MRC Biostatistics Unit

UNIVERSITY OF CAMBRIDGE

Assessing the impact of interim decisions in group sequential trials

G. Caruso*, W. F. Rosenberger, P. Mozgunov and N. Flournoy

PSI Conference 2026, Belfast, UK
16 June 2026

*gianmarco.caruso@mrc-bsu.cam.ac.uk

Group sequential designs in clinical trials

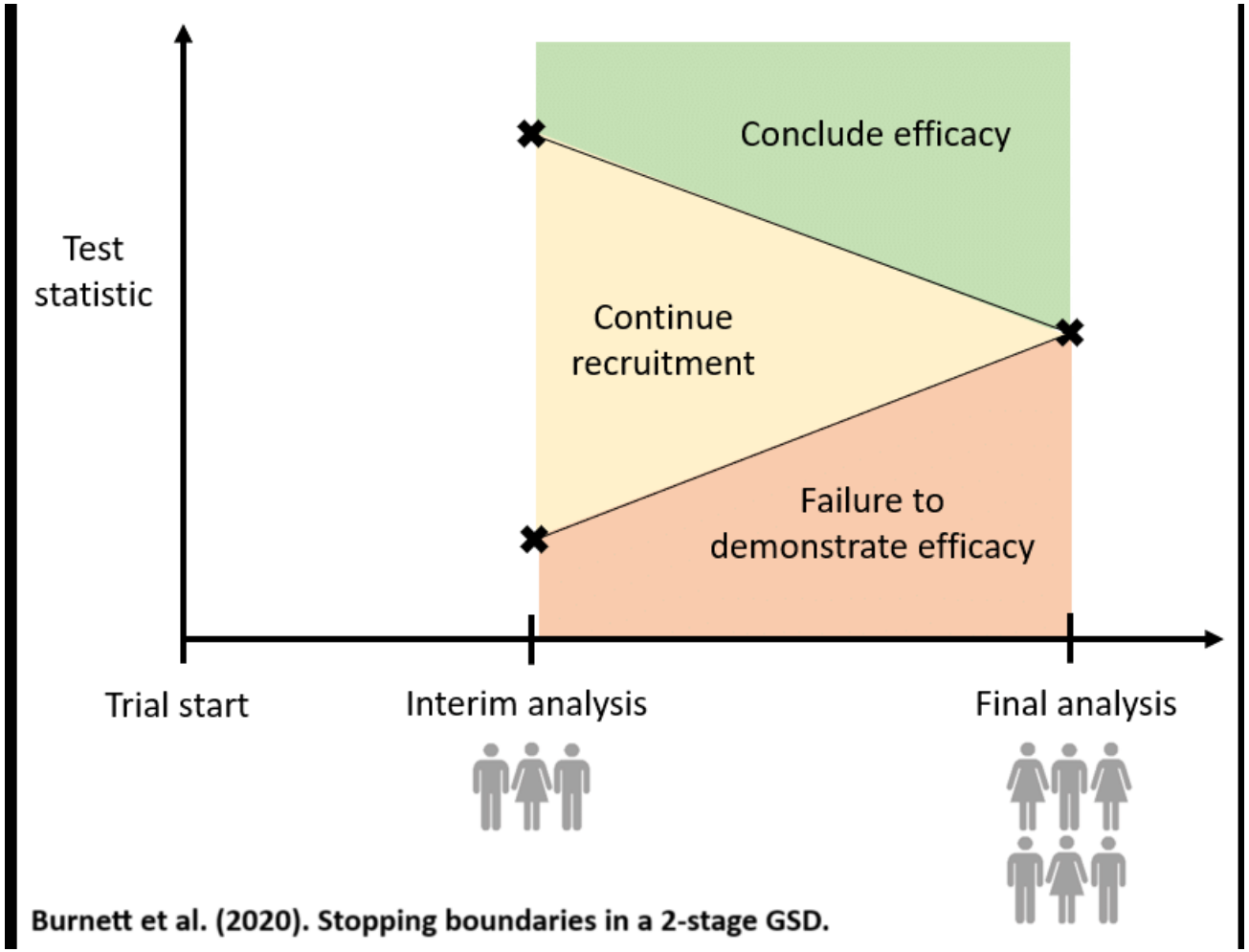
Group sequential designs (GSDs) are clinical trials divided into pre-planned **stages**, with periodic evaluation (**interim phases**) of accumulating data between stages.

