Estimation after an Adaptive Design

Christopher Jennison

Department of Mathematical Sciences,

University of Bath, UK

http://people.bath.ac.uk/mascj

CEN 2026 Conference

Warsaw

May 2026

ICH E20: Adaptive Designs for Clinical Trials

The ICH (International Council for Harmonisation) Guideline E20 gives principles for the design and analysis of adaptive clinical trials.

Examples of adaptive designs include

- Group sequential tests stopping for efficacy or futility,
- Adaptive trials with sample size re-assessment,
- Adaptive trials testing multiple hypotheses:
  - Seamless Phase 2-3 trials with treatment selection,
  - Multi-arm multi-stage (MAMS) designs,
  - Enrichment designs.

Principle 4 concerns Reliability of Estimation:

Estimates and confidence intervals for payers to use in making cost-benefit decisions.

Inference on termination of an adaptive design

Problems arise due to:

(i) Complex sample spaces

The data collected in an adaptive trial is data-dependent.

(ii) Flexible designs

Since the sample space is not completely known, properties such as the expected value of an estimate cannot be calculated.

(iii) Confidence intervals in multiple testing problems

A multiple testing procedure may "exhaust" the type I error probability α in testing null hypotheses, leaving nothing to test non-null parameter values.

This can lead to the situation where H_0: $\theta \leq 0$ is rejected but the upper confidence interval for θ is $(0, \infty)$.

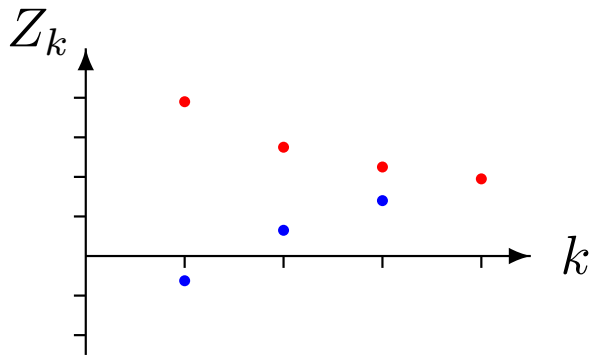
Inference on termination of an adaptive design

Methods are available for some, but not all, cases.

Type of design	Treatment effect estimate with negligible bias	Confidence interval, median unbiased estimate
Group sequential test (GST)	✓	✓
GST + sample size re-assessment (fixed)	✓	✓
GST + sample size re-assessment (flexible)	$\sim$ ✓	$\sim$ ✓
Phase 2/3 with treatment selection, always selecting 1 treatment for Phase 3	✓	✓
Phase 2/3 with treatment selection, selecting 1 or more treatments for Phase 3	✓	✓?
Multi-arm multi-stage design (fixed)	✓	✓?
Multi-arm multi-stage design (flexible)	?	?
Enrichment design (flexible)	✓	✓?

Point estimation after a group sequential test (GST)

Consider a two-treatment comparison where the treatment effect θ is the difference in mean responses between the experimental treatment and the control.



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

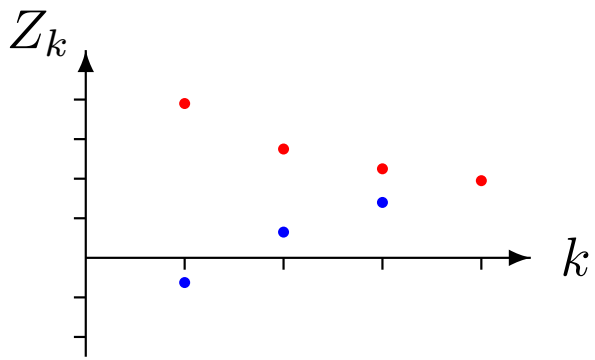
$$\widehat{\theta}_M = \bar{X}_{Ak} - \bar{X}_{Bk}.$$

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.

