD (Optimal Safe Dose) to be recommended for confirmation
- or to confirm efficacy over control in Phase III clinical trial

This clinical goal is usually determined by:
- the clinicians from the pharmaceutical industry
- practicing physicians
- key opinion leaders in the field, and
- the regulatory agency
Achieving the goals

- Once agreement has been reached on the objective, it is the statistician's responsibility to provide the appropriate design and statistical inferential structure required to achieve that goal.

There are plenty of available designs on statistician’s shelf.
- The greatest challenge is their implementation.
- Adaptive designs have much more to offer than the rigid conventional parallel group designs in clinical trials.
References

- Brannath W, Posch M, Bauer P. Recursive combination tests. JASA 2002;97: 236-244.

PARMA Adaptive Designs Working Group
Adaptive Seamless Designs for Phase IIb/III Clinical Trials

Jeff Maca, Ph.D.
Assoc. Director, Biostatistics
Novartis Pharmaceuticals
Primary PhRMA references

- PhRMA White Paper sections:
  
  

Outline

- Introduction and motivation of adaptive seamless designs (ASD)
- Statistical methodology for seamless designs
- Considerations for adaptive design implementation
- Simulations and comparisons of statistical methods
Introduction and Motivation

Reducing time to market is/has/will be a top priority in pharmaceutical development

- Brings valuable medicines to patients sooner
- Increases the value of the drug to the parent company

Adaptive seamless designs can help reduce this development time

Definitions

Seamless design
- A clinical trial design which combines into a single trial objectives which are traditionally addressed in separate trials

Adaptive Seamless design
- A seamless trial in which the final analysis will use data from patients enrolled before and after the adaptation (inferentially seamless)
Adaptive Seamless Designs

Primary objective – combine “dose selection” and “confirmation” into one trial

- Although dose is most common phase IIb objective, other choices could be made, e.g. population
- After dose selection, only change is to new enrollments (patients are generally not re-randomized)
- Patients on terminated treatment groups could be followed
- All data from the chosen group and comparator is used in the final analysis. Appropriate statistical methods must be used
Statistical methodology

Statistical methodology for Adaptive Seamless Designs must account for potential biases and statistical issues

- Selection bias (multiplicity)
- Multiple looks at the data (interim analysis)
- Combination of data from independent stages

Simple Bonferroni adjustment

Test final hypothesis at $\alpha / n_{trt}$

- Accounts for selection bias: multiplicity adjustment
- Multiple looks at the data: not considered
- Combination of data from stages by simple pooling
  - In some sense, ignores that there was an interim analysis at all
- Most conservative approach, simple to implement
- No other adjustments (i.e., sample size) can be made
Statistical Methodology – Closed Testing

And alternative and more powerful approach is a closed testing approach, and combination of p-values with inverse normal method

Methodology combines:

- Closed testing of hypothesis
- Simes adjustment of p-values for multiplicity
- Combines data (p-values) from stages via the inverse normal method (or Fisher’s combination)

Closed test procedure

- $n$ null hypotheses $H_1, \ldots, H_n$
- Closed test procedure considers all intersection hypotheses.
- $H_i$ is rejected at global level $\alpha$ if all hypotheses $H_i$ formed by intersection with $H_i$ are rejected at local level $\alpha$

$H_{12} = H_1 \cap H_2$

$H_1$ can only be rejected at $\alpha=0.05$ if $H_{12}$ is also rejected at $\alpha=0.05$
A typical study with 3 doses $\Rightarrow$ 3 pairwise hypotheses.

- Multiplicity can be handled by adjusting p-values from each stage using Simes procedure

$$q_s = \min_{i \in S} \frac{|S|}{i} p(i)$$

S is number of elements in Hypothesis, $p(i)$ is the ordered P-values

---

**Inverse Normal Method**

If $p_1$ and $p_2$ are generated from independent data, then

$$C(p_1, p_2) = \sqrt{t_0} \cdot \Phi^{-1}(1 - p_1) + \sqrt{1 - t_0} \cdot \Phi^{-1}(1 - p_2)$$

will yield a Z test statistic

Note: For adaptive designs, typical value for $t_0$ is $n_1/ (n_1+n_2)$
Statistical Methodology – Example

Example: Dose finding with 3 doses + control

- Stage sample sizes: $n_1 = 75$, $n_2 = 75$
- Unadjusted pairwise p-values from the first stage:
  - $p_{1,1} = 0.23$, $p_{1,2} = 0.18$, $p_{1,3} = 0.08$
- Dose 3 selected at interim
- Unadjusted p-value from second stage: $p_{2,3} = 0.01$

Three-way test:
- $q_{1,123} = \min(3 \times 0.08, 1.5 \times 0.18, 1 \times 0.23) = 0.23$
- $q_{2,123} = p_{2,3} = 0.01$
- $C(q_{1,123}, q_{2,123}) = 2.17 \Rightarrow P\text{-value} = 0.015$
Statistical Methodology – Example

Two-way tests:
- \( q_{1,13} = \min(2*0.08, 1*0.23) = 0.16 \)
- \( q_{1,23} = \min(2*0.08, 1*0.18) = 0.16 \)
- \( q_{2,13} = q_{2,23} = p_{2,3} = 0.01 \)
- \( C(q_{1,13}, q_{2,13}) = C(q_{1,23}, q_{2,23}) = 2.35 \rightarrow \text{P.value} = 0.0094 \)

