Unbiased estimation for adaptive designs

HRB-TMRN & MRC-NIHR-TMRP Webinar

May 2026

David Robertson

david.robertson@mrc-bsu.cam.ac.uk

MRC Biostatistics Unit

UNIVERSITY OF CAMBRIDGE

Outline

Introduction

Adaptive designs

- Adaptive clinical trials allow for preplanned opportunities to alter the course of the trial on the basis of accruing information
  - Increasing the recruitment target: *sample size re-estimation (SSR)* designs
  - Stopping the trial early for evidence of benefit or lack thereof: *group sequential* designs
  - Selecting the most promising treatment arm(s): *multi-arm multi-stage (MAMS)* designs
  - Selecting the most promising patient subpopulation: *adaptive enrichment* designs
  - Shifting the randomization ratio toward more promising arms: *response-adaptive randomization (RAR)* designs

Adaptive designs

- Adaptive designs (ADs) are increasingly being used in practice due to their attractive features, adding flexibility to a trial design while maintaining scientific rigour
- To date, most research has focused on the question of maintaining desirable operating characteristics related to hypothesis testing (type I error rate control and power)
- General methods (e.g. p-value combination and conditional error functions) have been proposed that are applicable to a wide range of ADs

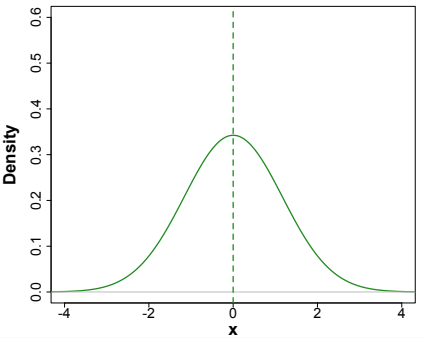
Estimation for adaptive designs

- Appropriate estimation of treatment effects is a key part of trial validity
- The question of estimation of treatment effects in ADs has received comparatively less attention
- One key issue with estimation after an AD is that the conventional point estimators can be prone to bias.
  - FDA (2019): *"A systematic tendency for the estimate of treatment effect to deviate from its true value" ... "Biased estimation in adaptive designs is currently a less well-studied phenomenon than Type I error probability inflation"*
  - ICH E20 (2025): *"Use of conventional analysis methods that would apply in non-adaptive designs usually lead to ... [a] biased treatment effect estimate."*

Estimation bias

Toy example

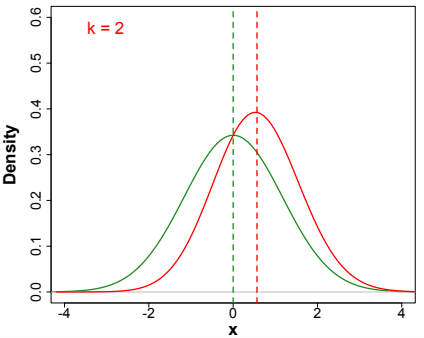
- Suppose outcome of experimental treatment follows a standard normal distribution
- What happens when we select the best-performing (i.e. treatment with highest observed mean) of k such treatments?



Estimation bias

Toy example

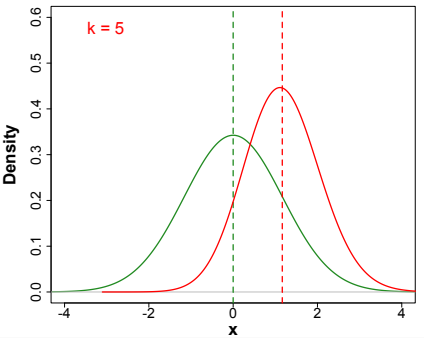
- Suppose outcome of experimental treatment follows a standard normal distribution
- What happens when we select the best-performing (i.e. treatment with highest observed mean) of k such treatments?



Estimation bias

Toy example

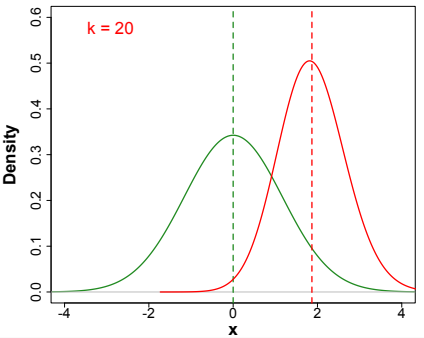
- Suppose outcome of experimental treatment follows a standard normal distribution
- What happens when we select the best-performing (i.e. treatment with highest observed mean) of k such treatments?



