2025

DEMING CONFERENCE

Atlantic City, NJ
December 10, 2025

Tutorial on ICH E20 Guidance on Adaptive Designs in Clinical Trials

Vlad Dragalin
Quantitative Sciences: Research & Consulting

Johnson&Johnson
Innovative Medicine

December 10, 2025

Tutorial on ICH E20 Guidance on Adaptive Designs in Clinical Trials

Christopher Jennison

Department of Mathematical Sciences,
University of Bath, UK
http://people.bath.ac.uk/mascj

Bruce Turnbull

Department of Statistical Science,
Cornell University, UK
https://people.orie.cornell.edu/bruce

Deming Conference, Atlantic City

December 2025

Agenda

01 **Introduction & Scope**

02 **Advantages & Challenges**

03 **Key Principles**

04 **Types of Adaptations**

05 **Special Considerations**

06 **Documentations**

07 **Types of adaptive designs**

Methods and Challenges

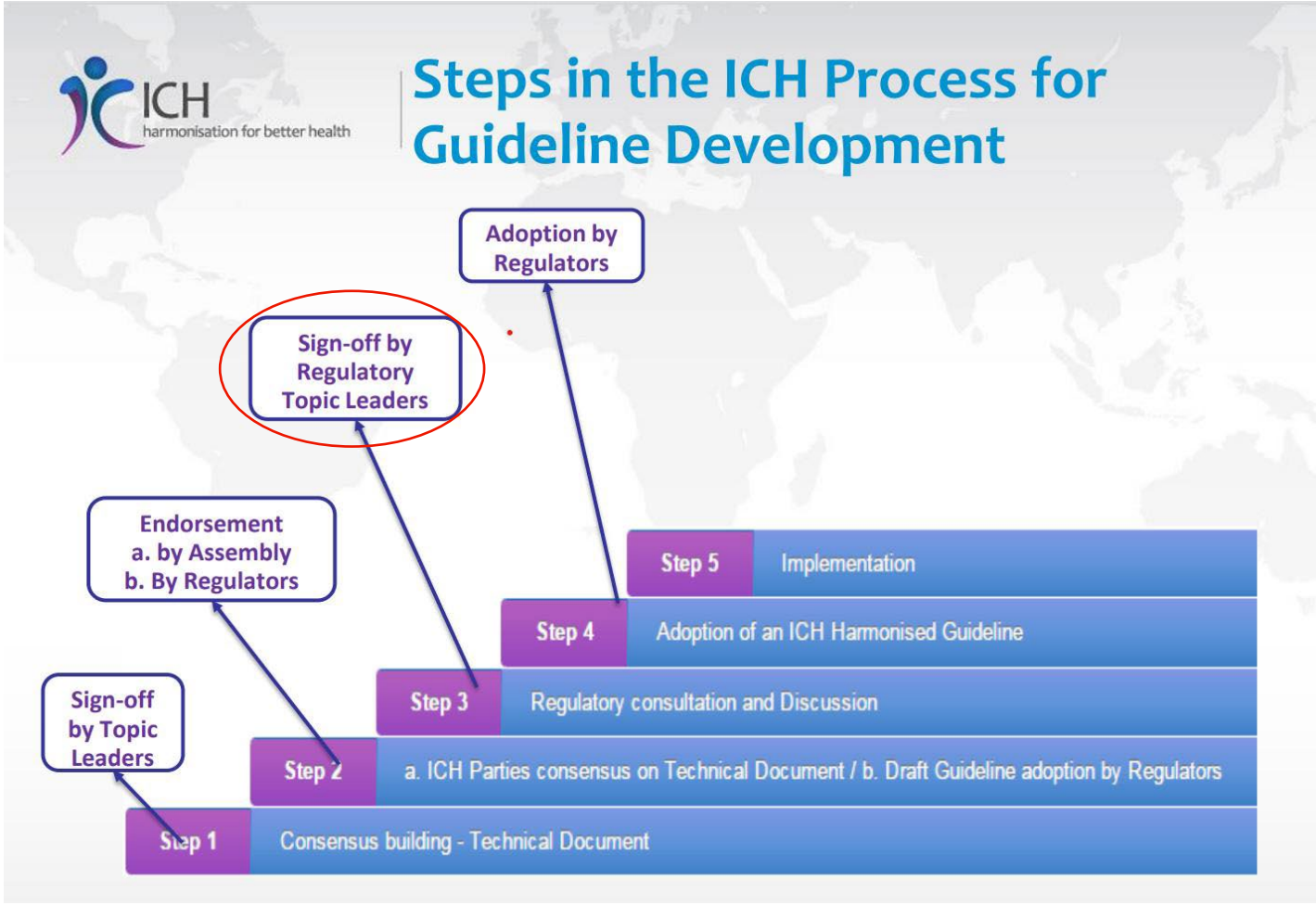
08 **Unbiased Estimation**

09 **Data Monitoring Committee**

10 **Adaptive Designs Using Bayesian Methods**

ICH Scope and Process for Guideline Development

- **Founded in 1990, ICH seeks to standardize regulatory requirements for pharmaceutical products worldwide. It was originally formed by members from Europe, Japan, and the US.**
- **A new topic is proposed by an ICH Member or Observer for approval by the ICH Assembly.**
- **An informal Working Group is formed to create a Concept Paper, offering additional context and outlining the objectives.**
- **An Expert Working Group (EWG) or Implementation Working Group is then established to develop a comprehensive Work Plan, setting milestones and deadlines.**



ICH E20 Draft Guideline

Adaptive Designs for Clinical Trials

- Development of ICH20 guidelines occurs through a **transparent standardized operating procedure**
- The ICH E20 is developed based on a **Concept Paper** (Nov 2019)
- The **Expert Working Group** (EWG) was established in November 2019, initiating guideline drafting through a small writing team and multiple specialized subteams. The writing process involved extensive drafting, review, and revision over several years.
- The Draft document was signed off as a Step 2b **'Draft'** document in June 2025 to be issued by the ICH **Regulatory** Members for public consultation.
- Focus is on confirmatory trials with an adaptive design.

ICH version availability online: https://www.ich.org/page/efficacy-guidelines#19-1

ICH E20 Expert Working Group

- **Rapporteur:** Z. John Zhong (PhRMA)
- **Regulatory Chair:** Gregory Levin (FDA)

- **PhRMA:** Amy Xia & Vlad Dragalin
- **EFPIA:** Bruno Flamion & Hans Ulrich Burger
- **BIO:** Frank Bretz & Erik Pulkstenis
- **JPMA:** Hideki Suganami & Masayo Miyata
- **IGBA:** Kevinkumar Kansagra
- **IFPMA:** Xiaoni Liu & Zhihong Lu

- **FDA:** John Scott
- **EC:** Armin Koch & Frank Petavy
- **MHLW/PMDA:** Yuki Ando & Naoto Kotani
- **Health Canada:** Roxana Alexa & Catherine Njue
- **Swissmedic:** Lorenzo Hess & Verena Gafner
- **ANVISA:** Carolina Pingret Cintra & Leonardo Fabio Costa Filho
- **HSA:** Lisa Tan & Tan Hui Xing
- **MFDS:** Myung Ah Chung
- **NMPA:** Jianhong Pan & Yunhong Huang

- **TFDA:** Lien-Cheng Chang & Wei-Lun Peng
- **GHC:** Turki Althunian

ICH Assembly Meeting in Singapore, Nov 17-20, 2019

Call for Comments

INTERNATIONAL COUNCIL FOR HARMONISATION OF TECHNICAL REQUIREMENTS FOR PHARMACEUTICALS FOR HUMAN USE

ICH HARMONISED GUIDELINE

ADAPTIVE DESIGNS FOR CLINICAL TRIALS
E20

Draft version

Endorsed on 25 June 2025

Currently under public consultation

At Step 2 of the ICH Process, a consensus draft text or guideline, agreed by the appropriate ICH Expert Working Group, is transmitted by the ICH Assembly to the regulatory authorities of the ICH regions for internal and external consultation, according to national or regional procedures.