Final test:
- \( q_{1,3} = p_{1,3} = 0.08 \)
- \( q_{2,3} = p_{2,3} = 0.01 \)
- \( C(q_{1,13}, q_{2,13}) = C(q_{1,23}, q_{2,23}) = 2.64 \rightarrow \text{P.value} = 0.0042 \)
- **Conclusion:** Dose 3 is effective
Choosing sample sizes

- There are two sample sizes to consider for a seamless design, \( n_1, n_2 \)
- If \( t \) is the number of treatments, the total size \( N \) is:
  \[
  N = t \times n_1 + 2 \times n_2
  \]
- The larger \( n_1 \), the better job of choosing the "right" dose. However, this makes the total much larger.
- Power can be determined by simulation, and is also a function of the (unknown) dose response

Simulation for power comparison

To compare the two methods for analyzing an adaptive seamless designs, the following parameters were used:

- Sample sizes were \( n_1 = n_2 = 75 \)
- Primary endpoint is normal, with \( \sigma = 12 \)
- One dose was selected for continuation
- Various dose responses were assumed
- 20,000 reps used for simulations (error \( = \pm 0.5\% \))
Statistical Methodology – Power

Simulation for power comparison

Selecting 1 treatment group from 2 possible treatments

<table>
<thead>
<tr>
<th>Dose Response (Δ placebo)</th>
<th>Power Bonferroni</th>
<th>Power Closed Test</th>
</tr>
</thead>
<tbody>
<tr>
<td>0 , 4.5</td>
<td>83.1%</td>
<td>83.2%</td>
</tr>
<tr>
<td>4.5, 4.5</td>
<td>91.0%</td>
<td>92.2%</td>
</tr>
</tbody>
</table>

Selecting 1 treatment group from 3 possible treatments

<table>
<thead>
<tr>
<th>Dose Response (Δ placebo)</th>
<th>Power Bonferroni</th>
<th>Power Closed Test</th>
</tr>
</thead>
<tbody>
<tr>
<td>0 ,0 , 4.5</td>
<td>79.4%</td>
<td>78.9%</td>
</tr>
<tr>
<td>4.5, 4.5, 4.5</td>
<td>90.8%</td>
<td>92.7%</td>
</tr>
</tbody>
</table>
Considerations for Seamless Designs

With the added flexibility of seamless designs, comes added complexity.

- Careful consideration should be given to the feasibility for a seamless design for the project.
- Not all projects can use seamless development
- Even if two programs can use seamless development, one might be better suited than the other
- Many characteristics add or subtract to the feasibility

Enrollment vs. Endpoint

- The length of time needed to make a decision relative to the time of enrollment must be small
  - Otherwise enrollment must be paused
- Endpoint must be well known and accepted
  - If the goal of Phase II is to determine the endpoint for registration, seamless development would be difficult
- If surrogate marker will be used for dose selection, it must be accepted, validated and well understood
Considerations for Seamless Designs

Clinical Development Time

- There will usually be two pivotal trials for registration
- Entire program must be completed in shorter timelines, not just the adaptive trial

Logistical considerations

- Helpful if final product is available for adaptive trial (otherwise bioequivalence study is needed)
- Decision process, and personnel must be carefully planned and pre-specified
Considerations for Seamless Designs

**Novel drug or indication**

- Decision process which will be overly complicated could be an issue with an external board
- If there are a lot of unknown issues with the indication or drug, a separate phase II trial would be better
- However, getting a novel drug to patients sooner increases the benefit of seamless development

**Conclusions**

- Adaptive seamless designs have an ability to improve the development process by reducing timelines for approval
- Statistical methods are available to account for adaptive trial designs
- Extra planning is necessary to implement an adaptive seamless design protocol
- Benefits should be carefully weighed against the challenges of such designs before implementation
References

- Schaid DJ, Wiegand S, Therneau TM. Optimal two stage screening designs for survival comparisons. *Biometrika* 1990;77:659-663.

References

- Inoue LYT, Thall PF, Berry DA. Seamlessly expanding a randomized phase II trial to phase III. *Biometrics* 2002;58:823-831.
Adaptive Designs
Sample size re-estimation: A review and recommendations

Keaven M. Anderson
Clinical Biostatistics and Research Decision Sciences
Merck Research Laboratories

Outline
- Introduction/background
- Methods
  - Fully sequential and group sequential designs
  - Adaptive sample size re-estimation
    - Background
    - Nuisance parameter estimation/internal pilot studies
      - Blinded sample size re-estimation
      - Unblinded sample size re-estimation
    - Conditional power and related methods
- Discussion and recommendations
- Case studies
- References
Background

- Origin
  - PhRMA Adaptive Design Working Group

- Focus
  - Late-stage (Phase III, IV) sample size re-estimation
  - Frequentist methods
  - Control of Type I error
  - Potential for bias is critical in these ‘confirmatory’ trials
    - Implications for logistical issues

Introduction

*Adaptive designs* allow design specifications to be changed based on accumulating data (and/or information external to the trial)

Extensive literature exists on adapting through sample size re-estimation, the topic of this talk

Since sample size in group sequential and fully sequential trials are data-dependent, we consider these to be included in a broad definition of adaptive design/sample size re-estimation
Introduction

Why consider sample size re-estimation?