Estimation bias

Toy example

- Suppose outcome of experimental treatment follows a standard normal distribution
- What happens when we select the best-performing (i.e. treatment with highest observed mean) of k such treatments?



Why is biased estimation a potential problem?

- Reporting substantially biased estimates for a primary outcome measure following an adaptive design can result in poor decisions
- Concern about over- or under-estimation of treatment effects affecting further research
  - In a phase II trial, ineffective treatments with exaggerated effects may be wrongly selected for further investigations in phase III trials or potentially effective treatments may not be pursued further when their effects are underestimated
- Impact on health economic analyses
  - Patients may be penalised when a treatment is not funded based on an underestimate of cost-effectiveness, or resources may be wasted based on an overestimate of cost-effectiveness

Methods for unbiased estimation

Bias and MSE

- Mean bias of an estimator $\hat{\theta}$ of the parameter θ is

$$\text{bias}(\hat{\theta}) = \text{E}(\hat{\theta}) - \theta$$

- An estimator that has mean bias identically equal to 0 (for all possible values of $\theta \in \Theta$) is said to be *mean-unbiased*

- The *mean squared error* (MSE) is

$$\text{mse}(\hat{\theta}) = \text{E}(\hat{\theta} - \theta)^2$$

- Can be decomposed into the sum of the variance and the bias squared:

$$\text{E}(\hat{\theta} - \theta)^2 = \text{var}(\hat{\theta}) + \left[\text{bias}(\hat{\theta})\right]^2$$

Desirable criteria for point estimators

- Adequately reflects the adaptive design used
- No or small bias
- Low MSE (reflecting a favourable bias-variance trade-off)
- Is easily computable

- ICH E20 regulatory guidelines (2025): *"Sponsors should evaluate bias and variability of treatment effect estimates, including measures such as the mean squared error. 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."*

Unbiased and bias-reduced estimators

- Methods to remove or reduce the bias in the usual maximum likelihood estimator (MLE):
  1. Unbiased estimators
  2. Bias-reduced estimators

- Also distinguish *conditionally* and *unconditionally* unbiased estimators:
  - *Unconditionally* unbiased if it is unbiased when averaged across all possible realizations of an adaptive trial
  - *Conditionally* unbiased if it is unbiased only conditional on the occurrence of a subset of trial realizations
    - E.g. one might be interested in an estimator only conditional on a particular arm being selected at an interim analysis

UMV(C)UEs

- If 'no bias' is the aim, can find the unbiased estimator with the smallest possible variance
- *Uniformly Minimum Variance Unbiased Estimator* (UMVUE)
- Use the Lehmann-Scheffé theorem:
  - Suppose S is a complete sufficient statistic and U is an unbiased estimator
  - Then the estimator $\hat{U} = E(U|S)$ is the (unique) UMVUE
- Can additionally condition on the selection or stopping rule used
- *Uniformly Minimum Variance Conditionally Unbiased Estimator* (UMVCUE)

Median unbiased estimators

- So far, have been considering *mean* unbiasedness (the usual unbiased property). A reasonable alternative is to consider *median* unbiasedness
- An estimator θ_{MU} is median unbiased if $Pr(\theta_{MU} \leq \theta) = Pr(\theta_{MU} \geq \theta)$
  - i.e. the estimator underestimates just as often as it overestimates
- Median-unbiased estimators can have smaller MSE than mean-unbiased estimators
- Median-unbiased estimators are invariant under one-to-one transformations (unlike mean-unbiased estimators)

Unbiased vs bias-reduced estimators

Why not always use an unbiased estimator?