Explanatory Note for ICH E20: The draft guideline makes statements on adaptive design approaches for clinical trials. The draft guideline acknowledges the high potential for adaptive designs to accelerate the process of drug development and to allocate resources more efficiently without lowering scientific and regulatory standards. Some of the approaches may affect the nature and timing of interactions between industry and regulators at confirmatory trial planning and assessment. The final guideline will indicate key adaptive design principles and approaches for which discussion of adaptive design features, and the rationale for their use, are particularly critical at the planning stage. To inform guideline finalization, specific feedback is sought on adaptive design principles and approaches and their impact on industry-regulatory interactions. Until a final guideline is agreed under Step 5 of the ICH process, the draft guideline should not be understood as confirming full regulatory acceptance from ICH parties of its contents, nor superseding current regional guidance, which remains valid. Public consultation comments on the draft guideline are sought.

critical at the planning stage. To inform guideline finalization, specific feedback is sought on adaptive design principles and approaches and their impact on industry-regulatory interactions. Until a final guideline is agreed under Step 5 of the ICH process, the draft guideline should not be understood as confirming full regulatory acceptance from ICH parties of its contents, nor superseding current regional guidance, which remains valid. Public consultation comments on the draft guideline are sought.

ICH HARMONISED GUIDELINE

ADAPTIVE DESIGNS FOR CLINICAL TRIALS

E20

ICH Consensus Guideline

TABLE OF CONTENTS

Definitions of adaptive design

Adaptive Design is one that uses *accumulating data from the ongoing trial* to modify aspects of the study without undermining the validity and integrity of the trial

- PhRMA ADWG (2006)

An adaptive design is defined as a clinical trial design that allows for prospectively planned modifications to one or more aspects of the trial based on interim analysis of accumulating data from participants in the trial

– ICH E20 (2025)

Adaptive design is defined as a clinical trial design that allows for *prospectively planned modifications* to one or more aspects of the design *based on accumulating data from subjects in the trial*

– FDA Guidance on AD (2019)

A clinical trial design that will have adaptations based on the accumulating data from the trial and/or *external data*. Modifications based on the accumulating data from the trial should be pre-specified prior to initiation of the trial

– Draft NMPA (2019)

A study design is adaptive if statistical methodology allows the modification of a design element (e.g. sample-size, randomization ratio, number of treatment arms) at an IA *with full control of the type I error*

– EMA reflection paper (2007)

1. Introduction and Scope

Guidance on confirmatory clinical trials with an adaptive design intended to evaluate a treatment for a given medical condition within the context of its overall development program.

Out of scope:

- Trials with unplanned modifications to the design
- Design changes based entirely on emerging information from a source external to the trial
- Routine monitoring of operational aspects

Focus on principles for the planning, conduct, analysis, and interpretation of trials with an adaptive design intended to confirm the efficacy and support the benefit-risk assessment of a treatment

2. Advantages and Challenges

Advantages:

- **Offer ethical advantages by potentially minimizing the number of participants subjected to less effective treatments through early termination of trials.**
- **Improve trial efficiency by boosting statistical power with the same sample size.**
- **Enhance comprehension of treatment outcomes and support informed decisions, including optimal dose selection and validation of effectiveness.**

Challenges:

- **Greater logistical complexities and heightened risks to trial integrity require more intricate planning and evaluation.**
- **Early termination for efficacy may introduce bias in treatment effect estimates and limit safety data, complicating benefit-risk evaluations.**
- **This approach may not be appropriate in all contexts, particularly when rapid enrollment or limited data availability impede reliable interim modifications.**
- **Complex adaptations can introduce additional uncertainty in trial outcomes and pose difficulties for regulatory decision-making.**
- **It is essential to provide comprehensive justification and assess the design's benefits, limitations, and its effects on trial integrity and interpretability.**

3. Key Principles

Adequacy within the development program:	Justifying the selected dose, etc
Adequacy of trial planning:	Pre-planned, as simple as possible, some flexibility
Limiting the chances of erroneous conclusions	Type I error control
Reliability of estimation	Estimates and confidence intervals for cost-benefit decisions
Maintenance of trial integrity	Blinding, avoiding information leakage, role of IDMCs.

"*All of these principles should be followed regardless of the type of adaptation and statistical approach (e.g., frequentist or Bayesian methods).*"

4. Types of Adaptations

1. Early Trial Stopping

- Allows for sequential analyses to stop trials early for efficacy or futility, using predefined boundaries to control Type I error.
- Early stopping may limit safety data and secondary endpoint information

2. Sample Size Adaptation

- Adjusts initial sample size based on interim estimates of nuisance parameters or treatment effects to ensure adequate power.
- Requires pre-specified rules, use of blinded data, and methods to control Type I error; bias in effect estimates should be evaluated and mitigated.

3. Population Selection

- Enables interim decisions to focus on specific subpopulations to improve trial efficiency and relevance.
- Needs thorough planning, justification, and statistical methods to control Type I error; bias in effect estimates should be addressed.

4. Types of Adaptations

4. Treatment Selection

- Uses interim data to select the most promising treatment doses or options for continued evaluation.
- Requires detailed planning, pre-specified rules, and methods to control Type I error; bias in effect estimates should be considered and corrected.

5. Adaptation to Participant Allocation

- Implements response-adaptive randomization (RAR), assigning more participants to better-performing treatments, potentially reducing exposure to inferior options.
- Challenges include bias, confounding from time trends, and ensuring valid statistical inference; deterministic adaptations are discouraged due to high bias risk.

5. Special Topics and Considerations

1. Further Considerations on Data Monitoring

- An IDMC should include expertise in interim monitoring, with access to unblinded efficacy and safety data, and operate under a detailed charter.
- An independent statistical group should conduct analyses and produce reports, with strict confidentiality and sole access to unblinded data to protect trial integrity.
- Sponsor access to unblinded interim results should be minimized and justified; any access should follow strict confidentiality protocols and be transparently documented.

2. Planning, Conducting, and Reporting Simulation Studies

- Simulations help evaluate operating characteristics of adaptive designs under various scenarios.
- Clear objectives, a broad range of design options, and justified assumptions are essential for meaningful simulations.
- Results must be comprehensively documented in a report, including key questions, design evaluations, assumptions, and limitations, to support regulatory review.

5. Special Topics and Considerations

3. Adaptive Designs Using Bayesian Methods

- Reinforces the ICH E9 principle emphasizing clear rationale and reliable conclusions
- Defines “Bayesian” as any approach combining prior information with study data to generate a posterior distribution
- Begins with an overview of Bayesian methods to guide trial adaptations, incorporating decision criteria to control the Type I error rate
- Sponsors must justify that the overall design achieves the intended operating characteristics
- Sponsors are expected to discuss and document the relevance of external data to the trial design, listing all potentially relevant sources and explaining reasons for excluding any

5. Special Topics and Considerations

4. Adaptive Designs in Time-to-Event Settings

- In time-to-event trials, the focus is on the number of events rather than the number of participants, often leading to interim analyses based on number of events, with possible adjustments to the number of events or follow-up time.
- Adaptive designs should ensure sufficient data for benefit-risk assessments, especially when increasing the number of participants or follow-up duration to observe more events.
- Maintaining independence between data collected before and after interim analyses is crucial; using participant data that contribute to both stages can inflate Type I error.
- Strategies to control error include pre-specified adaptation rules based only on the primary endpoint, defining participant sets for each stage, or planning early stopping options based on event counts.
- Similar considerations apply to longitudinal outcomes, where using interim surrogate or intermediate outcomes requires careful analysis methods to prevent increased Type I error.