- Minimize number of patients exposed to inferior or highly toxic treatment
- Right-size the trial to demonstrate efficacy
  - Reduce or increase sample size
- Stop the trial for futility if insufficient benefit
- Incorporate new internal or external information into a trial design during the course of the trial

Introduction

Reasons for Unplanned Adaptation

- Information that could not have been anticipated prior to trial start has become available.
  - Regulators change the primary efficacy endpoint from mean change in viral load to % of patients with viral load below the detectable limit in HIV patients.
  - Regulators decide that the new treatment needs to win on more than the one efficacy endpoint used to size the trial
- Adaptation, not planned at the design stage, is used to possibly to ‘bail out’ a trial
  - A competing therapy removed from market, allowing a lesser treatment benefit to be viable.
  - Sponsor ‘changes mind’ about minimal treatment effect of interest

This topic will receive minimal discussion here
The Problem

- In order to appropriately power a trial, you need to know:
  - The true effect size you wish to detect
  - Nuisance parameters such as
    - Variability of a continuous endpoint
    - Population event rate for a binary outcome or time to event
  - Other ancillary information (e.g., correlation between co-primary endpoints needed to evaluate study-level power)

- Inappropriate assumptions about any of these factors can lead to an underpowered trial

<table>
<thead>
<tr>
<th>Consequences of incorrect planning for treatment difference $\delta$ and/or standard deviation $\sigma$ ($\alpha=0.05$, planned Power=90%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Over-estimate $\delta$ or under-estimate $\sigma$ by 50%</td>
</tr>
<tr>
<td>Under-estimate $\delta$ or over-estimate $\sigma$ by 50%</td>
</tr>
<tr>
<td>Over-estimate $\delta$ AND under-estimate $\sigma$ by 50%</td>
</tr>
<tr>
<td>Under-estimate $\delta$ AND over-estimate $\sigma$ by 50%</td>
</tr>
<tr>
<td>Under-estimate $\delta$ AND under-estimate $\sigma$ by 50%</td>
</tr>
</tbody>
</table>
Solutions to the problem

- Plan a fixed trial conservatively
  - Pro: trial should be well-powered
  - Cons: Can lead to lengthy, over-powered, expensive trial

- Use group sequential design and plan conservatively
  - Pro: can power trial well and stop at appropriate, early interim analysis if your assumptions are too conservative
  - Con: over-enrollment occurs past definitive interim analysis because it takes time to collect, clean and analyze data

- Use adaptive design
  - Pro: can decide to alter trial size based on partial data or new, external information
  - Cons: methods used to adapt must be carefully chosen, regulatory scrutiny over methods and ‘partial unblinding,’ may not improve efficiency over group sequential design

Outline

- Introduction/background
- Methods
  - Fully sequential and group sequential designs
  - Adaptive sample size re-estimation
    - Background
    - Nuisance parameter estimation/internal pilot studies
      - Blinded sample size re-estimation
      - Unblinded sample size re-estimation
    - Conditional power and related methods
- Discussion and recommendations
- Case studies
- References
Fully sequential design

- Not commonly used due to continuous monitoring
- May be useful to continuously monitor a rare serious adverse effect
  - Intracranial hemorrhage in a thrombolytic/anti-platelet trial
  - Intussusception in rotavirus vaccine trial
- Unblinded analysis suggests need for an independent monitor or monitoring committee
- References
  - Wald (1947), Sequential Analysis
  - Siegmund (1985), Sequential Analysis: Tests and Confidence Intervals

Group sequential design

- Classic
  - Fixed sample sizes for interim and final analyses
  - Pre-defined cutoffs for superiority and futility/inferiority at each analysis
  - Trial stops (adapts) if sufficient evidence available to decide early
  - Independent data monitoring committee often used to review unblinded interim analyses
- Variations
  - Adjustment of interim analysis times (spending functions)
  - Adjustment of total sample size or follow-up based on, for example, number of events (information-based designs)
- Properties well understood and design is generally well-accepted by regulators
- See: Jennison and Turnbull (2000): Group Sequential Methods with Applications to Clinical Trials
Outline

- Introduction/background
- Methods
  - Fully sequential and group sequential designs
  - Adaptive sample size re-estimation
    - Background
    - Nuisance parameter estimation/internal pilot studies
      - Blinded sample size re-estimation
      - Unblinded sample size re-estimation
    - Conditional power and related methods
- Discussion and recommendations
- Case studies
- Evolving issues

The Opportunity

- Size the study appropriately to reach study objectives in an efficient manner based on interim data that offers more accurate information on
  - Nuisance parameter
    - Within-group variability (continuous data)
    - Event rate for the control group (binary data)
    - # of subjects and amount of exposure needed to capture adequate occurrences of time-to-event endpoint
  - Treatment effect
  - Other ancillary information (e.g., correlation between co-primary endpoints needed to evaluate study-level power)
- Ensure that we will have collected enough exposure data for safety evaluation by the end of the study
SSR Strategies

- Update sample size to ensure power as desired based on interim results
  - Internal pilot studies: Adjust for nuisance parameter estimates only
    - Blinded estimation
    - Unblinded estimation
    - Testing strategy: no adjustment from usual test statistics
  - Adjusting for interim test statistic/treatment effect
    - All methods adjust based on unblinded treatment difference
    - Adjust sample size to retain power based on interim test statistic
      - Assume observed treatment effect at interim
      - Assume original treatment effect
    - Testing strategy: adjust stage 2 critical value based on interim test statistic