At each interim, trial can be stopped for **efficacy** or **futility**.

Pros: improved *efficiency* and *ethics*

Cons: interim decisions introduce bias

- Need to control T1E rate
- Need to account for estimation bias and incorrect interval coverage (e.g., occasionally optimistic early results)



Burnett et al. (2020). Stopping boundaries in a 2-stage GSD.

MRC Biostatistics Unit

UNIVERSITY OF CAMBRIDGE

Likelihood decomposition

Let

- $\boldsymbol{x}_{(s)}$ be the vector of **responses** collected up to interim s ;
- $\boldsymbol{d}_{(s)}$ be the vector of **interim decisions** made up to interim s.

Following Marschner (2021), we can consider three types of likelihood functions:

- Full (**unconditional**) likelihood: $L_U\left(\theta; \boldsymbol{x}_{(s)}, \boldsymbol{d}_{(s)}\right) \propto \prod_{i=1}^{n_{(s)}} f_\theta(x_i) = L_U(\theta; \boldsymbol{x}_{(s)})$

  invariant wrt decisions

- **Decision** likelihood: $L_D(\theta; \boldsymbol{d}_{(s)}) = P\left(\boldsymbol{D}_{(s)} = \boldsymbol{d}_{(s)} | \theta\right)$

- **Conditional-on-decision** likelihood: defined by the decomposition

$$L_U(\theta; \boldsymbol{x}_{(s)}) = L_D(\theta; \boldsymbol{d}_{(s)}) \cdot L_C(\theta; \boldsymbol{x}_{(s)} | \, \boldsymbol{d}_{(s)})$$

Treating interim decisions as observed data

In Bayesian inference, posterior distributions should encapsulate all the information collected during the trial.

Flournoy and Tarima (2023) advocate incorporating informative stopping decisions $\boldsymbol{d}_{(s)}$ in the Bayesian posteriors as **part of the observed data**:

$$\pi_C\left(\theta \middle| \boldsymbol{x}_{(s)}, \boldsymbol{d}_{(s)}\right) \propto \pi(\theta) \cdot L_C(\theta; \boldsymbol{x}_{(s)} | \, \boldsymbol{d}_{(s)})$$

This can be contraposed to the "traditional" posterior which only conditions on $\boldsymbol{x}_{(s)}$:

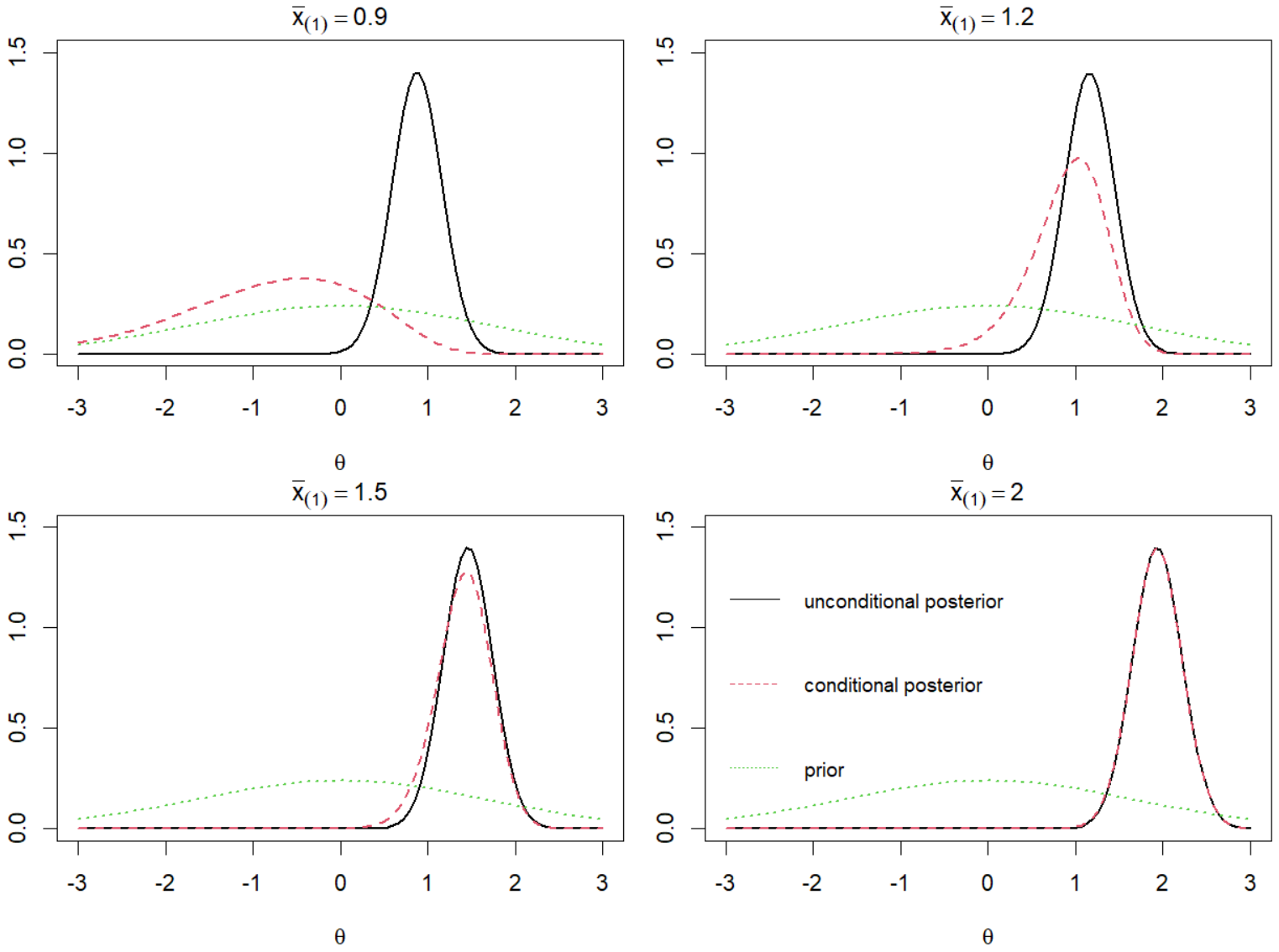
$$\pi_U\left(\theta \middle| \boldsymbol{x}_{(s)}\right) \propto \pi(\theta) \cdot L_U(\theta; \boldsymbol{x}_{(s)})$$

Illustration: normal endpoint with unknown mean

Trial was **stopped** after crossing the efficacy boundary of interim 1: $\bar{x}_{(1)} > 0.85$

Prior *10 times less informative* than stage-1 sample

Conditional-on-decision posterior **accounts for uncertainty on decisions made**



MRC Biostatistics Unit

UNIVERSITY OF CAMBRIDGE

Adaptation-Induced Posterior Divergence (AIPD)

The Kullback-Leibler (KL) divergence of π_C from π_U for a given $\boldsymbol{x}_{(s)}$,

$$D_{\boldsymbol{d}_{(s)},\boldsymbol{x}_{(s)}}(\pi_U||\pi_C) = \int \pi_U\left(\theta\middle|\boldsymbol{x}_{(s)}\right) \boldsymbol{log} \frac{\pi_U\left(\theta\middle|\boldsymbol{x}_{(s)}\right)}{\pi_C\left(\theta\middle|\boldsymbol{x}_{(s)},\boldsymbol{d}_{(s)}\right)}\, d\theta$$

is a **distribution-level metric** that summarizes **how posterior beliefs on θ have been modified by the adaptive decision rules** wrt a fixed-sample design (FSD).

N.B. $\pi_U\left(\theta\middle|\boldsymbol{x}_{(s)}\right)$ can be seen as a ***counterfactual*** posterior, i.e. the posterior we would have obtained if the data were collected under FSD.

Pre-experimental planning: expected AIPD

We can also define a **pre-experimental version** of this metric to assess the **expected impact of future adaptations** on the end-of-trial posterior inference:

$$\overline{D}(\theta) = E_{\boldsymbol{x}_{(s^*)}|\theta}\left[D_{\boldsymbol{d}_{(s^*)},\boldsymbol{x}_{(s^*)}}(\pi_U||\pi_C)\right],$$

where $\boldsymbol{x}_{(s^*)}$ is the sample collected up to the *stopping stage* s^*.