1. Not (yet) available for all classes of adaptive designs
2. Increased MSE (bias-variance trade-off) leading to large SEs and wide CIs

Bias-corrected MLE

- Whitehead (1986) proposed adjusting the MLE $\hat{\theta}_{\text{MLE}}$ to reduce bias
- Bias for the MLE as a function of the unknown parameter of interest θ is $b(\theta) = E(\hat{\theta}_{\text{MLE}}|\theta) - \theta$
- Bias-corrected MLE $\hat{\theta}_{\text{BC}}$ is the numerical solution to

$$\hat{\theta}_{\text{BC}} = \hat{\theta}_{\text{MLE}} - b(\hat{\theta}_{\text{BC}})$$

- Can also take conditional perspective and condition on selection or stopping time

Bias-reduced estimators

Other methods

- Bayesian and empirical Bayes approaches
- Resampling-based methods
  - Parametric and non-parametric bootstrap

Review papers: point estimation for ADs


Received: 25 June 2021 | Revised: 21 October 2022 | Accepted: 1 November 2022
DOI: 10.1002/sim.9605

RESEARCH ARTICLE

Statistics in Medicine WILEY

Point estimation for adaptive trial designs I: A methodological review

David S. Robertson[1] | Babak Choodari-Oskooei[2] | Munya Dimairo[3] | Laura Flight[3] | Philip Pallmann[4] | Thomas Jaki[1,5]


Received: 25 June 2021 | Revised: 20 January 2023 | Accepted: 18 March 2023
DOI: 10.1002/sim.9734

RESEARCH ARTICLE

Statistics in Medicine WILEY

Point estimation for adaptive trial designs II: Practical considerations and guidance

David S. Robertson[1] | Babak Choodari-Oskooei[2] | Munya Dimairo[3] | Laura Flight[3] | Philip Pallmann[4] | Thomas Jaki[1,5]

- Two-part paper series in *Statistics in Medicine*
  - **Part I**: Review methods for unbiased and bias-reduced estimation of treatment effects after an adaptive clinical trial and critically discuss different approaches
  - **Part II**: Point estimation for adaptive designs from a practical perspective, including a set of guidelines for best practice
- Available at `doi:10.1002/sim.9605` and `doi:10.1002/sim.9734`

Summary of systematic review of methods

- 145 methodological papers found
- Classified by type of design (Group sequential, SSR, MAMS, RAR, Adaptive enrichment)
- Literature on unbiased and bias-adjusted estimation for adaptive designs has grown rapidly in recent years
- However, only a few real-life trial examples/case studies, and lack of software/code.

Summary of systematic review of methods

- Group sequential designs have received the most attention by far, and some of the estimation methods are implemented in widely available statistical software
  - ADDPLAN, SAS, East
  - R (rpact, RCTdesign, AGSDest, OptGS)
- Little literature on estimation for RAR
- Some methods restricted to two-stage designs
- Relatively few methods for trials with time-to-event outcomes

Current practice

- **Group sequential designs**: Stevely et al. (2015) identified 68 trials that were published in leading medical journal
  - Only 7% (3/46) disclosed the use of some form of bias correction.

Current practice

- **Group sequential designs**: Stevely et al. (2015) identified 68 trials that were published in leading medical journal
  - Only 7% (3/46) disclosed the use of some form of bias correction.
- Systematic search of adaptive designs in MEDLINE database, from 1st January 2000 to 30th May 2022

Type of adaptive design	Number of records screened	Number of randomised trials that reported results	Number of randomised trials that reported an unbiased or bias-adjusted estimate
MAMS	773	22	2
RAR	59	17	0
Adaptive enrichment	530	3	0

Current practice

Why has there been so little uptake of adjusted estimators in practice?

- Belief that the bias of the MLE will typically be negligible in realistic trial scenarios
- Lack of awareness of the range of different unbiased and bias-adjusted estimators in the methodological literature
- Statistical software and code to calculate adjusted estimators is relatively sparse
- For more complex or novel adaptive designs, adjusted estimators may not exist

Unbiased estimation for group sequential trials

Example: two-stage group sequential trial

- When a group sequential trial is stopped at analysis j with $Z_j = z_j \rightarrow H_0 : \theta \leq 0$ rejected if $z_j > e_j$
- Stopping boundaries chosen to preserve type I error rate

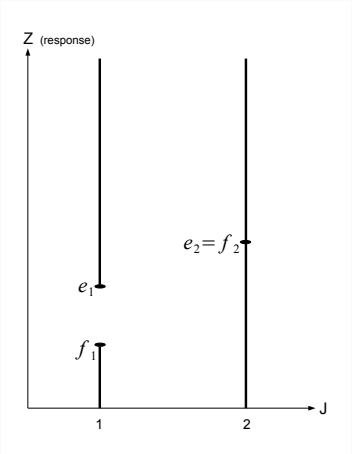
Example: two-stage group sequential trial

