ICH E20: Adaptive Designs for Clinical Trials

Comments by Christopher Jennison

Department of Mathematical Sciences,

University of Bath, UK

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

EFSPI Regulatory Statistics Workshop

Basel

September 2025

Scope of ICH E20: Adaptive Designs for Clinical Trials

The Guideline lays down general principles, rather than specifying particular designs.

Both frequentist and Bayesian approaches are considered.

While ICH E20 focuses on confirmatory trials, the Guideline also considers impact on the wider drug development process.

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.

ICH E20: General principles

The Guideline defines five principles.

1. Adequacy within the development program:

   Justifying the selected dose, etc.

2. Adequacy of trial planning:

   Pre-planned, as simple as possible, some flexibility.

3. Limiting the chances of erroneous conclusions:

   Type I error control.

4. Reliability of estimation:

   Estimates and confidence intervals for cost-benefit decisions.

5. Maintenance of trial integrity:

   Blinding, avoiding information leakage, role of IDMCs.

Emerging topics

1. Bayesian Designs

Section 5.3 states that views on Bayesian methods are "not fully harmonized" and "public consultation comments are sought".

Bayesian designs combine a prior distribution for unknown parameters with decision making criteria, possibly optimizing a specified gain function.

In doing this, the type I error rate is not a primary concern.

Perhaps ICH E20 should be more explicit in referring to the type I error rate.

However, a Bayesian design can be "calibrated" to satisfy Principle 3 and "limit the chance of erroneous conclusions".

Demonstrating control of type I error can be difficult, particularly when there are multiple parameters and the family-wise type I error rate is to be controlled.

Emerging topics: Bayesian Designs

Section 5.2 describes 12 steps that should be followed when conducting a large-scale simulation study to determine the operating characteristics of a proposed design.

This process may well be computer-intensive and time-consuming.

When there are multiple parameters, it may not be obvious where an adaptive procedure's maximum type I error probability occurs.

The process of calibrating a Bayesian design may nullify the effort taken to define a good prior or a realistic gain function.

Another possibility is to use a hybrid Bayes-frequentist design.

In this approach, frequentist tools are used to control the family-wise error rate. Then, where flexibility remains, Bayesian criteria are applied to carry out adaptations — using a prior and gain function that investigators believe in.

2. Inference on termination

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

Type of design	Treatment effect estimate with little or no 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)	✓	✓?

Emerging topics: Inference on termination

Two common difficulties arise.

(i) Flexible designs

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

(ii) Confidence intervals in multiple testing problems

A confidence interval for a parameter is the result of testing a family of hypotheses concerning all possible parameter values.

A multiple testing procedure may "exhaust" the type I error probability α, 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 $(0, \infty)$ — an uninformative confidence interval

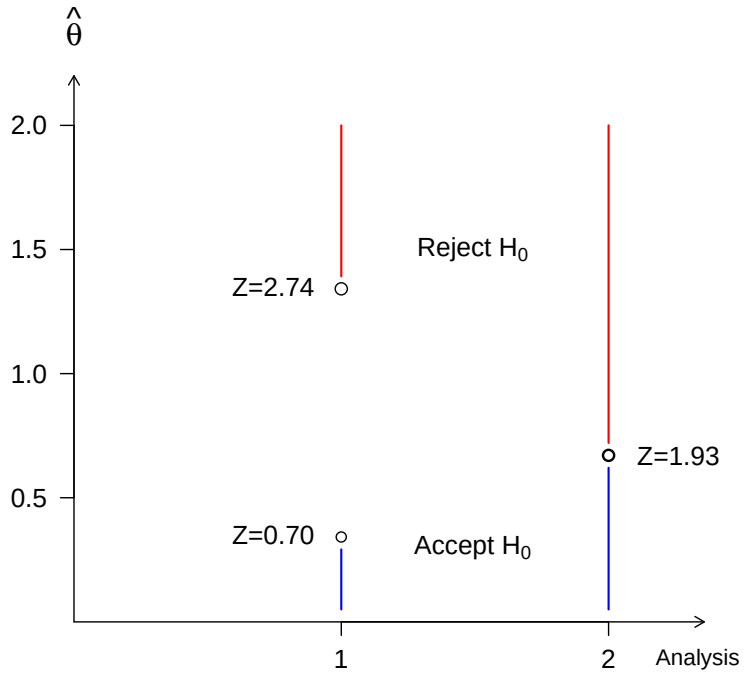
Emerging topics: Inference on termination

Point estimates

ICH E20 refers to estimates that are unbiased or have small bias.

While being exactly unbiased may seem desirable, estimators that achieve this can have some strange properties.