5. Special Topics and Considerations

5. Adaptive Designs in Exploratory Trials

- The guideline emphasizes applying key principles to adaptive designs in confirmatory trials to ensure reliable evidence for benefit-risk assessment
- Adaptive designs can also be used early in development for dose, regimen, population, or endpoint decisions, but principles still apply to maintain interpretability
- Exploratory trials may allow general adaptation principles rather than strict rules, but they must still provide a solid basis for subsequent confirmatory phases
- Sponsors should balance involvement in interim decisions with maintaining trial integrity, ensuring participant safety and minimizing bias

5. Special Topics and Considerations

6. Operational Considerations

- Operational challenges of adaptive designs, such as maintaining trial integrity, should be addressed during trial planning, including measures to limit inference from interim analyses.
- Informed consent forms must explain the possibility of adaptive changes, their purpose, and the continued protection of participants' rights and safety.
- Data management systems, like interactive randomization platforms, should be fully integrated and capable of handling scenario changes with minimal sponsor involvement.
- Drug supply logistics can be strained by rapid adaptations, especially across multiple countries, necessitating careful planning and simulation to support supply chain decisions.
- Processes for timely data validation and cleaning, including formal interim database locks, are essential to ensure high-quality data for adaptation decisions

6. Documentations

A rationale for the proposed adaptive design
A description of the adaptations being proposed
A description of the statistical analysis methods
A description of how the adaptive design will be implemented
A description of steps to maintain confidentiality of interim results and protect trial integrity, among other details of the operational execution
A description of important operating characteristics of the design

Types of adaptive design, methods and challenges

There are many research papers and several books that describe methods which can meet some of the ICH E20 guidelines.

See the recently published:

Group Sequential and Adaptive Methods for Clinical Trials,

Jennison & Turnbull, CRC Press, December 2025

(hereafter JT) and references therein — and our short course.

Existing methods may have to be applied in particular ways to satisfy the E20 principles.

Some methods many need further development to satisfy E20 principles.

1. Group sequential tests

Objective

Facilitate early stopping for efficacy or futility

Methods

To protect the type I error rate

Error spending tests

Computations for inference on termination

One-sided error spending tests

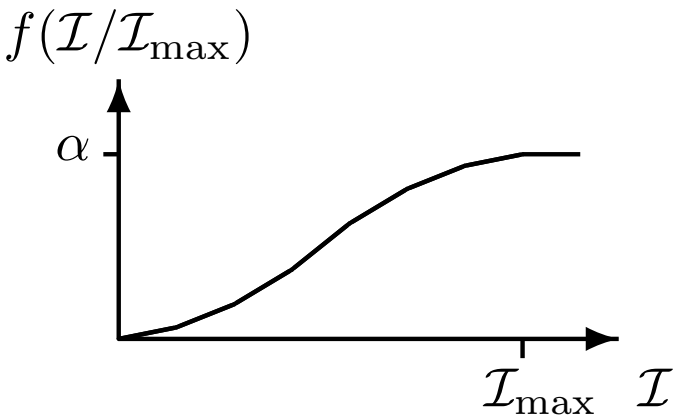
For treatment effect θ, we test H_0: $\theta \leq 0$ against $\theta > 0$ with

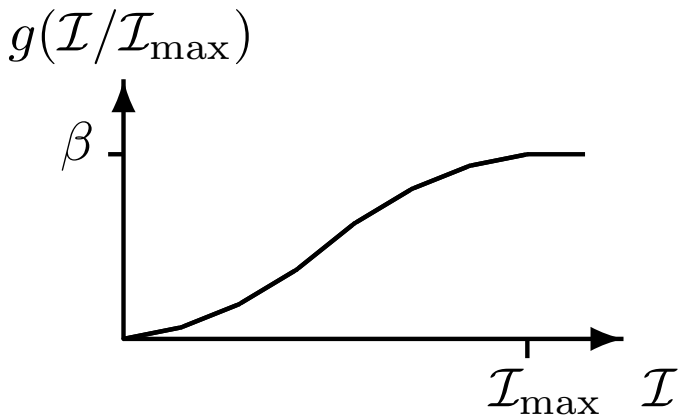
Type I error probability α at $\theta = 0$,

Type II error probability β at $\theta = \delta$.

Let $\mathcal{I}_k = \{\text{Var}(\widehat{\theta}_k)\}^{-1}$ where $\widehat{\theta}_k$ is the estimate of θ at analysis k.

We specify two error spending functions





Type I error probability α is spent according to the function $f(\mathcal{I}/\mathcal{I}_{\max})$, and type II error probability β according to $g(\mathcal{I}/\mathcal{I}_{\max})$.

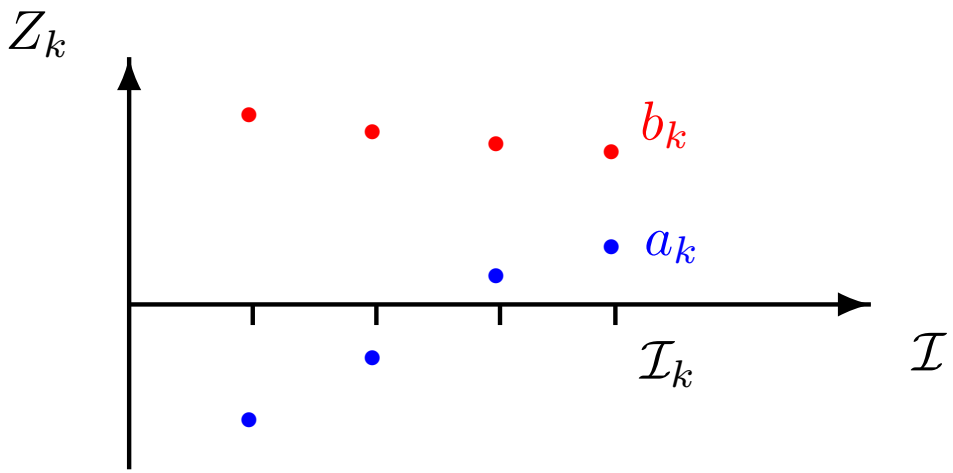
One-sided error spending tests

Analysis k: With observed information $\mathcal{I}_k$, we find a_k and b_k to satisfy

$$P_{\theta=0}\{a_1 < Z_1 < b_1, \ldots, a_{k-1} < Z_{k-1} < b_{k-1}, Z_k > b_k\}$$
$$= f(\mathcal{I}_k/\mathcal{I}_{\max}) - f(\mathcal{I}_{k-1}/\mathcal{I}_{\max}),$$

and

$$P_{\theta=\delta}\{a_1 < Z_1 < b_1, \ldots, a_{k-1} < Z_{k-1} < b_{k-1}, Z_k < a_k\}$$
$$= g(\mathcal{I}_k/\mathcal{I}_{\max}) - g(\mathcal{I}_{k-1}/\mathcal{I}_{\max}).$$



Group sequential tests

Challenges

Early stopping for efficacy should only happen when there is:

- Adequate safety data
- The information needed to make cost-benefit decisions

Avoiding bias in estimates of treatment effects

2. Sample size adaptation

Objective

Respond to interim estimates of:

- Nuisance parameters
- The treatment effect on the primary endpoint

Methods

Combination tests

"Conditional probability of rejection" principle

Combination tests

Before the trial commences, define the null hypothesis.

Let θ denote the treatment effect *vs* control for a specified form of the treatment, patient population and endpoint.

Suppose we wish to test H_0: $\theta \leq 0$ against $\theta > 0$, with type I error rate α at $\theta = 0$ when sample size may be re-assessed after Stage 1.

Define one-sided P-values $P^{(1)}$ and $P^{(2)}$ from hypothesis tests of H_0 based on Stage 1 and Stage 2 data, respectively.

Then, under $\theta = 0$

$P^{(1)} \sim U(0,1)$.

Conditionally on Stage 1 data and the Stage 2 design (informed by Stage 1 data), $P^{(2)} \sim U(0,1)$.

Hence, if $\theta = 0$, $P^{(1)}$ and $P^{(2)}$ are independent $U(0,1)$ variates.