SSR - Issues

- Planned vs Unplanned (at the design stage)
- Control of Type I error rate and power
- If we have a choice, do we do it blinded or unblinded?
  - If we do it unblinded, how do we maintain confidentiality?
    - Who will know the exact SSR rule?
    - Who will do it, a third party?
    - Who will make the recommendation, a DMC?
    - How will the results be shared?
    - Who will know the results, the sponsors, investigators?
- When is a good time to do SSR?
- Regulatory acceptance
SSR reviews

- These all concern what might be considered ‘internal pilot’ studies
  - Friede and Kieser, Statistics in Medicine, 2001; 20:3861-73
    - Also Biometrical Journal, 2006; 48:537-555
  - Gould, Statistics in Medicine, 2001; 20:2625-43
  - Jennison and Turnbull, 2000, Chapter 14
  - Zucker, Wittes, Schabenberger, Brittain, 1999; Statistics in Medicine, 18:3493-3509

Blinded SSR

- When SSR is based on nuisance parameters
  - Overall variability (continuous data)
  - Overall rate (binary data)

- Advantage
  - No need to break the blind.
  - In-house personnel can do it.
  - Minimal implication for Type I error rate.

- Disadvantage
  - The estimate of the nuisance parameter could be wrong, leading to incorrect readjustment.
Blinded SSR

- Internal pilot studies to estimate nuisance parameter without adjustment of final test statistic/critical value
- Gould and Shih (1992)
  - Uses EM algorithm to estimate individual group means or event rates
  - Estimates variance (continuous case)
  - Updates estimate of sample size required for adequate power
  - Software: Wang, 1999
- Friede and Kieser (2001)
  - Assume treatment difference known (no EM algorithm required)
  - Adjust within group sum of squares using this constant
- Type I error and power appear good
  - Some controversy over appropriateness of EM (Friede and Kieser, 2002?; Gould and Shih, 2005?)
- Question to ask:
  - How well will this work if treatment effect is different than you have assumed for the EM procedure?
  - Will it be under- or over-powered?
  - Group sequential version (Gould and Shih, 1998) may bail you out of this

Blinded SSR gone wrong?

- Observed combined event rate
  - Assuming 20% placebo event rate
  - Assuming 25% reduction

- 20% vs. 12%: N=436
- 17.8% vs. 14.2% n=1346

90% power, 2-sided Type I error 5%
Unblinded SSR

- **Advantage**
  - Could provide more accurate sample-size estimate.

- **Disadvantages**
  - Re-estimate sample size in a continuous fashion can reveal interim difference.
  - There could be concerns over bias resulting from knowledge of interim observed treatment effect.
  - Typically require an external group to conduct SSR for registration trials.
    - Interim treatment differences can be misleading
      - Due to random variation or
      - If trial conditions change

Internal Pilot Design: Continuous Data

- Adjusts sample size using only nuisance parameter estimate
  - Question to ask: does updated sample size reveal observed treatment effect?

- Use some fraction of the planned observations to estimate error variance for continuous data, modify final sample size, allow observations used to estimate the variance in the final analysis.

- Plug the new estimate into the SS formula and obtain a new SS. If the SS re-estimation involves at least 40 patients per group, simulations have shown (Wittes et al, SIM 1999,18:3481-3491; Zucker et al, SIM 1999,18:3493-3509)
  - The type I error rate of the unadjusted (naïve) test is at about the desirable level if we do not allow SS to go down
  - The unadjusted test could lead to non-trivial bias in the type I error rate if we allow the SS to go down
  - Power OK

- Coffey and Muller (Biometrics, 2001, 57:625-631) investigated ways to control the type I error rate (including different ways to do SSR).

- Denne and Jennison, (Biometrika, 1999) provide a group sequential version
Internal pilot design: binary data

- Estimate control group event rate at interim
  - Type I error OK if interim n large enough
- Options (see Jennison and Turnbull, 2000 for power study)
  - Assume $p_1 - p_2$ fixed
    - Power appears OK
  - Assume $p_1/p_2$ fixed
    - Can be underpowered

Combination tests

- Methods for controlling Type I error
- The invariance principle – calculate separate standardized test statistics from different stages and combine them in a predefined way to make decisions.
  - Weighting of a stage does not increase if sample size for that stage is increased, meaning that individual observations for that stage are down-weighted in the final test statistic
  - Efficiency issue (Tsiatis and Mehta, 2003)
- Many methods available, including
  - Fisher's combination test (Bauer, 1989)
  - Conditional error functions (Proschan and Hunsberger, 1995; Liu and Chi, 2001)
  - Inverse normal method (Lehmacher and Wassmer, 1999)
  - Variance spending (Fisher, 1998)
Combination tests

- Apply combination test method to determine the critical value for the second stage based on the observed data from the first stage.

- Make assumption on treatment effect; options include:
  - Observed effect (highly variable)
  - External estimate
  - Original treatment effect used for sample size planning

- Compute next stage sample size based on critical value, set conditional power to originally desired power given interim test statistic and assumed second stage treatment effect
  - Generally, will only raise sample size – not lower

Outline

- Introduction/background
- Methods
  - Fully sequential and group sequential designs
  - Adaptive sample size re-estimation
    - Background
    - Nuisance parameter estimation/internal pilot studies
      - Blinded sample size re-estimation
      - Unblinded sample size re-estimation
    - Conditional power and related methods
- Discussion and recommendations
- Case studies
- References
Blinded vs Unblinded SSR

- For SSR due to improved estimate on variance (continuous data), Friede and Kieser (Stat in Med, 2001) conclude that there is not much gain in conducting SSR unblinded.
  - They only studied a constant treatment effect

- Statistical approaches to control Type I error rate particularly important when adjusting sample size to power for observed treatment difference

- Decisions related to SSR because of inaccurate assumption on the nuisance parameters can differ significantly from those due to inaccurate assumption on the treatment effect.