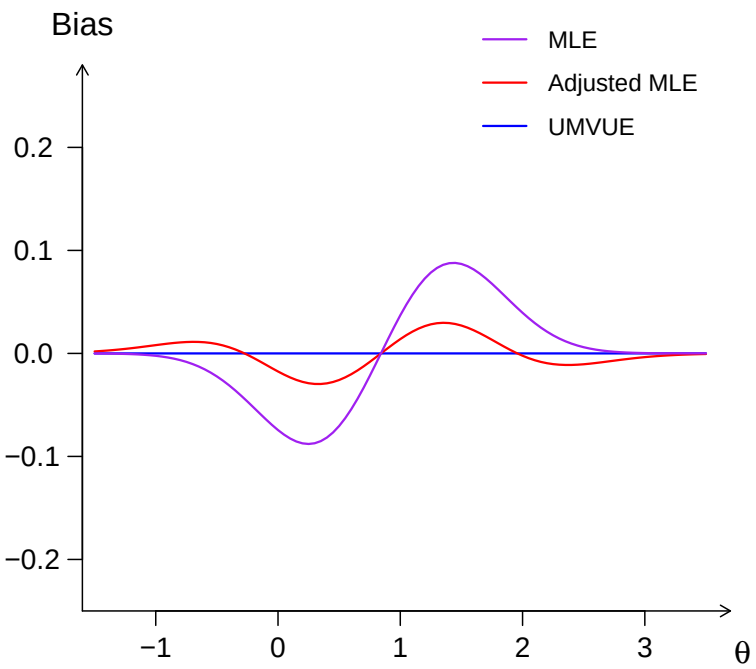
Consider a group sequential test with two analyses.



The test of H_0: $\theta \leq 0$ against $\theta > 0$ has power 0.8 for $\theta = 1$.

Emerging topics: Point estimates on termination

The bias of several estimates:



We denote the maximum likelihood estimate (MLE) on termination by $\widehat{\theta}$.

The Adjusted MLE is formed by subtracting the bias when $\theta = \widehat{\theta}$ from $\widehat{\theta}$.

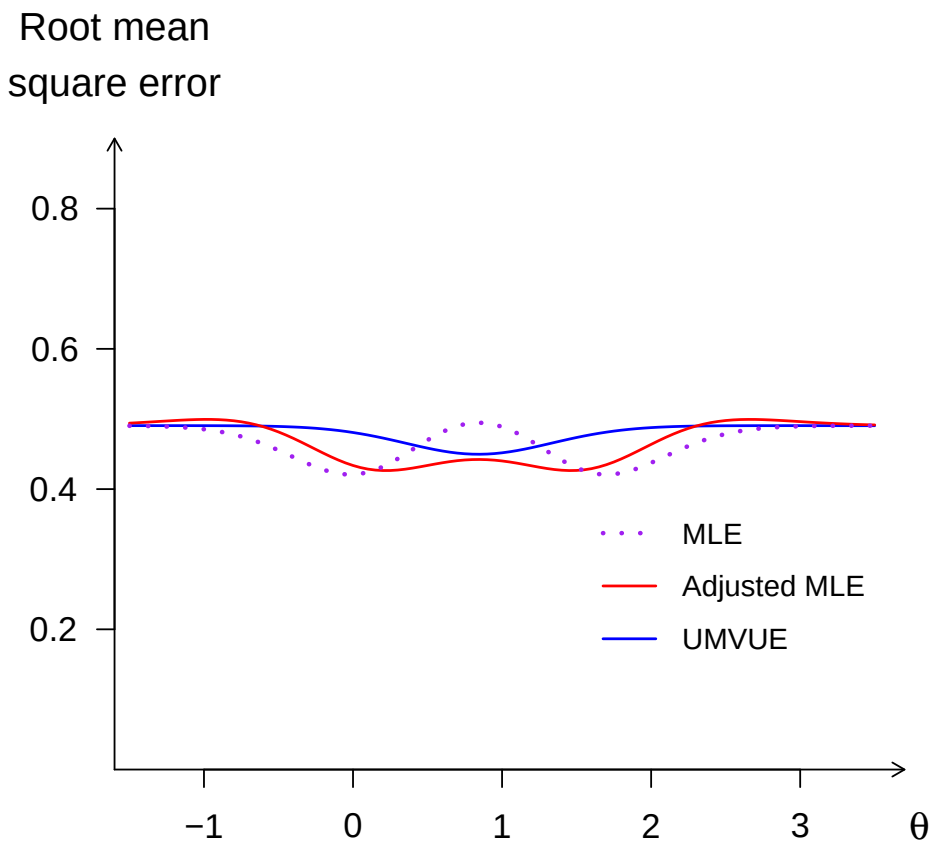
(Whitehead, *Bmka*, 1986)

The Uniform Minimum Variance Unbiased Estimate (UMVUE) uses the fact that $\widehat{\theta}_1$, the MLE at analysis 1, is unbiased for θ.