The inverse normal combination test

Initial design

Specify the **inverse normal test** for null hypothesis H_0, with weights w_1 and w_2 where $w_1^2 + w_2^2 = 1$.

Design Stage 1, fixing sample size and test statistic.

Stage 1

Observe the one-sided P-value, $P^{(1)}$, based on Stage 1 data.

Compute $Z^{(1)} = \Phi^{-1}(1 - P^{(1)})$.

Design Stage 2 in the light of Stage 1 data.

Stage 2

Observe the P-value, $P^{(2)}$, based **only** on Stage 2 data.

Compute $Z^{(2)} = \Phi^{-1}(1 - P^{(2)})$.

The inverse normal combination test

Under $\theta = 0$

We know $P^{(1)} \sim U(0,1)$ and $P^{(2)} \sim U(0,1)$ are independent.

Hence $Z^{(1)} \sim N(0,1)$ and $Z^{(2)} \sim N(0,1)$ are independent and

$$w_1 Z^{(1)} + w_2 Z^{(2)} \sim N(0,1).$$

For a one-sided test with type I error rate α, we reject H_0 if

$$w_1 Z^{(1)} + w_2 Z^{(2)} > \Phi^{-1}(1-\alpha).$$

If $\theta < 0$, then $Z^{(1)}$ and $Z^{(2)}$ are stochastically smaller than $N(0,1)$ random variables and the type I error rate is less than α.

Here, it is crucial that w_1 and w_2 are pre-specified and not changed in response to observed data.

Sample size adaptation

Challenges

Trial integrity: blinding, information leakage

Avoiding bias in estimates of treatment effects

3. Population selection

Objective

Focus on the sub-population in which a new treatment is most effective, "enriching" the sample size in that sub-population

Methods

Combination tests and a **closed testing procedure** (CTP) to control the family-wise error rate (FWER)

An effective (possibly Bayes) rule to decide when to enrich

Testing multiple hypotheses: the family-wise error rate

In an enrichment trial, we may test for a treatment effect in the full population and various sub-populations.

Adaptations are to drop certain sub-populations and concentrate on subjects with the best response to the new treatment.

In analysing the data, we wish to control the overall probability of a false positive conclusion.

The family-wise error rate

Suppose we have h null hypotheses, H_i: $\theta_i \leq 0$ for $i = 1, \ldots, h$.

A procedure's **family-wise error rate** when $\boldsymbol{\theta} = (\theta_1, \ldots, \theta_h)$ is

$$P_{\boldsymbol{\theta}}\{\text{Reject } H_i \text{ for some } i \text{ with } \theta_i \leq 0\}.$$

The family-wise error rate is controlled **strongly** at level α if this error rate is at most α for all possible combinations of θ_i values, so

$$P_{\boldsymbol{\theta}}\{\text{Reject any true } H_i\} \leq \alpha \quad \text{for all } \boldsymbol{\theta} = (\theta_1, \ldots, \theta_h).$$

Controlling family-wise error: closed testing procedures

Marcus et al. (*Biometrika*, 1976) introduced a **closed testing procedure** which provides strong control of FWER by combining level α tests of each H_i and of intersections of these hypotheses.

Suppose we have null hypotheses H_i, $i = 1, \ldots, h$.

For each subset I of $\{1, \ldots, h\}$, define the intersection hypothesis

$$H_I = \cap_{i \in I} H_i.$$

— a simple hypothesis H_j is a special case where $I = \{j\}$.

Construct a level α test of each intersection hypothesis H_I, i.e., a test which rejects H_I with probability at most α whenever all hypotheses specified in H_I are true.

Closed testing procedure

The simple hypothesis H_j: $\theta_j \leq 0$ is rejected overall if, and only if, H_I is rejected for every set I containing index j.

Closed testing procedures

Proof of strong control of family-wise error rate

In the closed testing procedure, overall rejection of the simple hypothesis H_j can only occur if H_I is rejected for every set I containing index j.

Let $\tilde{I}$ be the set of indices of all true hypotheses H_i.

Since $H_{\tilde{I}}$ is true, $P\{\text{Reject } H_{\tilde{I}}\} = \alpha$.

For a family-wise error to be committed, $H_{\tilde{I}}$ must be rejected.

Hence, the probability of a family-wise error is no greater than α.

Using combination tests and a closed testing procedure

Suppose an enrichment trial is conducted in two stages and adaptation may occur at the end of Stage 1.

Initially, there are h populations or sub-populations for which a null hypothesis of no treatment effect may be tested.

We need to define a level α test for each intersection hypothesis

$$H_I = \cap_{i \in I} H_i$$

In a two-stage adaptive trial, each stage provides a P-value for H_I, $P_I^{(1)}$ in Stage 1 and $P_I^{(2)}$ in Stage 2.

The way in which the $P_I^{(1)}$ are to be calculated is specified at the start of the trial and the way in which the $P_I^{(2)}$ are to be calculated must be stated before commencing Stage 2.

We combine these P-values across stages by a pre-specified method, e.g., an inverse normal combination test.

Population selection

Challenges

- Justifying the choice of sub-population
- Adequate information on (lack of) treatment efficacy in the complementary population
- Avoiding bias in estimates of treatment effects

4(a) Treatment selection: Seamless Phase 2/3 trials

Objective

Use interim data to choose a dose to take to the Phase 3 stage, then test for a difference between the selected treatment and control using Phase 2 and Phase 3 data

Methods

Analyse Phase 2 and Phase 3 data with combination tests and a closed testing procedure (CTP) to control FWER

Dunnett tests for intersection hypotheses in the CTP

Treatment selection: Seamless Phase 2/3 trials

Suppose the Phase 2 stage of the study has treatments $i = 1, \ldots, p$ with treatment effects $\theta_1, \ldots, \theta_p$ when compared with the control.

We shall select one treatment, i^* say, to proceed to the Phase 3 stage and test H_{i^*}: $\theta_{i^*} \leq 0$.

With a different choice after Phase 2, we could have tested any one of the hypotheses H_i: $\theta_i \leq 0$, $i = 1, \ldots, p$ at the end of the trial.

We use a closed testing procedure to protect the family-wise error rate for this set of hypotheses.

The level α test for intersection hypothesis H_I will be based on stage-wise P-values $P_I^{(1)}$ from the Phase 2 stage and $P_I^{(2)}$ from the Phase 3 stage.

Then, $P_I^{(1)}$ and $P_I^{(2)}$ will be combined using, say, an inverse normal combination test.

Testing an intersection hypothesis

Suppose the intersection hypothesis $H_I = \cap_{i \in I} H_i$ is the intersection of m simple hypotheses.

For each $i \in I$, let P_i be the 1-sided P-value for testing H_i.

Denote the ordered values of the P_i by $P_{[1]} \leq P_{[2]} \leq \ldots \leq P_{[m]}$.

Bonferroni adjustment

The overall P-value for testing H_I is $P_I = m\, P_{[1]}$.

Simes' method (Biometrika, 1986)

The overall P-value for H_I is

$$P_I = \min_{k=1,\ldots,m} (m\, P_{[k]}/k).$$

The Simes method is valid — usually a little conservative — when the P_i are independent or positively dependent.

Dunnett's method (JASA, 1955)

Suppose m treatments are compared with a control, responses are normal with known variance, and sample sizes on each treatment and the control are equal.

Each null hypothesis H_i says treatment i is no better than control.

We are to test the intersection hypothesis $H_I = \cap_{i \in I} H_i$.

Denote the Z-statistic arising from the test of H_i by Z_i.

When each treatment effect for an $H_i \in H_I$ is zero,

$$Z_i \sim N(0,1), \quad i \in I, \quad \text{Cov}(Z_i, Z_{i'}) = 0.5, \quad i \neq i'.$$

The P-value for testing H_I using Dunnett's test is

$$P\{\max_{i \in I} Z_i > z^*\},$$

where z^* is the observed value of $\max_{i \in I} Z_i$, and the probability is under the above multivariate normal distribution for $\{Z_i,\, i \in I\}$.