Relative efficiency of SSR methods

- Internal estimates of treatment effect lead to very inefficient trials (Jennison and Turnbull, 2003) due to the variability of the estimates.

- External or pre-determined minimal treatment effect assumptions can yield comparable efficiency to group sequential (Liu and Chi, 2001, Anderson et. al, 2004)
  - Adding in a maximum sample size adjustment limit can improve over group sequential (Posch et al, 2003)
  - Based on comparison of optimal group sequential and adaptive designs, improvement of adaptive designs over group sequential is minimal (Jennison and Turnbull, SIM 2006; see also Anderson, 2006)

- Use of sufficient statistic design rather than weighted combination test improves efficiency (Lokhnygina, 2004)
Group Sequential vs SSR Debate

- **Efficiency**
  - The adaptive designs for SSR using combination tests with fixed weights are generally inefficient.
  - Efficient adaptive designs for SSR have little to offer over efficient group sequential designs in terms of sample size. However, the latter might require more interim analyses and offer minimum gain. In addition, the comparisons were made as if we knew the truth.

- **Flexibility and upfront resource commitment**
  - SSR offers flexibility and reduces upfront resource commitment. The flip side is the need to renegotiate budget and request additional drug supply when an increase in SS is necessary.
  - SSR addresses uncertainty at the design stage.

SSR is fluid and can respond to changing environment both in terms of medical care and the primary endpoint to assess treatment effect.
- The above is important for trials lasting 3-5 years when environmental changes are expected.

Need to ascertain treatment effect in major subgroups even though the subgroups are not the primary analysis populations
- Xigris for disease severity groups
- Cozaar for race groups
Recommendation #1

- Before considering adaptive sample-size re-estimation, evaluate whether or not group sequential design is adequate
  - Pros:
    - Regulatory acceptance
    - Well-understood methods allow substantial flexibility
    - Experienced monitoring committee members available
  - Cons:
    - May not work well in some situations when trial cannot be stopped promptly (long follow-up, slow data collection, cleaning or analysis)

Recommendation #2

- Anticipate as much as possible at the planning stage the need to do SSR to incorporate information that will accumulate during the trial
  - Treatment effect size
  - Nuisance parameters
  - The effect of environmental changes on the design assumptions

- Do not use SSR to
  - Avoid up-front decisions about planning
  - As a ‘bait-and-switch’ technique where a low initial budget can be presented with a later upward sample size adjustment.
Recommendation #3

- For SSR based on variance, consider using blinded SSR
  - However, when there is much uncertainty about the treatment effect, consider using unblinded SSR.

- For a binary outcome, one can either do blinded SSR based on the overall event rate or an unblinded SSR based on the event rate of the control group. There is no clear preference, choice dependant on several factors.
  - If there is much uncertainty about treatment effect, unblinded SSR using conditional power methods (see next slides).
  - If SSR is blinded, consider conducting interim analysis to capture higher than expected treatment effect early.

Recommendation #4

- To help maintain confidentiality of the interim results, we recommend
  - Do not reveal exact method for adjusting sample size.
  - Make the outcome of SSR discrete with only 2-3 options.

- Under the first approach, details on SSR methodology will not be described in the protocol, but documented in a stand-alone statistical analysis plan for SSR not available to study personnel.

- For SSR based on observed treat effect (continuous case), it will be beneficial to base SSR on both variability and effect.

- We recommend that the protocol include the maximum sample size allowed to minimize the need to go back to the IRB.
Recommendation #5

- For unblinded SSR
  - Invite a third party to do the calculations following a pre-specified rule.
  - If possible, combine SSR with a group sequential design where SSR will be conducted at the same time with an interim analysis.
    - Convene a DMC (or preferably an IDMC) to review the SSR recommendation from the third party. If an IDMC is used, the IDMC statistician can carry out the SSR.
  - Assuming Recommendation #4 is followed, the new sample size will be communicated to the sponsor. The investigators will be told to continue enrollment.

Recommendation #6

- Carefully consider the number of times to do SSR.
  - E.g., for variance estimation, is once enough?
- Timing of the SSR should be based on multiple considerations such as
  - available info at the design stage,
  - disease,
  - logistics
    - delay from enrollment until follow-up complete and data available
    - enrollment rate,
  - Method
    - whether the SSR will be based on variance or treatment effect
  - Gould and Shih (1992) recommend early update as soon as variance estimate stable due to administrative considerations, while Sandvik et al. (1996) recommend as late as possible to get accurate variance estimate
Recommendation #7

- Acceptance of SSR by regulators varies, depending on the reasons for SSR. In general, blinded SSR based on a nuisance parameter is acceptable.

- When proposing unblinded SSR, should include
  - The objective for SSR
  - Statistical methodology including the control of Type I error
  - When to do the SSR
  - How to implement (e.g., DMC, third party)
  - How to maintain confidentiality
  - How will the results be shared
  - Efficiency (power/sample size) considerations

- Discuss the plan with regulatory agencies in advance.