- When a group sequential trial is stopped at analysis j with $Z_j = z_j \rightarrow H_0 : \theta \leq 0$ rejected if $z_j > e_j$
- Stopping boundaries chosen to preserve type I error rate
- MLE $\hat{\theta}$ generally reported: $\hat{\theta}_j = \frac{Z_j}{\sqrt{\mathcal{I}_j}}$
- The MLE ignores the sequential nature of the trial and can be *biased*

Example: two-stage group sequential trial

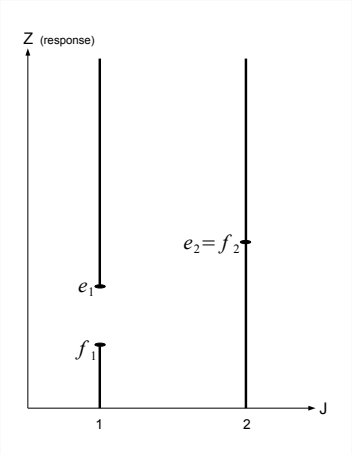
- When a group sequential trial is stopped at analysis j with $Z_j = z_j \rightarrow H_0 : \theta \leq 0$ rejected if $z_j > e_j$
- Stopping boundaries chosen to preserve type I error rate
- MLE $\hat{\theta}$ generally reported: $\hat{\theta}_j = \frac{Z_j}{\sqrt{\mathcal{I}_j}}$
- The MLE ignores the sequential nature of the trial and can be *biased*
- Earlier a trial stopped for efficacy, the more likely it is that the observed treatment effect is over-estimated
- Controlling type I error rate does not control bias

Example 1: Two-stage trial



- Want to estimate θ regardless of stage the trial stops

Example 1: Two-stage trial



- Want to estimate θ regardless of stage the trial stops
- Define bias of an estimator $\hat{\theta}$ as

$$E[\hat{\theta}] - \theta = \sum_{k=1}^{2} E[\hat{\theta}|j = k]Pr(j = k) - \theta$$

- This is *unconditional* bias

Median unbiased estimator

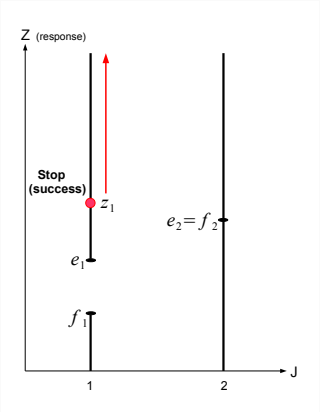
- Seek an estimator $\hat{\theta}_{MU}$: $Pr(\hat{\theta}_{MU} \leq \theta) = Pr(\hat{\theta}_{MU} \geq \theta)$
- In order to achieve this, use the following procedure:
  1. Define an **ordering** of design space w.r.t evidence against $H_0 : \theta \leq 0$
  2. Define a **p-value function** $P_z(\theta)$: gives probability that, at the stage the trial stopped, even more extreme evidence against H_0 could have been observed
  3. At the point the trial stops, find median unbiased estimator (MUE) $\hat{\theta}_{MU}$: $P_z(\hat{\theta}_{MU}) = \frac{1}{2}$

Ordering the design space

- **Stage-wise** ordering most common
- For two test statistics (z_j, z_k), $P_{z_j}(\theta) \leq P_{z_k}(\theta)$ if:
  - $j < k$ and $z_j > e_j$
  - **or** if $j = k$ and $z_j \geq z_k$
  - **or** if $j > k$ and $z_k \leq f_k$

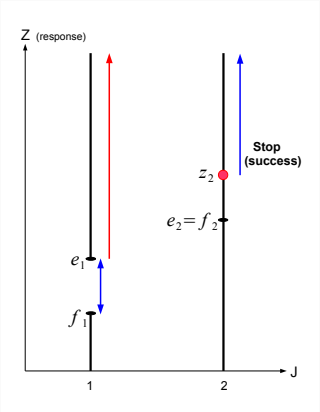
Example: Case (1) stop at stage 1 for efficacy