Point estimation after a group sequential test (GST)



Similarly, for negative values of θ:

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

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

Thus, the MLE is biased with

$$E_\theta(\widehat{\theta}_M) > \theta \text{ for high values of } \theta$$

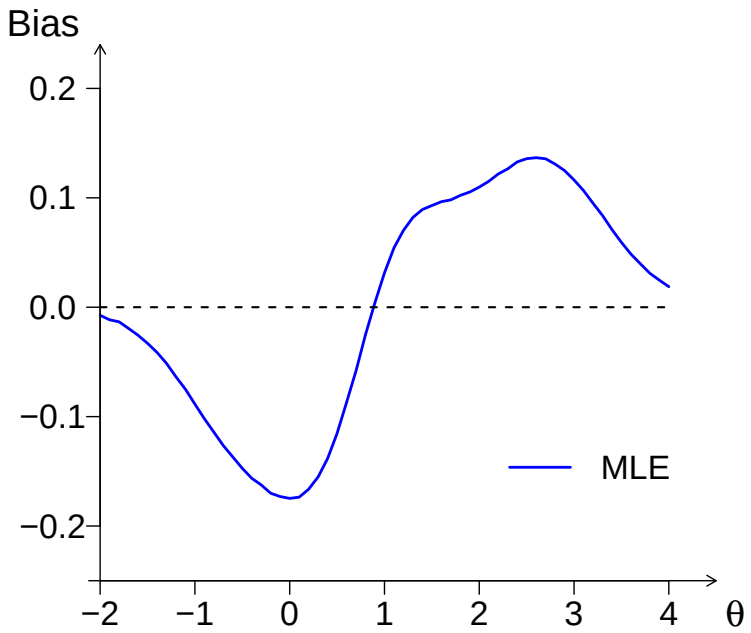
and

$$E_\theta(\widehat{\theta}_M) < \theta \text{ for low values of } \theta.$$

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

Consider a Pampallona & Tsiatis GST with 4 analyses, $\Delta = 0$, $\alpha = 0.025$ and **power $1 - \beta = 0.8$ at $\theta = 1$** (see Jennison & Turnbull, 2026, 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 $b(\widehat{\theta}_M)$.

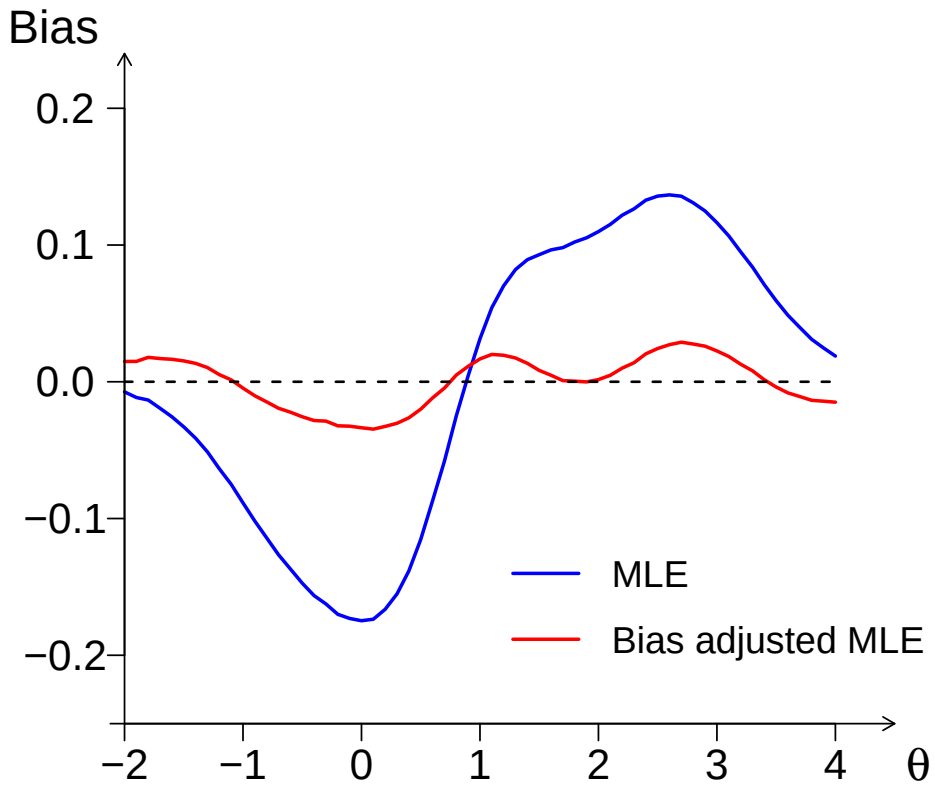
The adjusted estimator is then

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

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, unobserved information levels.