Outline

- Introduction/background
- Methods
  - Fully sequential and group sequential designs
  - Adaptive sample size re-estimation
    - Background
    - Nuisance parameter estimation/internal pilot studies
      - Blinded sample size re-estimation
      - Unblinded sample size re-estimation
    - Conditional power and related methods
- Discussion and recommendations
- Case studies
- References
Case Study #1: Blinded SSR Based on Variance

- Drug X low-dose, high-dose and placebo
- Main efficacy endpoint - percent change in continuous primary outcome
- N=270 provides 90% power to detect a 10% difference versus placebo
  - estimate of the variability obtained from study performed in a different setting (SD estimated at 20%)
  - seasonal disease: interim analysis performed during long pause in enrollment
- Recruitment was anticipated to be difficult,
  - specified in protocol that a blinded estimate of the variability of primary outcome would be computed when the sample size is 100
  - If the variability is less than anticipated (e.g., SD ≤15%) then the final sample size could be reduced
  - If the variability is greater than anticipated (e.g., > 25%), the main comparison would be the pooled Drug X groups (low and high-dose) vs. placebo

Case Study: REST Study Design

- Sample size: Minimum of 60,000 (1V:1P)
- Age: 6 to 12 weeks at enrollment
- Dose regimen: 3 oral doses of Rotavirus every 4-10 wks
- Formulation: Refrigerated liquid buffer/stabilizer intended for licensure
- Potency: Release range intended for licensure
- Study Period: 2001 to 2005
Primary Safety Hypothesis

- Oral RotaTeq™ will not increase the risk of intussusception relative to placebo within 42 days after any dose
- To satisfy the primary safety hypothesis, 2 criteria must be met:
  1. During the study, the vaccine/placebo case ratio does not reach predefined unsafe boundaries being monitored by the DSMB
     - 1 to 42 days following any dose
     - 1 to 7 days following any dose
  2. At the end of the study, the upper bound of the 95% CI estimate of the relative risk of intussusception must be ≤10

Safety Monitoring for Intussusception (IT)

<table>
<thead>
<tr>
<th>IT Surveillance at Study Sites</th>
<th>Safety Endpoint Adjudication Committee</th>
<th>Data and Safety Monitoring Board (DSMB)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Active surveillance - contacts on day 7, 14, and 42</td>
<td>Pediatric surgeon, radiologist, &amp; emergency department specialists</td>
<td>Unblind each case as it occurs and make recommendations about continuing</td>
</tr>
<tr>
<td>Passive surveillance - parent education</td>
<td>Use specific case definition</td>
<td>Review all safety data every 6 months</td>
</tr>
<tr>
<td>Intense surveillance during 6 weeks after each dose</td>
<td>Individual &amp; collaborative adjudications</td>
<td></td>
</tr>
</tbody>
</table>

Potential IT Case → Positively Adjudicated IT Case
Safety Monitoring for Intussusception

- Trial utilizes two predefined stopping boundary graphs for the 1 to 7 and 1 to 42 day ranges after each dose

- Stopping boundaries were developed to ensure that the trial will be stopped if there is an increased risk of intussusception within these day ranges

- DSMB plots intussusception cases on graphs and makes recommendations about continuing the study

Intussusception 42 days post-dose

Unsafe Boundary (LB on 95% CI > 1.0)

Acceptable Safety Profile (UB on 95% CI ≤ 10)
REST Group Sequential Study Design

Enroll subjects

Monitor continuously for intussusception (IT) → Stop trial early if detect increased risk of IT

Evaluate statistical criteria with 60,000 subjects

Monitor continuously and stop early if detect increased risk of IT

Primary hypothesis satisfied: Stop

Data inconclusive: Enroll 10,000 more infants

Evaluate statistical criteria with 70,000 subjects

Comments on REST Study Design

- The goal of the REST study design and the extensive safety monitoring was to provide:
  - High probability that a safe vaccine would meet the end of study criteria; and simultaneously
  - High probability that a vaccine with increased intussusception risk would stop early due to ongoing safety monitoring

- The statistical operating characteristics of REST were estimated using Monte Carlo simulation
**Statistical Operating Characteristics of REST***

<table>
<thead>
<tr>
<th>Risk Scenario</th>
<th>Probability of reaching unsafe monitoring boundary</th>
<th>Probability of meeting end of study safety criteria</th>
</tr>
</thead>
<tbody>
<tr>
<td>Safe Vaccine (RR=1)</td>
<td>~6%</td>
<td>~94%</td>
</tr>
<tr>
<td>RRV-TV Risk Profile**</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Case-control study</td>
<td>~91%</td>
<td>~9%</td>
</tr>
<tr>
<td>Case-series study</td>
<td>~85%</td>
<td>~15%</td>
</tr>
</tbody>
</table>

* Assumes background intussusception rate of 1/2000 infant years and 102 days of safety follow-up over three doses.