Treatment selection: Seamless Phase 2/3 trials

Challenges

- Operational aspects of combining phases of testing a new drug
  - Handing over decision making to the IDMC
  - Pre-specifying details of the Phase 3 stage
- Avoiding bias in estimates of treatment effects

4(b) Treatment selection: Multi-arm multi-stage trials

Objective

Use interim data to focus on the most promising treatments

Methods

Combination tests and a closed testing procedure (CTP) to control the family-wise error rate

Dunnett tests for intersection hypotheses in the CTP

Using combination tests in a closed testing procedure

When comparing h treatments with the control, we have null hypotheses H_i, $i = 1, \ldots, h$.

We wish to test these in a closed testing procedure with FWER α.

We need to define a level α test for each intersection hypothesis

$$H_I = \cap_{i \in I} H_i.$$

In a multi-stage trial, treatments can be dropped at a sequence of interim analyses.

Before each new stage we specify how the P-value for each (relevant) intersection hypothesis H_I will be calculated.

We combine these P-values by a pre-specified method, e.g., a multi-stage combination test when there are more than 2 analyses at which stopping may occur.

Treatment selection: Multi-arm multi-stage trials

Challenges

Operational aspects

- Handing over decision making to the IDMC
- Pre-specifying rules for dropping treatment arms

Avoiding bias in estimates of treatment effects — when the sample space is complex and, with flexible decision making, not clearly defined.

5. Trials with response adaptive randomisation (RAR)

Objective

Use interim data to identify better-performing treatments and allocate more patients to these treatments

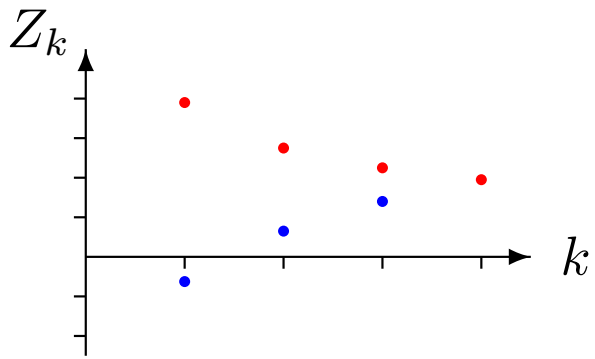
Methods

- Methods for "bandit" problems
- RAR can be incorporated neatly in a group sequential test (JT book, Chapter 30)

Challenges

- Confounding from time trends
- Maintaining uncertainty in treatment allocation
- Valid statistical inference on completion of the trial

Adaptations 1 to 5: Estimation after an adaptive trial



In a two-treatment comparison, the maximum likelihood estimate (MLE) of θ when a group sequential trial stops at analysis k is

$$\widehat{\theta}_M \;=\; \bar{X}_{Ak} \;-\; \bar{X}_{Bk}.$$

For large, positive values of θ:

high values of $\widehat{\theta}$ lead to early stopping,

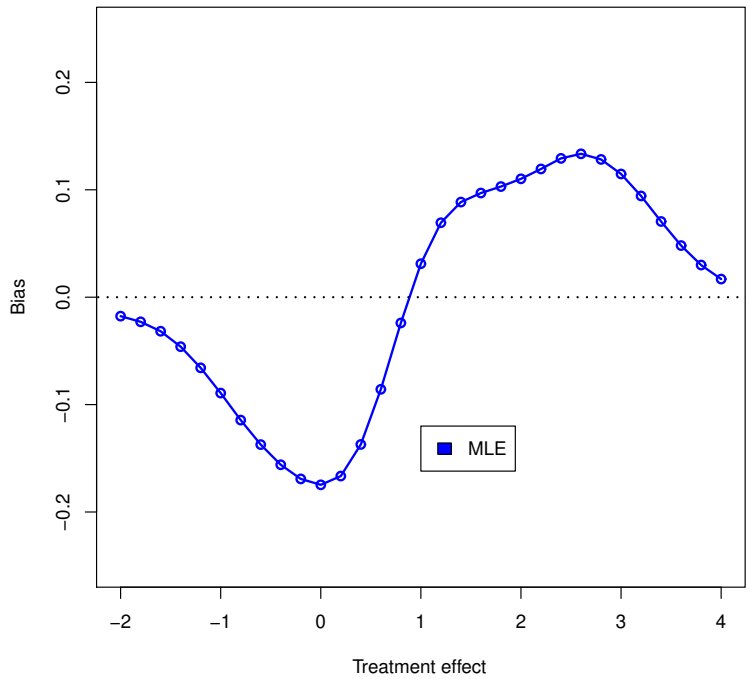
lower values of $\widehat{\theta}$ result in more observations, so $\widehat{\theta}$ can increase.

Thus, the MLE is biased with $E_\theta(\widehat{\theta}_M) > \theta$ for high values of θ and $E_\theta(\widehat{\theta}_M) < \theta$ for low values of θ.

Bias of the MLE of θ after a Pampallona & Tsiatis test

Consider the Pampallona & Tsiatis GST with 4 analyses, $\Delta = 0$, $\alpha = 0.025$ and power $1 - \beta = 0.8$ at $\theta = 1$ (see JT, Chapter 5)

The bias of the MLE can be calculated as a function of the true effect size, θ.



The bias of the MLE is around 0.1 at values of θ just above 1.

Correcting the bias of the MLE

Denote the bias function of the MLE by

$$b(\theta) = E_\theta(\widehat{\theta}_M) - \theta.$$

Whitehead (*Biometrika*, 1986) suggested correcting the MLE by subtracting an estimate of its bias.

Although the true θ is unknown, the bias of the MLE can be estimated by $b(\widehat{\theta}_M)$.

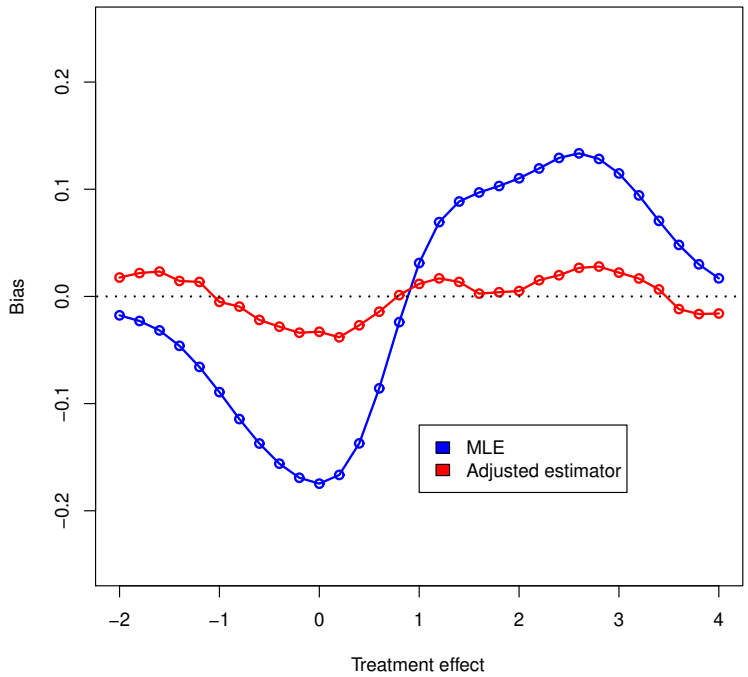
The adjusted estimator is then

$$\widehat{\theta}_{adj} = \widehat{\theta}_M - b(\widehat{\theta}_M).$$

Bias of the MLE of θ after a Pampallona & Tsiatis test

Simulation results show that Whitehead's adjusted estimator has much smaller bias than the MLE on which it is based.

For our example:



The adjustment almost completely removes the bias in the MLE.

Unbiased estimation after a group sequential test

ICH E20:

Key principle number 4 concerns "Reliability of estimation".