- $P_{z_1}(\theta) = P(Z_1 \geq z_1|\theta)$



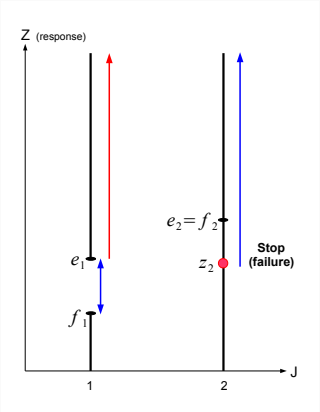
Example : Case (2) stop at stage 2

- $P_{z_2}(\theta) = P(Z_1 > e_1|\theta) + P(Z_1 \in (f_1, e_1] \cap Z_2 \geq z_2|\theta)$



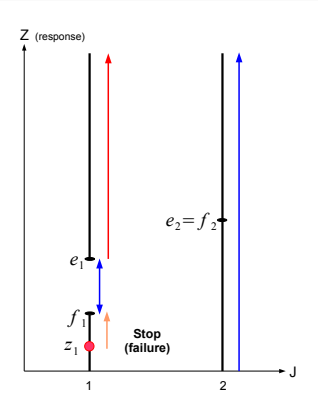
Example : Case (2) stop at stage 2

- $P_{z_2}(\theta) = P(Z_1 > e_1|\theta) + P(Z_1 \in (f_1, e_1] \cap Z_2 \geq z_2|\theta)$



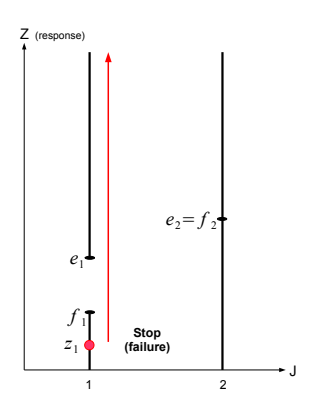
Example : Case (3): stop at stage 1 for futility

- $P_{z_1}(\theta) = P(Z_1 \in (z_1, f_1)|\theta) + P(Z_1 \in (f_1, e_1) \cap Z_2 \geq -\infty|\theta)$
  $+ P(Z_1 \in (e_1, \infty)|\theta)$



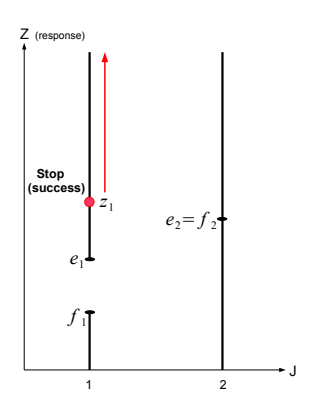
Example : Case (3) = Case (1)

- $P_{z_1}(\theta) = P(Z_1 \geq z_1|\theta)$



Example : Case (3) = Case (1)

- $P_{z_1}(\theta) = P(Z_1 \geq z_1|\theta)$



MUE and confidence intervals

$$P_z(\theta) = \begin{cases} P_{z_1}(\theta) = \Phi(\theta\sqrt{\mathcal{I}_1} - z_1) & \text{if } z_1 \notin (f_1, e_1] \\ P_{z_2}(\theta) = \Phi(\theta\sqrt{\mathcal{I}_1} - e_1) + P(Z_1 \in (f_1, e_1) \cap Z_2 \geq z_2|\theta) & \text{if } z_1 \in (f_1, e_1] \end{cases}$$

- Find $\hat{\theta}_{MU}$: $P_z(\hat{\theta}_{MU}) = \frac{1}{2}$ for MUE

MUE and confidence intervals

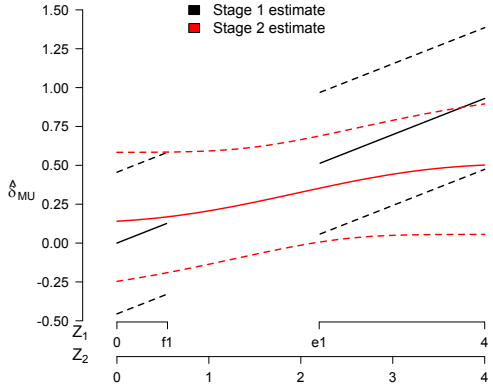
$$P_z(\theta) = \begin{cases} P_{z_1}(\theta) = \Phi(\theta\sqrt{\mathcal{I}_1} - z_1) & \text{if } z_1 \notin (f_1, e_1] \\ P_{z_2}(\theta) = \Phi(\theta\sqrt{\mathcal{I}_1} - e_1) + P(Z_1 \in (f_1, e_1) \cap Z_2 \geq z_2|\theta) & \text{if } z_1 \in (f_1, e_1] \end{cases}$$