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

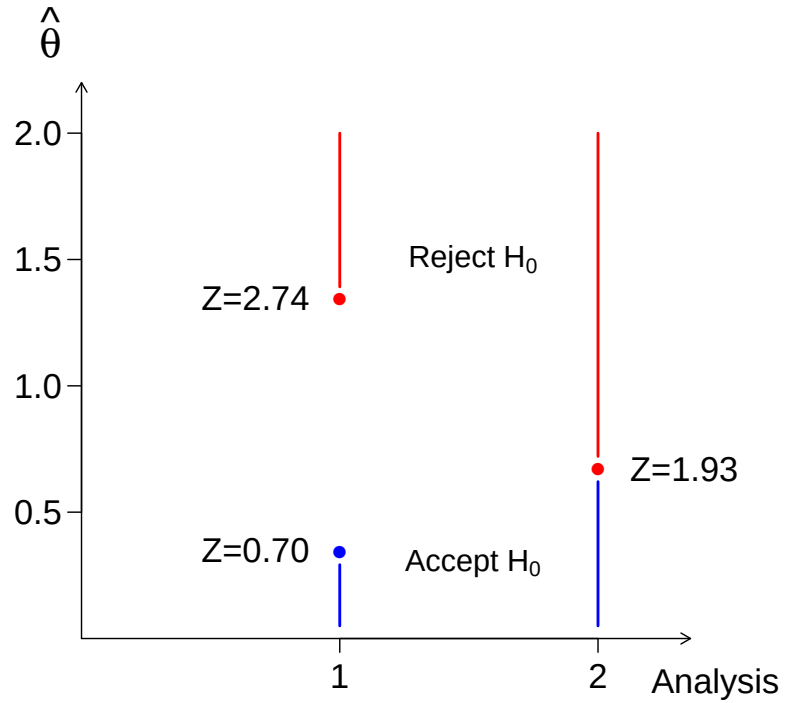
However, unbiased estimators may have a large variance and a relatively high mean square error.

The UMVUE may also be rather strange!

Unbiased estimation after a group sequential test

Clinical trial designs with just two analyses are common.

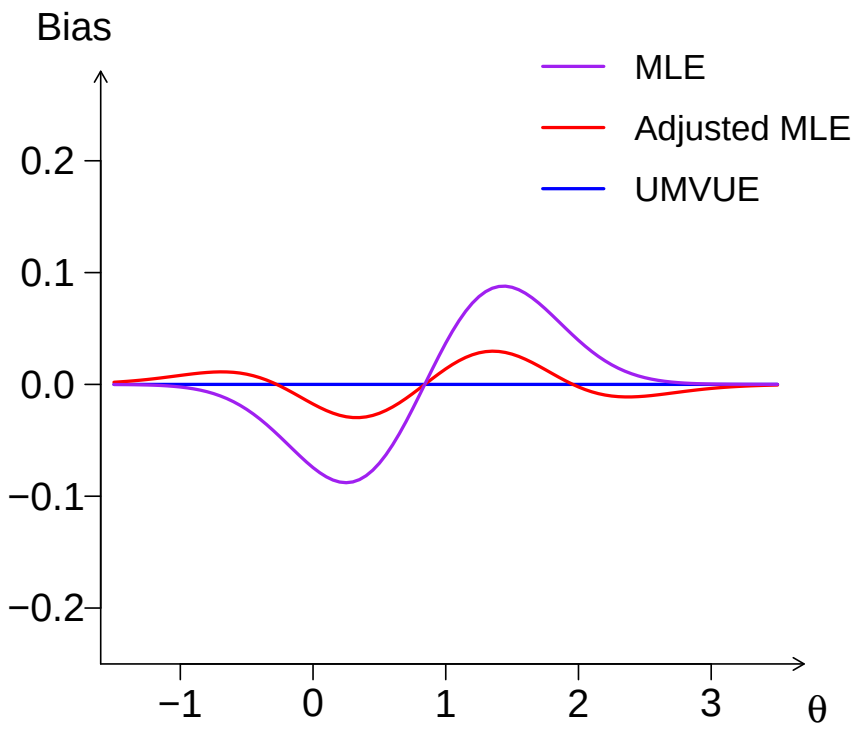
Consider the following one-sided, group sequential test of H_0: $\theta \leq 0$ against $\theta > 0$.