Given a set of efficacy (or futility) boundaries, we can compute the associated value of $\overline{D}(\theta)$ for all values of θ, similar as for the expected sample size (ESS).

Illustration: 3-stage GSD for treating depression

- **Superiority trial** (Dmitrienko and Koch, 2017): treatment vs placebo
- **Endpoint:** 8-week mean reduction in depression score, $X_i \sim N(\delta, 1)$
- **Weakly informative prior:** $\delta \sim N(0, 5)$

We want to compare *shapes of efficacy boundaries* and *interim schedules* based on

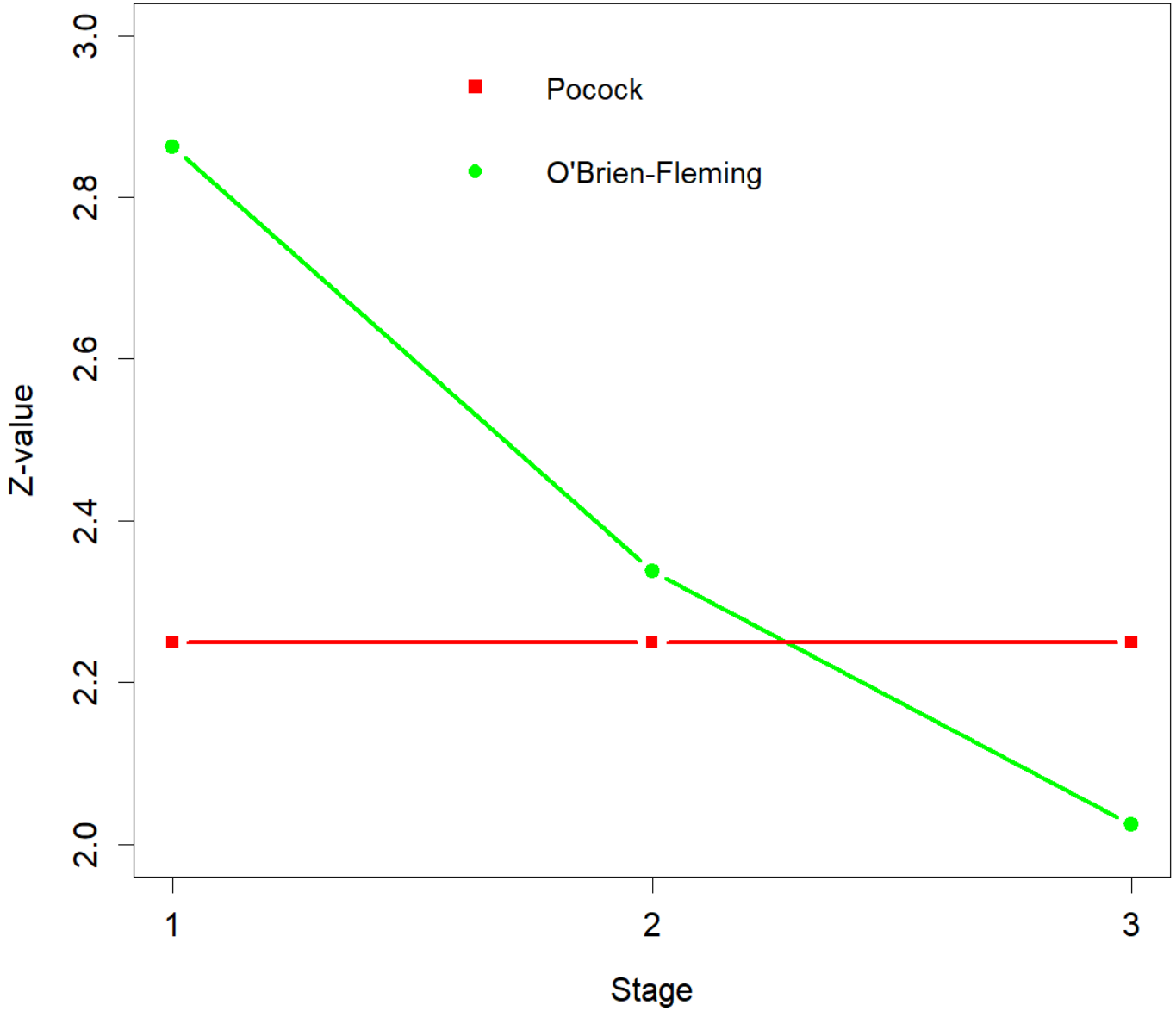
- Expected Sample Size, $\mathrm{ESS}(\delta)$ $\longrightarrow$ trial's efficiency
- Expected AIPD, $\overline{D}(\delta)$ $\longrightarrow$ robustness of posterior inference to adaptation