- Find $\hat{\theta}_{MU}$: $P_z(\hat{\theta}_{MU}) = \frac{1}{2}$ for MUE
- $100\times(1-\alpha)\%$ two-sided adjusted confidence interval for θ: find lb, ub where $P(\hat{\theta}_{MU}(lb)) = \frac{\alpha}{2}$ and $P(\hat{\theta}_{MU}(ub)) = 1 - \frac{\alpha}{2}$
- Overall p-value for trial is given by $p = P_z(0)$

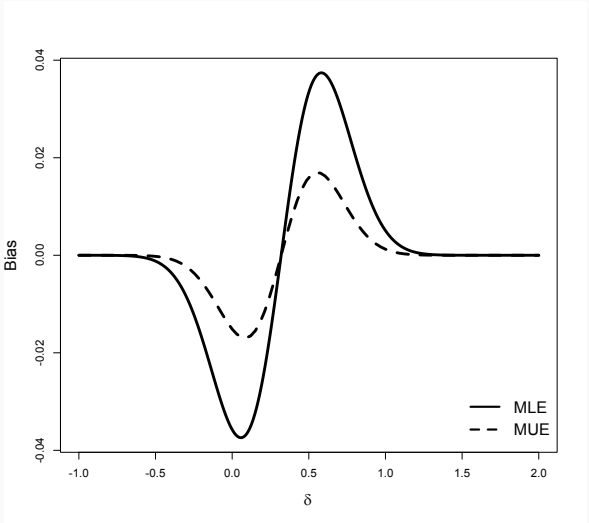
Example

- Normally distributed endpoint with known variance
- Suppose we design a two-stage trial with:
  - Early stopping for efficacy and futility
  - 90% power to detect standardised mean difference $\frac{\theta}{\sigma} = 0.5$
  - 5% type I error rate
- A possible trial design is:
  - $\boldsymbol{n} = (37, 74)$, $(f_1, f_2) = (0.55, 1.65)$ and $(e_1, e_2) = (2.2, 1.65)$
  - Hence $\mathcal{I}_1 = 18.5$ and $\mathcal{I}_2 = 37$

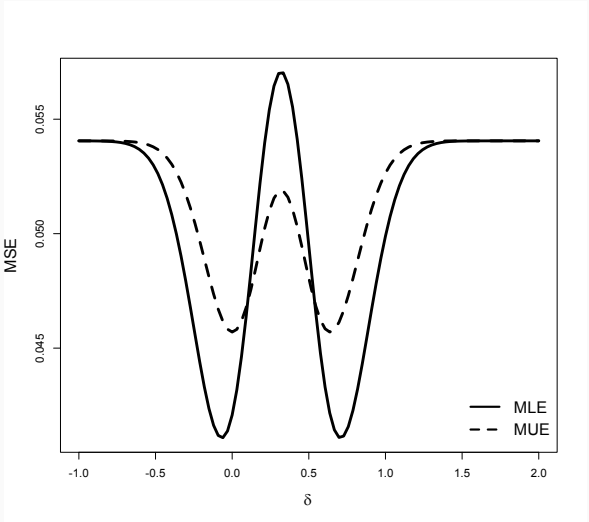
Estimate


Stage 1 estimate
Stage 2 estimate
1.50
1.25
1.00
0.75
0.50
0.25
0.00
-0.25
-0.50
$\hat{\delta}_{MU}$
Z_1
Z_2
0
f1
e1
4
0
1
2
3
4

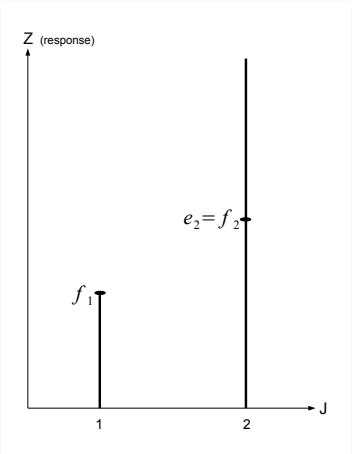
Unconditional bias



Unconditional MSE



Example 2: Two-stage trial, futility stopping only



- Trial stops at stage 1 if $z_1 \leq f_1$ (for futility)
- Common in early phase oncology trials
- Only interested in estimation if trial continues to stage 2
- Define *conditional* bias as $E[\hat{\theta}|j = 2] - \theta$