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:



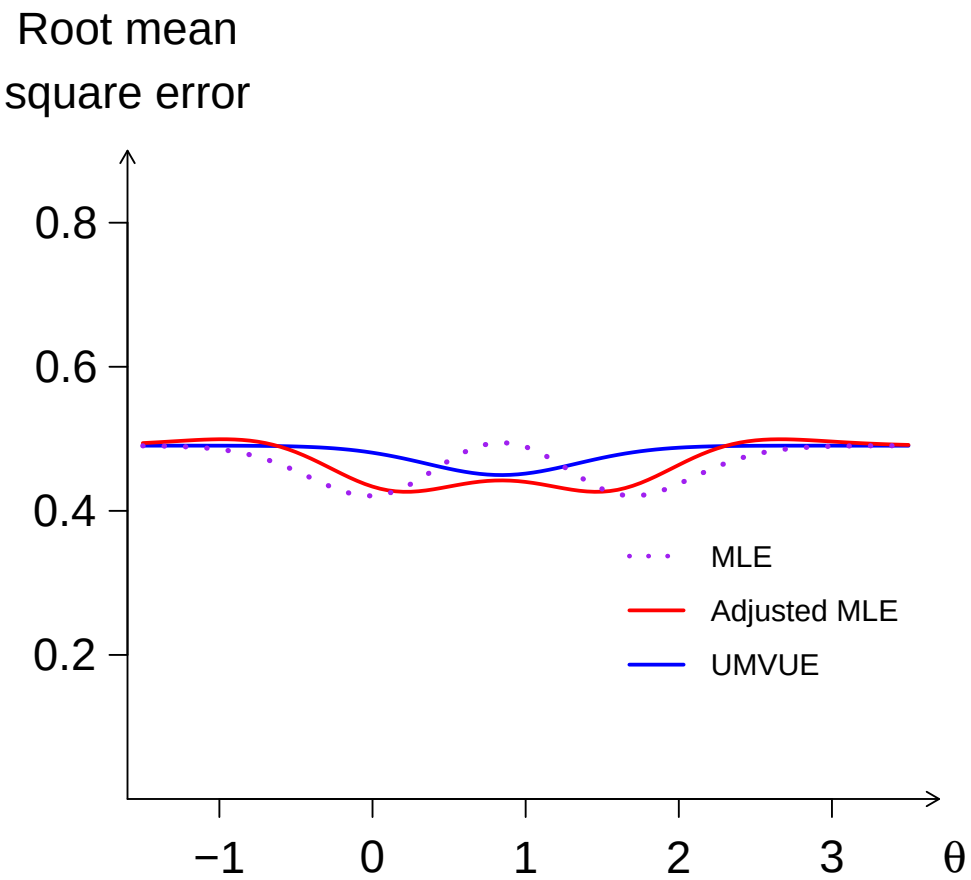
The MLE on termination, $\widehat{\theta}_M$, has bias function $b(\widehat{\theta})$.