Shapes of efficacy stopping boundaries

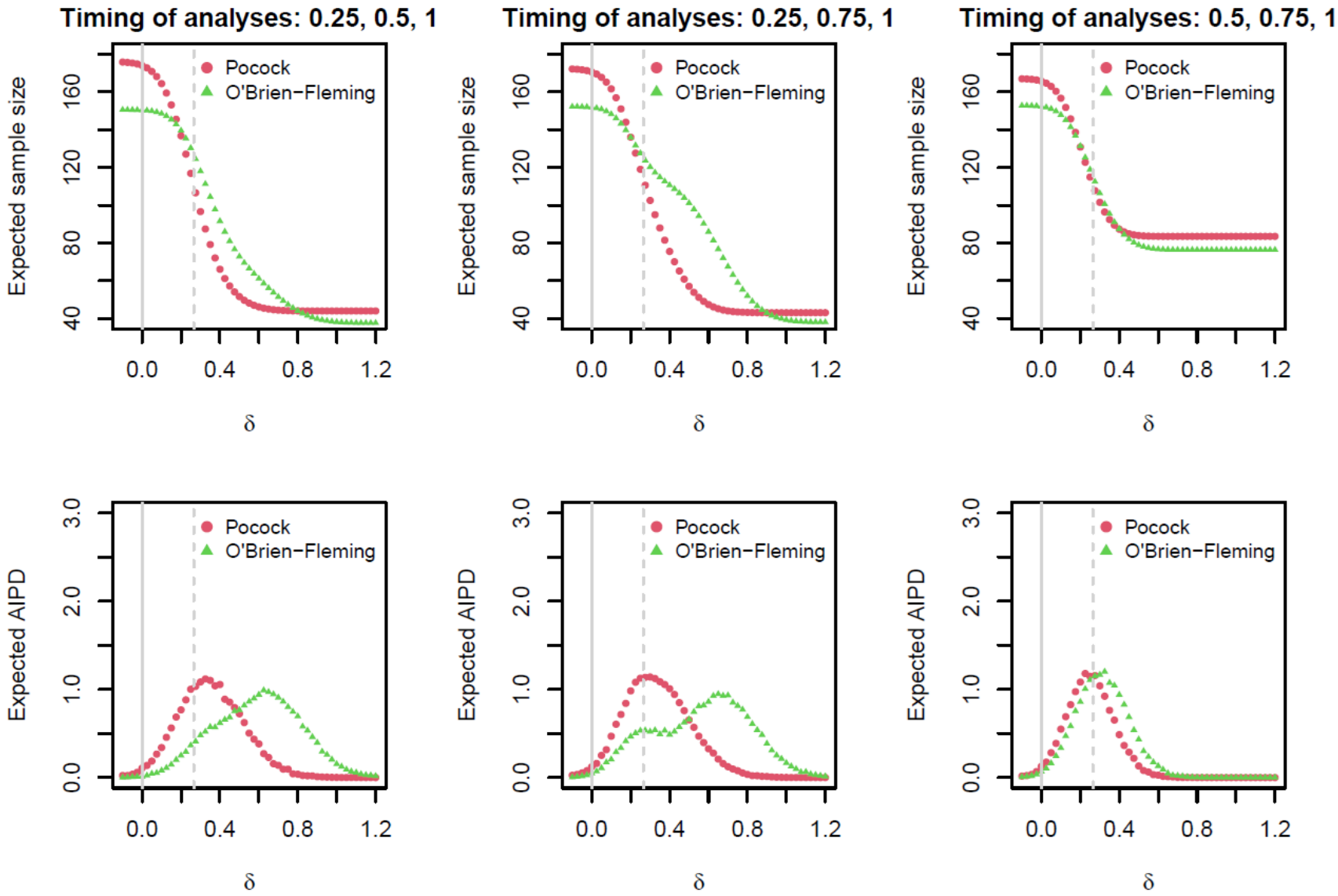
For **fixed overall T1E rate**, we can have different sets of efficacy boundaries, e.g.,

- **Pocock (1977):** same chance to stop early after each stage.
- **O'Brien-Fleming (1979):** protects against premature stopping.