It is noted that, while controlling the chances of false positive efficacy conclusions is expected,

> *In addition, reliable estimation of treatment effects for the primary efficacy endpoint and other key efficacy and safety outcomes is important.*
>
> *In the trade-off between bias and variance, the expectation is generally for* **limited to no bias** *in the primary estimate of the treatment effect.*

It may be surprising to suppose it is possible to give an estimate with "no bias" after a group sequential or adaptive trial.

Unbiased estimation after a group sequential test

Emerson & Fleming (*Biometrika*, 1990) noted that $\widehat{\theta}_1$, the MLE based on the data at analysis 1, is unbiased for θ.

Applying "Rao-Blackwellization", we can calculate the conditional expectation of $\widehat{\theta}_1$ given the final set of data to obtain an unbiased estimate of θ.

This is the Uniform Minimum Variance Unbiased Estimate (UMVUE) among estimators that do not require knowledge of future information levels.

The numerical methods used to compute properties of a group sequential test can be adapted to compute this UMVUE.

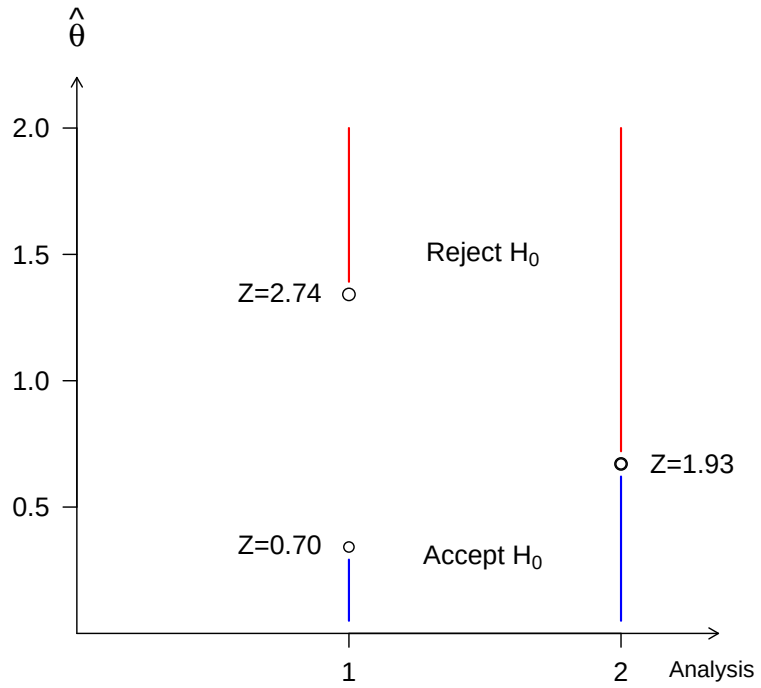
However, unbiased estimators may have a large variance and a relatively high mean square error.

The UMVUE may also be rather strange!

Unbiased estimation after a group sequential test

Clinical trial designs with just two analyses are common.

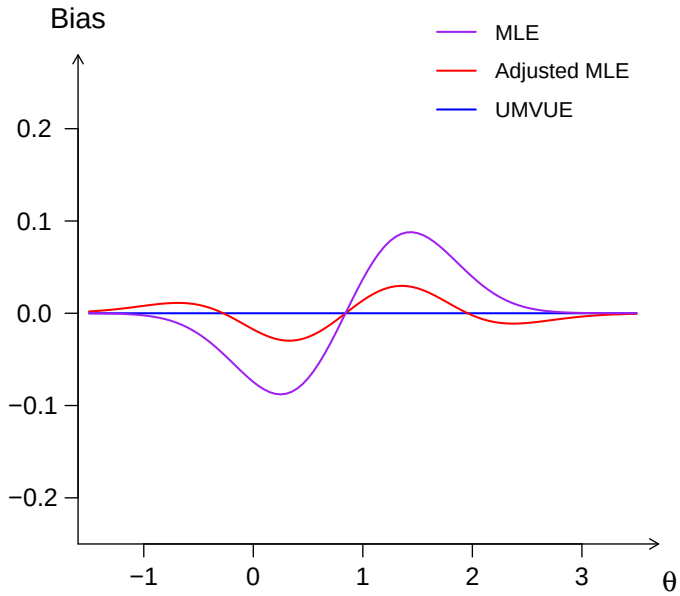
Consider the following one-sided, group sequential test of H_0: $\theta \leq 0$ against $\theta > 0$.



This test has type I error probability $\alpha = 0.025$ and power 0.8 is achieved if $\theta = 1$.

Unbiased estimation after a group sequential test

The bias of several estimates:



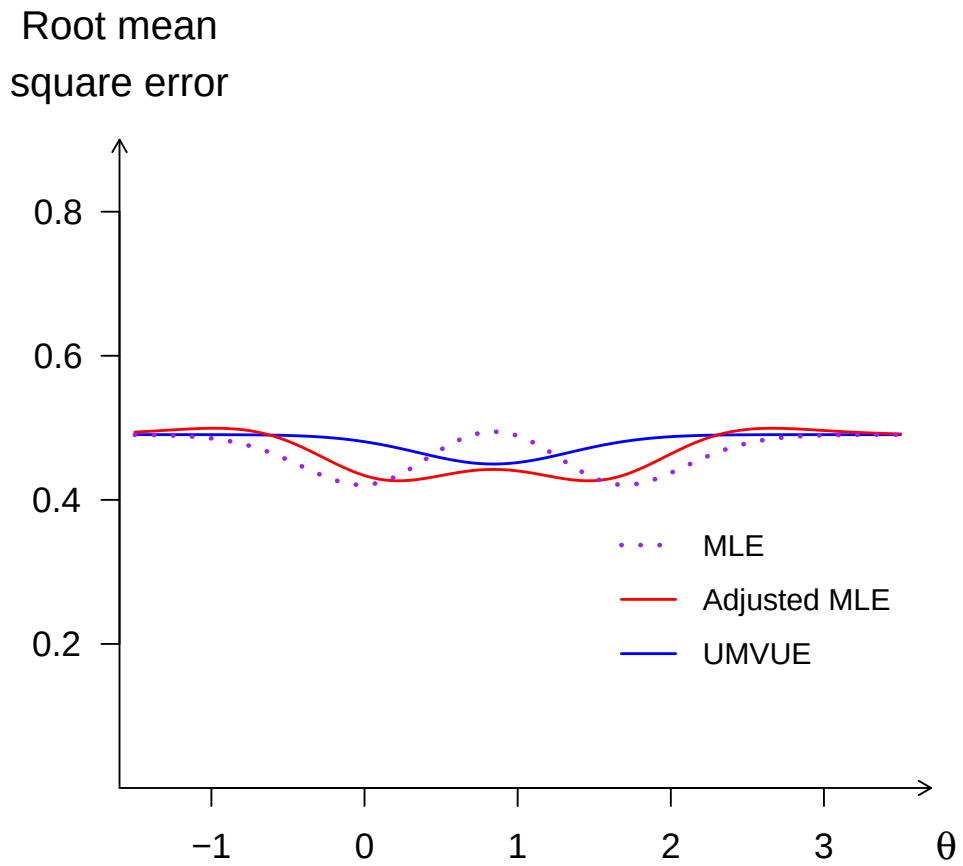
Denote the maximum likelihood estimate (MLE) on termination by $\widehat{\theta}_M$ and its bias function by $b(\widehat{\theta})$.

The Adjusted MLE is $\widehat{\theta}_{adj} = \widehat{\theta}_M - b(\widehat{\theta}_M)$.

The UMVUE is the conditional expectation of $\widehat{\theta}_1$ given the final set of data.