UMVCUE

- The UMVCUE is given by

$$\tilde{\theta} = \hat{\theta}_2 - \frac{(\mathcal{I}_2^*)^{-1}}{\sqrt{\mathcal{I}_1^{-1} + (\mathcal{I}_2^*)^{-1}}} \frac{\phi\left[\frac{\sqrt{\mathcal{I}_1^{-1}+(\mathcal{I}_2^*)^{-1}}}{\mathcal{I}_1^{-1}}\left(\hat{\theta}_2 - \frac{f_1}{\sqrt{\mathcal{I}_1}}\right)\right]}{\Phi\left[\frac{\sqrt{\mathcal{I}_1^{-1}+(\mathcal{I}_2^*)^{-1}}}{\mathcal{I}_1^{-1}}\left(\hat{\theta}_2 - \frac{f_1}{\sqrt{\mathcal{I}_1}}\right)\right]}$$

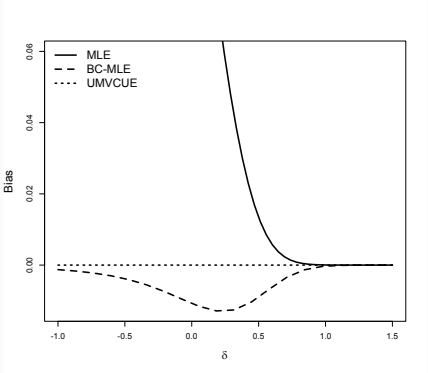
  where $\mathcal{I}_2^* = \mathcal{I}_2 - \mathcal{I}_1$

- Has the form UMVCUE = MLE – Bias

Example 2

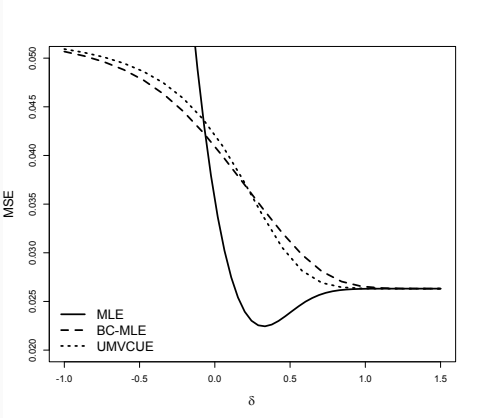
- Normally distributed endpoint with known variance
- Suppose we design a two-stage trial with:
  - 90% power to detect standardised mean difference $\frac{\theta}{\sigma} = 0.5$
  - 5% type I error rate
- A possible trial design is:
  - $\boldsymbol{n} = (38, 76)$, $(f_1, f_2) = (0.5775, 1.7325)$ and $(e_1, e_2) = (\infty, 1.7325)$
  - Hence $\mathcal{I}_1 = 19$ and $\mathcal{I}_2 = 38$

Example: conditional bias and MSE



- MLE can have a massive positive bias
- Bias-corrected MLE generally overcorrects for bias
- UMVCUE uniformly unbiased (by definition)

Example: conditional bias and MSE



- MLE can have a large MSE, but can also have the smallest MSE when $\theta > 0$
- Bias-corrected MLE and UMVCUE perform similarly

Case study: MUSEC trial

Case study: MUSEC trial

- Two-arm trial of oral cannabis extract vs placebo in adults with MS
- Primary endpoint: relief from muscle stiffness (yes/no) after 12 weeks
- Two-stage group-sequential design (maximum $n = 300$) with interim analysis (after $n = 200$)
  - O'Brien-Fleming efficacy stopping boundaries

	Interim data		**Final data**	
	Placebo	**CE arm**	**Placebo**	**CE arm**
Number of subjects with relief from muscle stiffness	12	27	21	42
Total number of subjects	97	101	134	143
Standardized test statistic	2.540		2.718	
OBF boundary	2.797		1.977	

Case study: MUSEC trial

Type of estimator	Estimator	Estimate (SE)	Relative difference
Standard/naive	MLE	0.137 (0.054)	–
Unconditional	MUE	0.134 (0.054)	−2%
	UMVUE	0.128 (0.054)	−7%
	Bias-corrected MLE	0.133 (0.055)	−3%
Conditional	MUE	0.185 (0.080)	+35%
	UMVCUE	0.172 (0.071)	+26%
	Bias-corrected MLE	0.191 (0.073)	+39%