Key difference is how the overall T1E rate is *spent* across stages.



- Little difference between boundaries when first interim is delayed (~50% of patients)
- OBF preferable when the effect is small ($\delta < 0.25$)
- Pocock preferable when larger effects ($\delta > 0.5$) are plausible
- Area under expected AIPD curve summarizes average performance across all δ



MRC Biostatistics Unit

UNIVERSITY OF CAMBRIDGE

Error rates: $\alpha = 2.5\%$ under $H_0: \delta = 0$ and $1 - \beta = 90\%$ under $H_1: \delta = 0.265$

Final remarks

Takeaways:

- Interim decisions carry information that may change Bayesian estimates compared to fixed-sample trials.
- We proposed a global metric of how interim decisions alter the posterior inference.
- A pre-experimental version of this measure can guide choice of boundaries and schedules of interim analyses.

Extensions:

- analyze interaction between prior and design (in the arXived manuscript).
- study more complex endpoints and trial designs.
- decision-theoretic framework to formally balance the trade-off between ESS and expected AIPD.

Main references

Burnett, T., Mozgunov, P., Pallmann, P., Villar, S. S., Wheeler, G. M., & Jaki, T. (2020). Adding flexibility to clinical trial designs: an example-based guide to the practical use of adaptive designs. *BMC medicine, 18(1), 352*.

Flournoy, N. and Tarima, S. (2023). Posterior alternatives with informative early stopping. *Statistical papers, 64*.

Marschner, I. C. (2021). A general framework for the analysis of adaptive experiments. *Statistical Science, 36(3).*

O'Brien, P. C. and Fleming, T. R. (1979). A multiple testing procedure for clinical trials. *Biometrics, 35.*

Pocock, S. J. (1977). Group sequential methods in the design and analysis of clinical trials. *Biometrika, 64.*

Robertson, D. S., Choodari-Oskooei, B., Dimairo, M., Flight, L., Pallmann, P., & Jaki, T. (2023). Point estimation for adaptive trial designs I: a methodological review. *Statistics in medicine, 42(2)*.

Robertson, D. S., Burnett, T., Choodari-Oskooei, B., Dimairo, M., Grayling, M., Pallmann, P., & Jaki, T. (2025). Confidence intervals for adaptive trial designs I: a methodological review. *Statistics in Medicine, 44(18-19).*

MRC Biostatistics Unit

UNIVERSITY OF CAMBRIDGE

Thank you for your attention!

Link to arXived manuscript

Appendix: alternative interpretation of AIPD

$L_D^{-1}\left(\theta; \boldsymbol{d}_{(\boldsymbol{s})}\right) = L_D^{-1}$ quantifies how *surprising* θ is given the realized design.

AIPD can be decomposed as a *Jensen's gap*:

$$D_{\boldsymbol{d}_{(\boldsymbol{s})}, \boldsymbol{x}_{(\boldsymbol{s})}}(\pi_U || \pi_C) = \log E_{\pi_U}[L_D^{-1}] - E_{\pi_U}[\log L_D^{-1}]$$