Unbiased estimation after a group sequential test

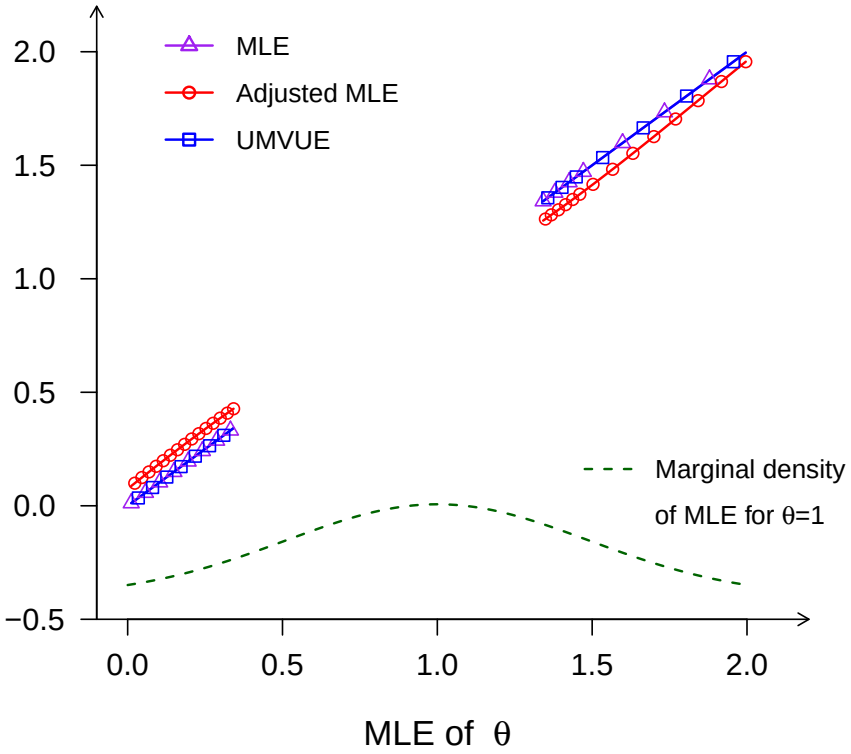
The root mean square error of several estimates:



The UMVUE has a higher variance than the Adjusted MLE, and this results in a higher mean square error.

Unbiased estimation after a group sequential test

Estimates on termination at analysis 1

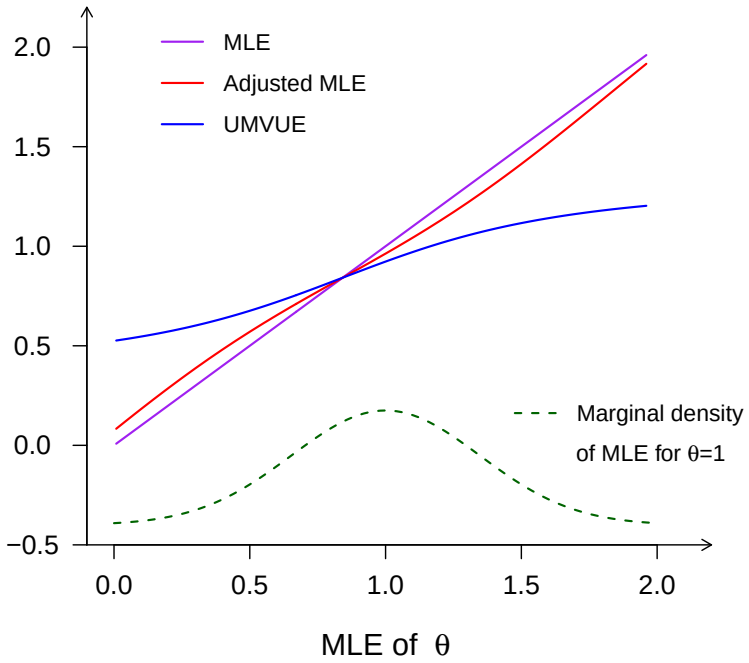


If the trial stops at analysis 1 the UMVUE is $\widehat{\theta}_1$, the MLE of θ.

So there is no "adjustment for bias" in the UMVUE!

Unbiased estimation after a group sequential test

Estimates on termination at analysis 2



When stopping at analysis 2, the UMVUE can be substantially lower than the MLE.

If the MLE is $\widehat{\theta}_2 = 1.5$, the UMVUE is only 1.12 — but the bias in the MLE is at most 0.09 for any value of θ.

Unbiased estimation after a group sequential test

In more complex adaptive designs, bias may arise

(i) from selecting a treatment arm or patient sub-population based on promising early results,

(ii) from early stopping on a "random high".

Some of the methods proposed for estimation after such trials also use Rao-Blackwellization to find UMVUE or Uniform Minimum Variance Conditionally Unbiased Estimates.

Given the behaviour of the UMVUE estimate in our simple example of a two-stage group single trial with a single parameter to estimate, we should look more closely at how these estimates may behave.

Adjusted estimates with a small bias may well be preferable.

The independent data monitoring committee (IDMC)

Some reflections of experience on IDMCs

Suggestions for sponsors

- Imagine you are an IDMC member
- Consider the expertise you need as the IDMC statistician
- After the kick off meeting, a firewall will be in place: what questions would you want to ask at this meeting?
- List the "what if" scenarios you would want to discuss

Remember: A typical company statistician may not have experience as an IDMC member.

Adaptive Designs Using Bayesian Methods

Borrowing external information

- “Incorporating external data to inform inference requires comprehensive scientific justification, including evaluation of alternative approaches that do not involve borrowing”
- When borrowing external information via Bayesian approaches, thorough scientific justification, careful selection of external data, and detailed documentation are essential to avoid bias and to control false positives.
- External data should ideally come from relevant, recent, and patient-level sources, with expert input crucial for evaluating their relevance and addressing potential conflicts with current trial data.

Adaptive Designs Using Bayesian Methods

Prior Distribution

- Proper pre-specification of the prior distribution, including the extent of data borrowing and success criteria, along with sensitivity analyses, helps ensure the reliability and robustness of trial conclusions.
- Sponsors should pre-specify and justify the degree of borrowing, success criteria, adequacy of trial data, and address potential conflicts between prior information and data
- Simulations should be conducted to assess the risk of incorrect conclusions, including false positives
- Sensitivity analyses to test the robustness of results relative to the choice of prior should be planned

Adaptive Designs Using Bayesian Methods

Success Criteria

- Bayesian methods can be used in adaptive trials when their application is well-justified, ensuring that decision criteria control Type I error and maintain robust operating characteristics.
- Carefully chosen success criteria are important to trial interpretability and efficiency
  - For frequentist trials, success criteria are almost always chosen to control familywise error rate at 0.025 one-sided
  - In some trials with Bayesian methodology, especially when borrowing information, this may not be applicable or appropriate
- Potential approaches to defining Bayesian success criteria:
  - Calibration to Type I error rate
  - Direct interpretation of posterior as probability statement
  - Criteria based on benefit-risk assessment or decision theory

Adaptive Designs Using Bayesian Methods

Operating Characteristics

- Always important to understand how a trial is likely to perform in terms of supporting correct conclusions and reliable estimation of treatment effects
- In frequentist inference, important characteristics include FWER and power, bias and MSE of effect estimates, and coverage probabilities
  - No different for Bayesian trials calibrated to frequentist characteristics
- For Bayesian trials not calibrated to Type I error, still important to understand what conclusions could be drawn under alternative potential prior distributions (design priors)

Adaptive Designs Using Bayesian Methods

Informative Priors

- The prior construction process should be designed, implemented, and documented in a systematic and transparent manner
- Sponsors should pre-specify and justify the full details of the proposed prior distribution in the protocol
  - Justification should address the appropriateness of the prior distribution's influence and the operating characteristics of the design
- Noninformative and minimally informative priors
- Skeptical priors
- Informative prior construction generally more complex
  - Depends on nature of data being used
  - Need a greater amount of justification
- Informative priors to borrow external information
  - Sponsors should provide strong justification that considers feasibility (e.g., of alternative approaches that do not involve borrowing) and the relevance of the available information
  - Important to consider the possibility of prior-data conflict
  - Identifying relevant information is a multidisciplinary effort