** References (Blinded SSR)**

References (Others)

- Wittes, Brittain (1990) Stat in Medicine, 9:65-72
- Bristol (1993) J of Biopharm Stat, 3:159-166
- Birkett, Day (1994) Stat in Medicine, 13:2455-2463
- Shih, Zhao (1997) Stat in Medicine, 16:1913-1923
- Fisher (1998) Stat in Medicine, 17:1551-1562
- Denne, Jennison (1999) Stat in Medicine, 18:1575-1585
- Kieser, Friede (2000) Stat in Medicine, 19:901-911
- Shun, Yuan, Brady, Hsu (2001) Stat in Medicine, 20:497-513
References (Others)

- Sandvik, Erikssen, Mowinckel, Rodland (1995), Statistics in Medicine, 15:1587-90
- Denne and Jennison (2000), Biometrika, 87:125-134
Adaptive Designs
Logistic, Operational and Regulatory Issues

Paul Gallo
Biostatistics and Statistical Reporting
Novartis

Primary PhRMA references

- PhRMA White Paper sections:
Outline

- Motivations and opportunities
- Cautions and challenges
- Logistic and feasibility issues
- Interim monitoring and confidentiality issues
  - review of current conventions
  - monitoring processes for adaptive designs
  - information conveyed by adaptive designs

General motivation

- The greater flexibility offered within the adaptive design framework has the potential to translate into more ethical treatment of patients within trials (possibly including the use of fewer patients), more efficient drug development, and better focusing of available resources.

- The potential appeal of adaptive designs is understandable, and motivates the current high level of interest in this topic.
Cautions

- But, being too eager, and proceeding without all relevant issues being fully thought out, is not advisable either.

- The question should be:
  - "What is the most appropriate (e.g., ethical, efficient) means at hand to address the research questions of importance?"

- rather than:
  - "How can adaptive designs be integrated into our program at all costs?"

Challenges

- Clearly, there will be many challenges to be addressed or overcome before adaptive designs become more widely utilized.
  - Statistical
  - Logistic
  - Procedural / regulatory
General considerations

- Like any new technology with challenges, some resistance is to be expected.
- Closer scrutiny is natural, and constructive.
- But we should not make “the perfect be the enemy of the good”.
- Can we address the challenges to a sufficient extent so that in particular situations the advantages outweigh the drawbacks?

Planning

- Adaptive designs are not a substitute for poor planning, and in fact will generally require more planning.
- They are part of a rational strategy to achieve research objectives more efficiently and ethically:
  - by utilizing knowledge gained from the study
  - in a manner which maintains the validity and interpretability of the results.
Feasibility issues

- Endpoint follow-up time vs recruitment speed
  - Shorter read-out time is generally favorable to adaptive designs.
  - *Surrogates* / early predictors can have a role.

- Timely data collection is important, as well as efficient analysis and decision-making processes.
  - *Electronic Data Capture* should be helpful.

Data quality

- All else being equal, *cleaner* is *better*

- But the usual trade-off exists:
  - cleaner takes *longer*, and results in less data being available for decisions
  - lack of data is a source of noise also!

- There is no requirement that data must be fully cleaned for adaptive designs.
  - Details of data quality requirements should be considered on a case-by-case basis.
Opportunities

- Early-phase trials may in the short term be the most favorable arena for wider-scale implementation of adaptive designs.
  - More uncertainties, and thus more opportunity for considering adaptation
  - Lesser regulatory concerns
  - Lower-risk opportunities to gain experience with ADs to learn, solve operational problems, and set the stage for more important applications.

Simulation

- Simulation will play an important role in planning of adaptive trials.
  - Detailed simulation scenarios should be of broad interest in evaluating adaptive design proposals (e.g., health authorities).
  - Main simulation results may be included in the protocol or analysis plan.
Monitoring / confidentiality issues

Issues relating to

- **monitoring** of accruing data
- **restriction of knowledge** of interim results
- and the **processes** of data review, decision-making and implementation

are likely to be critical in determining the extent and shaping the nature of adaptive design utilization in clinical trials.

Current monitoring conventions

- Monitoring of accruing data is of course a common feature in clinical trials. Most frequently for:
  - safety monitoring
  - formal group sequential plan allowing stopping for efficacy
  - lack of effect / futility judgments.

- Current procedures and conventions governing monitoring are a sensible starting point for addressing similar issues in trials with adaptive designs.
Current monitoring conventions

- As described in the FDA DMC guidance (2006): Comparative interim results and access to unblinded data should not be accessible to trial personnel, sponsor, investigators.

- Access to interim results diminishes the ability of trial personnel to manage the trial in a manner which is (and which will be seen by interested parties to be) completely objective.

Knowledge of interim results could introduce subtle, unknown biases into the trial, perhaps causing slight changes in characteristics of patients recruited, administration of the intervention, endpoint assessments, etc.

- Changes in “investigator enthusiasm”?

- The *equipoise* argument: knowledge of interim results violates equipoise.
Current conventions - sponsor

- FDA (2006): “Sponsor exposure to unblinded interim data . . . can present substantial risk to the integrity of the trial.”

- Risks include lack of objectivity in trial management; further unblinding, even if inadvertent; SEC requirements and fiduciary responsibilities, etc.

- Sponsor is thus usually *not* involved in monitoring of confirmatory trials.

Issues for adaptive designs

I. Adaptive designs will certainly require review of accruing data.
   - Who will be involved in the analysis, review, and decision-making processes?
   - Will operational models differ from those we’ve become familiar with?
   - Will sponsor perspective and input be relevant or necessary for some types of adaptations?
   - Will sponsors accept and trust decisions made confidentially by external DMCs in long-term trials / projects with important business implications (e.g., seamless Phase II / III)?
Issues for adaptive designs

II. An important distinction versus common monitoring situations: the results will be used to implement adaptation(s) which will govern some aspect of the conduct of the remainder of the trial.