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

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

Unbiased estimation after a group sequential test

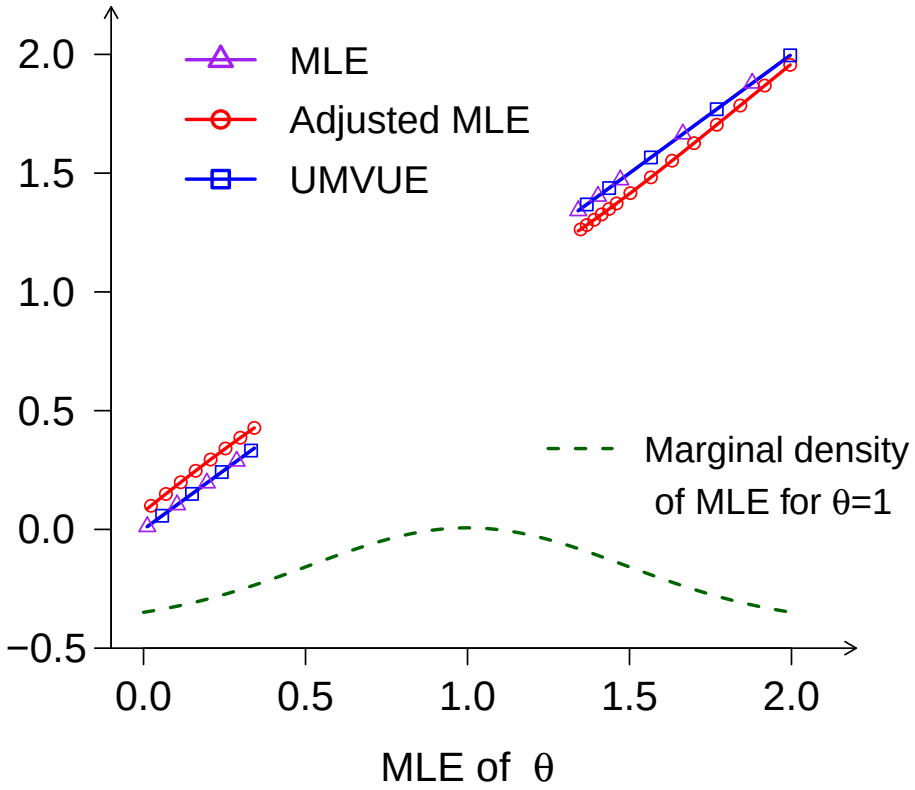
The root mean square error of several estimates:



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

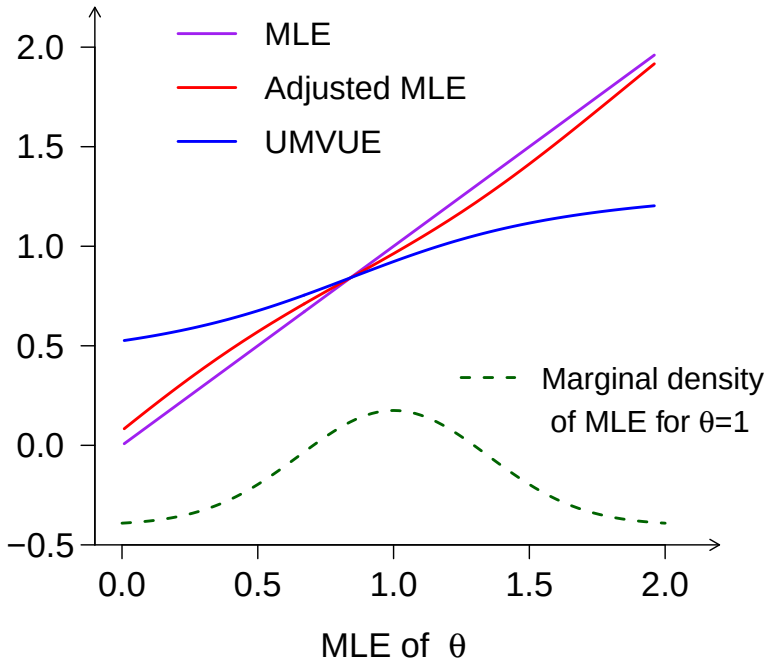


If the trial stops at analysis 1 the UMVUE is $\widehat{\theta}_1$, the MLE of θ.

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 complex adaptive design

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 sequential 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.

Conclusions

Simple estimates of a treatment effect after a group sequential or adaptive design are liable to be biased.

In many cases, methods are available to reduce bias or even to remove it completely.

However, methods that eliminate bias completely may have undesirable features.

Further research is needed into methods for creating point estimates and confidence intervals after more complex adaptive designs.