Applying "Rao-Blackwellization", the UMVUE is the conditional expectation of $\widehat{\theta}_1$, given the final data.

Emerging topics: Point estimates on termination

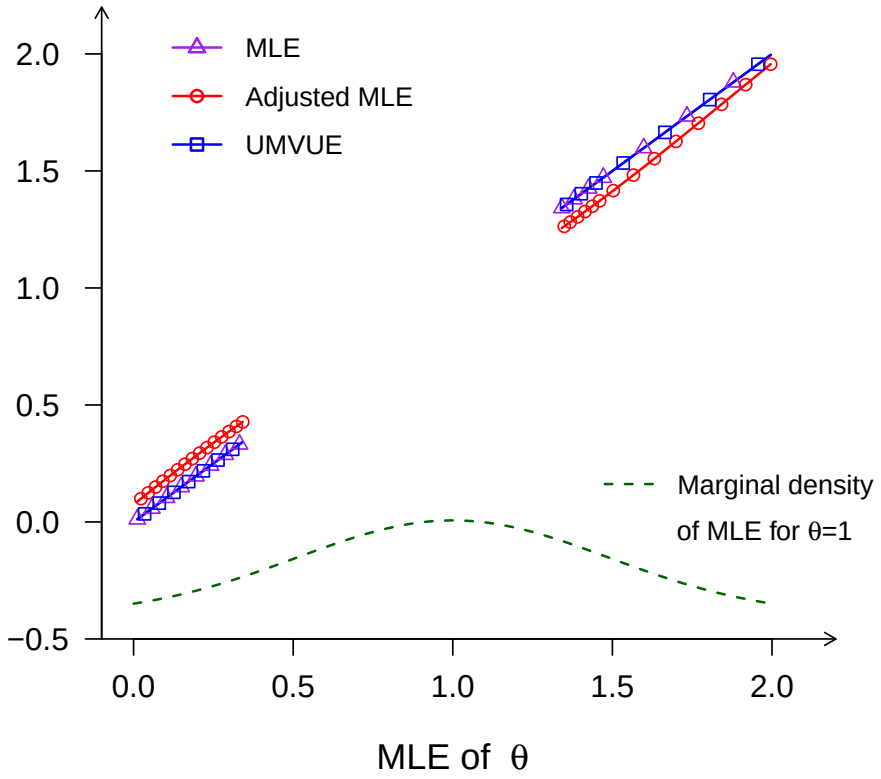
The root mean square error of several estimates:



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

Emerging topics: Point estimates on termination

Estimates on termination at analysis 1

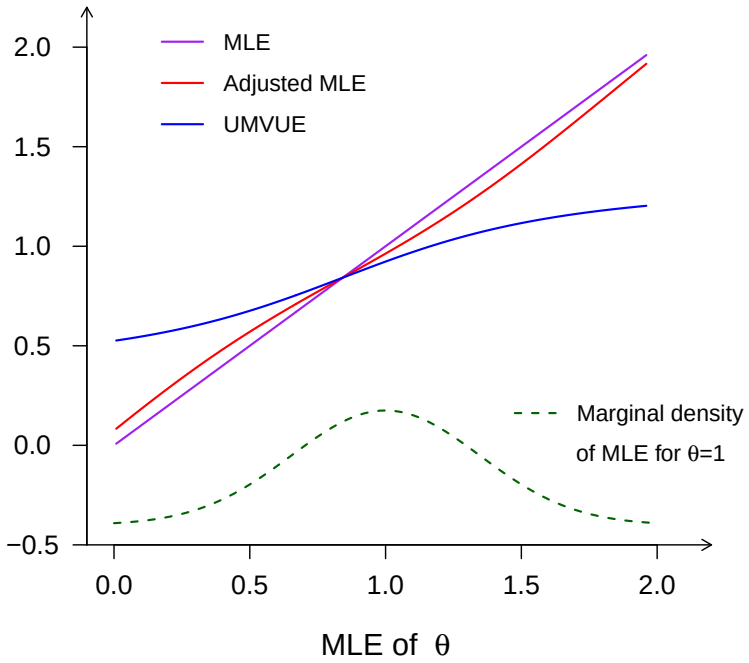


When stopping at analysis 1, the UMVUE is $\widehat{\theta}_1$, the MLE of θ.

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

Emerging topics: Point estimates on termination

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.

Emerging topics: Point estimates on termination

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.