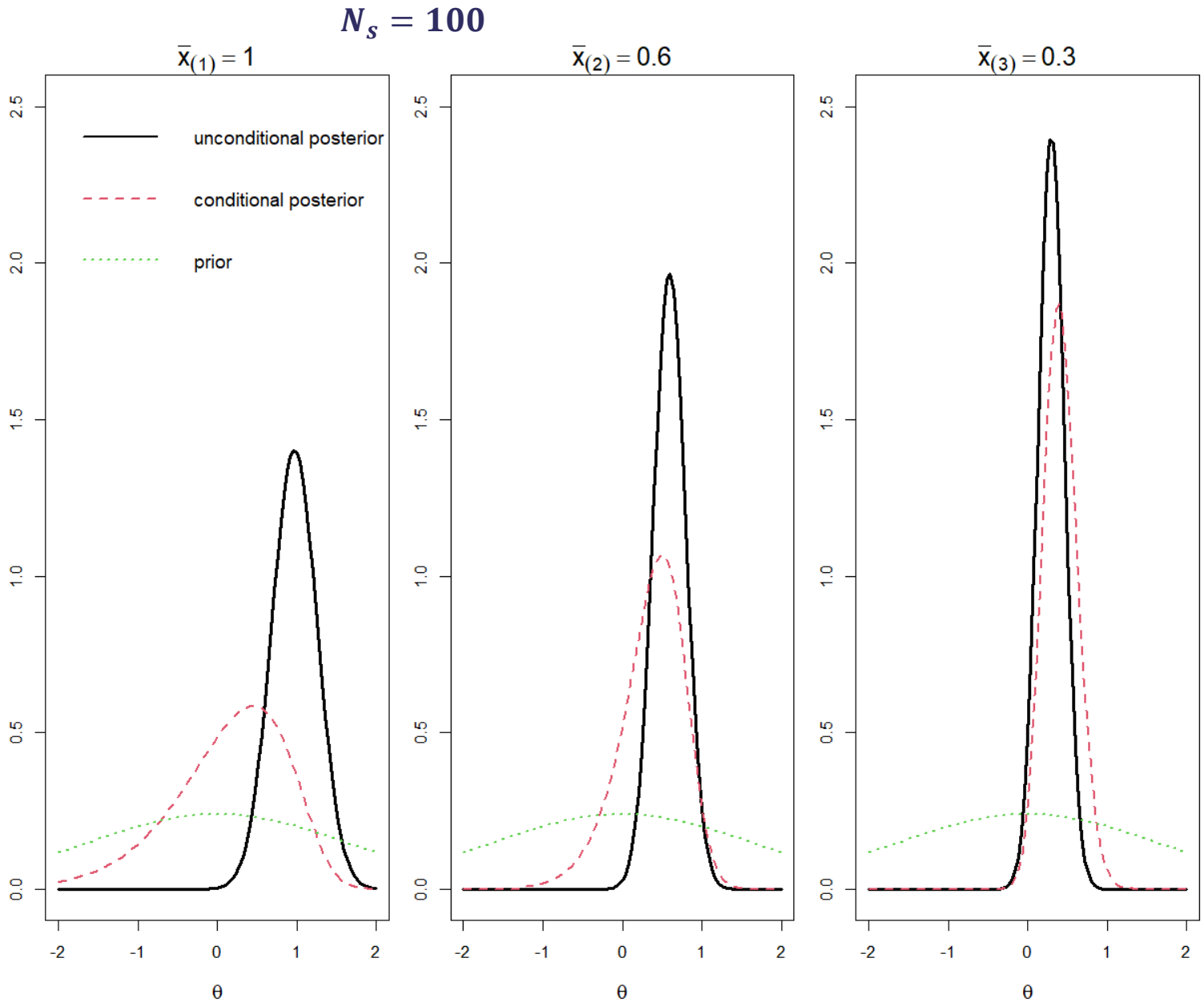
where

- $\log E_{\pi_U}[L_D^{-1}]$ coincides with the Bayes factor of the conditional model relative to the unconditional model;
- $E_{\pi_U}[\log L_D^{-1}]$ averages the measure of surprise on θ wrt its unconditional posterior.

N.B. $D_{\boldsymbol{d}_{(\boldsymbol{s})}, \boldsymbol{x}_{(\boldsymbol{s})}}(\pi_U || \pi_C) \approx \frac{1}{2} V_{\pi_U}[L_D^{-1}] \cdot E_{\pi_U}^{-2}[\log L_D^{-1}]$, thus AIPD is related to the variability of the surprise of θ.

Appendix: 3-stage trial ended for efficacy at interim s

Stopping stage (s*)	1	2	3
Efficacy boundary	0.85	0.43	0.28
$\overline{x}_{(s)}$	1.00	0.60	0.30
AIPD	1.04	0.35	0.19
95% CIs overlapping	31%	51%	73%
Posterior variances ratio (C/U)	2.60	1.99	1.30
Posterior modes diff. (C-U)	-0.52	-0.09	0.09



MRC Biostatistics Unit

UNIVERSITY OF CAMBRIDGE