- Can observers infer from viewing the actions taken information about the results which might be perceived to rise to an unacceptable level?

Analysis / review / decision process

- Concerns about confidentiality to ensure objective trial management, and potential bias from broad knowledge of interim results, should be no less relevant for adaptive designs than in other settings.

- The key principles to adhere to would seem to be:
  - separation / independence of the DMC from other trial activities
  - limitation of knowledge about interim treatment effects.
Analysis / review / decision process

- Adaptive design trials may utilize a single monitoring board for adaptations and other responsibilities (e.g., safety); or else a separate board may be considered for the adaptation decisions.

- DMCs in adaptive design trials may require additional expertise not traditionally represented on DMCs; perhaps to monitor the adaptation algorithm, or to make the type of decision called for in the adaptation plan (e.g., dose selection).

Sponsor participation

- Sponsor participation and knowledge of interim results in confirmatory adaptive trials may be a hard sell.
  - The “objective trial management” issue - sponsor can have some influence on trial management activities, even for individuals not directly participating in the trial.
  - There seems to be an assumption that information, once within the sponsor organization, may not be controlled, whether inadvertently or otherwise.
Sponsor participation

- **Proposal** - There should be potential for sponsor involvement in certain types of decisions if:
  - a strong rationale can be described whereby these individuals are needed for the best decision
  - the individuals are not involved in trial operations
  - all involved clearly understand the issues and risks to the trial, and adequate firewalls are in place
  - sponsor exposure to results is “minimal” for the needed decision, i.e., *only* at the adaptation point, *only* the relevant data (e.g., unlike a DMC with whom they may be working, which may have a broader ongoing role).

Information apparent to observers

- Adaptive designs may lead to changes in a trial which will be *apparent* to some extent - sample size, randomization allocation, population, dosage, treatment arm selection, etc., etc. - and can thus be viewed as providing some information to observers about the results which led to those changes.

- Considering the concerns which are the basis for the confidentiality conventions: can we distinguish between *types* and *amounts* of information, and how risky they would be in this regard?
Information apparent to observers

- **Note**: conventional monitoring is not immune from this issue.

- It has never been the case that *no information* can be inferred from monitoring; i.e., all monitoring has some potential action thresholds, and lack of action usually implies that such thresholds have not been reached.

Example – *Triangular test*

- **Design:**
  - Normal data, 2 group comparison
  - Study designed to detect $\Delta = 0.15$
  - 4 equally-spaced analyses
  - will require about 2276 patients.
Example – *Triangular test*

‘Christmas tree’ boundary

\[ \hat{\Delta} = \frac{Z}{V} \]

Continuation beyond the 3rd look would imply (barring over-ruling of the boundary) that the point estimate is between 0.076 and 0.106.

Doesn’t that convey quite a bit of information about the interim results?

In conventional GS design practice, this issue seems not to be perceived to compromise trials nor to discourage monitoring.
Information apparent to observers

- Presumably, in GS practice it’s viewed that reasonable balance is struck between the objectives and benefits of the monitoring and any slight potential for risk to the trial, with appropriate and feasible safeguards in place to minimize that risk.

- The same type of standard should make sense for adaptive designs.

Information apparent to observers

- In some cases we may have opportunities to lessen this concern by withholding certain details of the strategy from the protocol, and placing them in another document of more limited circulation.

- For example, if some type of selection is to be made based upon predictive probabilities, do full details and thresholds need to be described in the protocol?
Information apparent to observers

A proposal:

- **Selection decisions** (choice of dose, subgroup, etc. for continuation) generally do NOT give away an amount of information that would be considered to compromise or influence the trial, as long as the specific numerical results on which the decisions were based remain confidential.

Consider the alternative -

- In a seamless Phase II / III design, we might instead have run a conventional separate-phase program.

- Phase II results would be widely known (what about *equipoise*??)

- In this sense, maybe the adaptive design offers a further advantage relative to the traditional paradigm?
Algorithmic changes

- More problematic - changes based in an algorithmic manner on interim treatment effect estimates in effect provide knowledge of those estimates to anyone who knows the algorithm and the change.

- Most typical example - certain approaches to sample size re-estimation:
  - $SS_{new} = f$ (interim treatment effect estimate)
  - $\Rightarrow$ estimate $= f^{-1}(SS_{new})$

Mitigating the concerns

- Perhaps the adaptation can be made based upon a combination of factors in order to mask the observed treatment effect.
  - e.g., SS re-estimation using the treatment effect, the observed variance, and external information.
- If possible, “discretize” the potential actions, i.e., a small number of potential actions correspond to ranges of the treatment effects.
- We may at times try to quantify that the knowledge which can be inferred is comparable to that of accepted group sequential plans.
Summary

- Adaptive designs suggest real benefits for the clinical development process.
- Achieving this promise will require full investigation and understanding of the relevant issues, trade-offs, and challenges.
- Advantages should be considered in balance against any perceived risks or complexities.
- This should be expected to require more planning, not less.
- We can expect that adaptive designs will inevitably be scrutinized closely because of their novelty.

Summary

- We should not aim to broadly undo established monitoring conventions, but rather to fine-tune them to achieve their sound underlying principles.
- To justify sponsor participation in monitoring, provide convincing rationale and “minimize” this involvement, and enforce strict control of information.
- Some types of adaptations convey limited information for which it seems difficult to envision how the trial might be compromised.
- Others convey more information, but perhaps we can implement extra steps to mask this.
References