Adaptive Designs Using Bayesian Methods

Additional Considerations

- Identifying and reviewing available information
  - Relevant information may be clinical, PK, PD, non-clinical, RWD
  - Should consider data quality, study design, relevance, availability of patient-level data
- Prior construction
  - Variety of data sources will affect complexity of modeling
  - Consider whether exchangeability can be assumed
- Discounting: Static, dynamic
- Quantifying influence of the prior
- Sensitivity analyses
- Estimands and missing data
  - Look for alignment on estimands between data sources
- Software and computation
  - Reliable software, appropriate documentation and diagnostics
- Trial documentation
  - Describing Bayesian design in pre-study documents
  - Describing Bayesian results in study reports

Maintaining independence between data collected before and after interim analyses

Example: A study in oncology with treatment selection (GATSBY)

Experimental Treatment 1: Intensive dosing

Experimental Treatment 2: More frequent lower doses

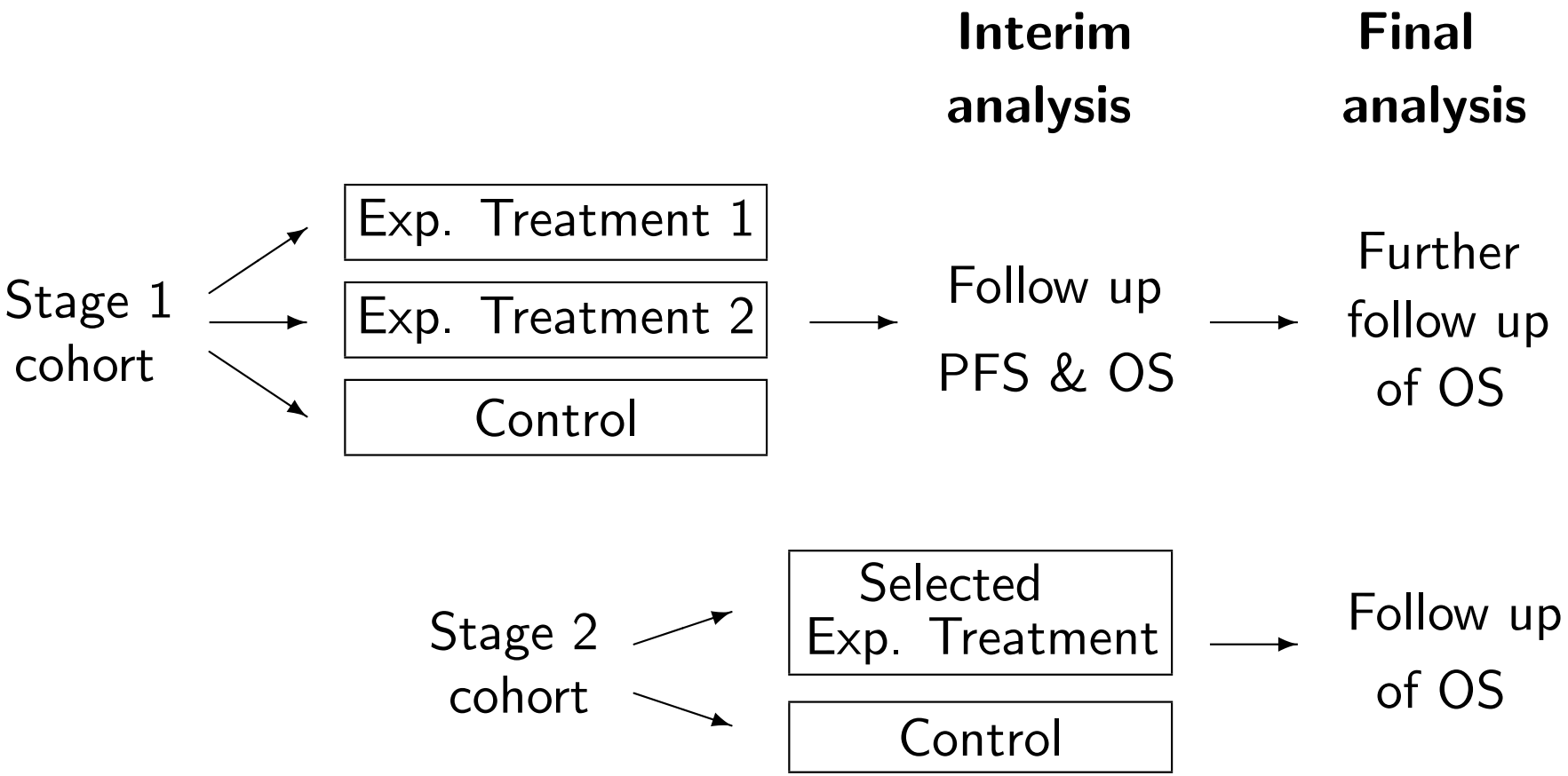
Control treatment

The primary endpoint is Overall Survival (OS).

At an interim analysis, information on OS, Progression Free Survival (PFS), PK measurements and safety will be used to choose between the two experimental treatments.

After the interim analysis, patients will only be recruited to the selected treatment and the control.

Overall plan of the trial



At the final analysis, we test the null hypothesis that OS on the selected treatment is no better than OS on the control treatment.

A combination test for survival data

We form logrank statistics to compare the selected treatment and the control.

Based on data at the interim analysis:

$$Z_1 = \frac{S_1}{\sqrt{\mathcal{I}_1}},$$

Based on data accrued **between** the interim and final analyses:

$$Z_2 = \frac{S_2 - S_1}{\sqrt{\mathcal{I}_2 - \mathcal{I}_1}}.$$

Standard theory or logrank statistics tells us that, if $\theta_1 = 0$, then $Z_1 \sim N(0,1)$ and $Z_2 \sim N(0,1)$ are independent.

So, we can use $Z = w_1 Z_1 + w_2 Z_2$ in an inverse normal combination test of $H_{0,1}$: $\theta_1 \leq 0$.

A combination test for survival data

The above distribution theory for logrank statistics of a single comparison requires

$$Z_2 \;=\; \frac{S_2 - S_1}{\sqrt{\mathcal{I}_2 - \mathcal{I}_1}} \;\sim\; N(0,1) \quad \text{under } \theta_1 = 0,$$

regardless of decisions taken at the interim analysis.

Bauer & Posch (*Statistics in Medicine*, 2004) note this implies that the conduct of the second part of the trial should not depend on the prognosis of Stage 1 patients at the interim analysis.

However, we select the better of the two experimental treatments based on good PFS results.

Hence, the prognoses for patients on the selected arm arm are liable to better than "average".

Jenkins, Stone & Jennison (*Pharm. Statist.*, 2011)

If we base a combination test on the two parts of the data accrued before and after the interim analysis, bias can result:

	Z_1	Z_2
Stage 1 cohort	Overall survival (during Stage 1)	**Overall survival (during Stage 2)**
Stage 2 cohort		Overall survival (during Stage 2)

Instead, we divide the data into the parts from the two cohorts:

Stage 1 cohort	Overall survival (during Stage 1)	Overall survival (during Stage 2)	Z_1
Stage 2 cohort		Overall survival (during Stage 2)	Z_2

Partitioning data for a combination test

To avoid bias: All patients in the Stage 1 cohort are followed for overall survival up to a fixed time, shortly before the final analysis.

"Stage 1" statistics are based on Stage 1 cohort's **final** OS data

$Z_{1,1}$ from log-rank test of Experimental Tr 1 vs Control

$Z_{1,2}$ from log-rank test of Experimental Tr 2 vs Control

$Z_{1,12}$ (for intersection hypothesis) from, say, a Dunnett test.

"Stage 2" statistics are based on OS data for the Stage 2 cohort

If Exp Treatment 1 is selected:

$Z_{2,1}$ from log-rank test of Exp Tr 1 vs Control, $Z_{2,12} = Z_{2,1}$

If Exp Treatment 2 is selected:

$Z_{2,2}$ from log-rank test of Exp Tr 2 vs Control, $Z_{2,12} = Z_{2,2}$.

Questions?

Thank you!