MLE = Maximum Likelihood Estimator; MUE = Median Unbiased Estimator; UMVUE = Uniform Minimum Variance Unbiased Estimator; UMVCUE = Uniform Minimum Variance Conditionally Unbiased Estimator

Estimation in practice

How to do estimation in practice

- The issue of estimation should be considered throughout the whole lifecycle of an adaptive trial
- The design and analysis of an adaptive trial are closely linked, and one should not be considered without the other
- Context, aims and design of an adaptive trial should all inform the analysis strategy used, which includes the choice of point estimators
- A review of the literature may be sufficient to compare their performance
- Otherwise, conduct simulations to explore their properties
- Statistical analysis plan should include a description and justification of the point estimators used

How to do estimation in practice

Planning stage

- Decide on what exactly is to be estimated (i.e. the estimands of interest)
- Decide on desired characteristics of point estimators
  - Conditional versus unconditional perspective (see Marschner 2021)
  - Bias-variance trade-off for point estimators
  - Link between point estimators and confidence intervals
- A review of the literature may be sufficient. Otherwise, conduct simulations to explore

Pre-specification of analyses

- Statistical analysis plan (SAP) and health economic analysis plan (HEAP) should include a description and justification of the estimators
- When available, unbiased or bias-reduced estimators should be used and reported alongside the standard MLE
- If multiple adjusted estimators are available and are of interest, one adjusted estimatorshould be designated the 'primary' adjusted estimator
  - Others included as sensitivity or supplementary analyses

Proposed guidance

Data Monitoring Committees

- When presenting interim results to DMCs, the issue of potential bias should also be considered
- The sensitivity of the standard MLE to potential bias should be reported
- When unbiased or bias-reduced estimators are available, these should also be presented to the DMC

Reporting results for a completed trial

- There should be a clear description of the statistical methods used to estimate treatment effects
- Adjusted point estimates taking the trial design into account are to be preferred (see FDA, EMA and ICH E20 guidance!)
- FDA guidance: *"if naive estimates such as unadjusted sample means are used, the extent of bias should be evaluated, and estimates should be presented with appropriate cautions regarding their interpretation"*

Discussion

Discussion

- Point estimates typically need adjusting after an adaptive design
- There is a growing body of methodological literature proposing unbiased estimators for a variety of adaptive trial designs
- Methodological gaps remain:
  - Designs beyond group sequential
  - Endpoints beyond normal and binary
  - Conditional vs unconditional perspective (see Marschner 2021)
- There has been relatively little uptake of adjusted point estimators in practice
- Need for further development of user-friendly software and code
- The reporting of appropriate measures of uncertainty for the estimators, such as confidence intervals, is also key (see Robertson et al. 2025a,b)

References i

EMA (2007)
Reflection paper on methodological issues in confirmatory clinical trials planned with an adaptive design
https://www.ema.europa.eu/en/documents/scientific-guideline/reflection-paper-methodological-issues-confirmatory-clinical-trials-planned-adaptive-design_en.pdf

US FDA (2019)
Adaptive Designs for Clinical Trials of Drugs and Biologics: Guidance for Industry.
https://www.fda.gov/media/78495/download

C. Jennison and B. Turnbull
Group sequential methods with applications to clinical trials
Chapman & Hall, 2000.

Marschner IC (2021)
A General Framework for the Analysis of Adaptive Experiments
Statistical Science, 36(3):465–492.

References ii

Robertson DS et al. (2023a)
Point estimation for adaptive trial designs I: A methodological review
Statistics in Medicine, 42(2):122–145

Robertson DS et al. (2023b)
Point estimation for adaptive trial designs II: Practical considerations and guidance
Statistics in Medicine, 42(14):2496–2520

Robertson DS et al. (2025a)
Confidence intervals for adaptive trial designs I: A methodological review
Statistics in Medicine, 44 (18-19), e70174

Robertson DS et al. (2025b)
Confidence intervals for adaptive trial designs II: Practical considerations and guidance
Statistics in Medicine, 44 (18-19), e70202

Steveley A et al. (2015)
An Investigation of the Shortcomings of the CONSORT 2010 Statement for the Reporting of Group Sequential Randomised Controlled Trials: A Methodological Systematic Review.
PLOS ONE, 10(11):e0141104. doi:10.1371/journal.pone.0141104