Assurance and power-like friends in group-sequential designs

Marta Bofill Roig

July 17, 2025

UNIVERSITAT POLITÈCNICA
DE CATALUNYA
BARCELONATECH

Outline

Trial design considerations

- Trial design depends on objectives and statistical parameters.
- **Frequentist design** is based on:
  - Null hypothesis (H_0) and alternative hypothesis (H_1)
  - Type I error: reject H_0 when true (Significance level)
  - Type II error: accept H_0 when false (1−Power)
- Type I and II errors guide test reliability and decision-making.

Power and sample size in trial design

- **Power** is a key element in frequentist trial design.
- After defining the hypothesis, design, and significance level (typically 5%):
  - **Sample size** is calculated to achieve a desired power.
- Power = Probability of rejecting H_0 when the true treatment effect equals a specified value.
- Common power levels: 80% or 90%.
- Higher power is recommended, especially for hard-to-repeat trials.

Specifying the treatment effect

- The treatment effect to be detected is based on:
  - Minimal clinically relevant effect, or
  - Anticipated effect of the new treatment
- True treatment effect may differ from the assumed value.
- **Power does not represent the actual chance of rejecting H_0 in practice.**
- For sponsors, the true probability of trial success is crucial for:
  - Deciding to conduct the trial
  - Choosing the appropriate sample size

So what's the power?

Consider:

$$H_0 : \theta = 0 \qquad \text{vs} \qquad H_1 : \theta > 0$$

Let the event R denote rejection of the null hypothesis.

So what's the power?

Consider:

$$H_0 : \theta = 0 \qquad \text{vs} \qquad H_1 : \theta > 0$$

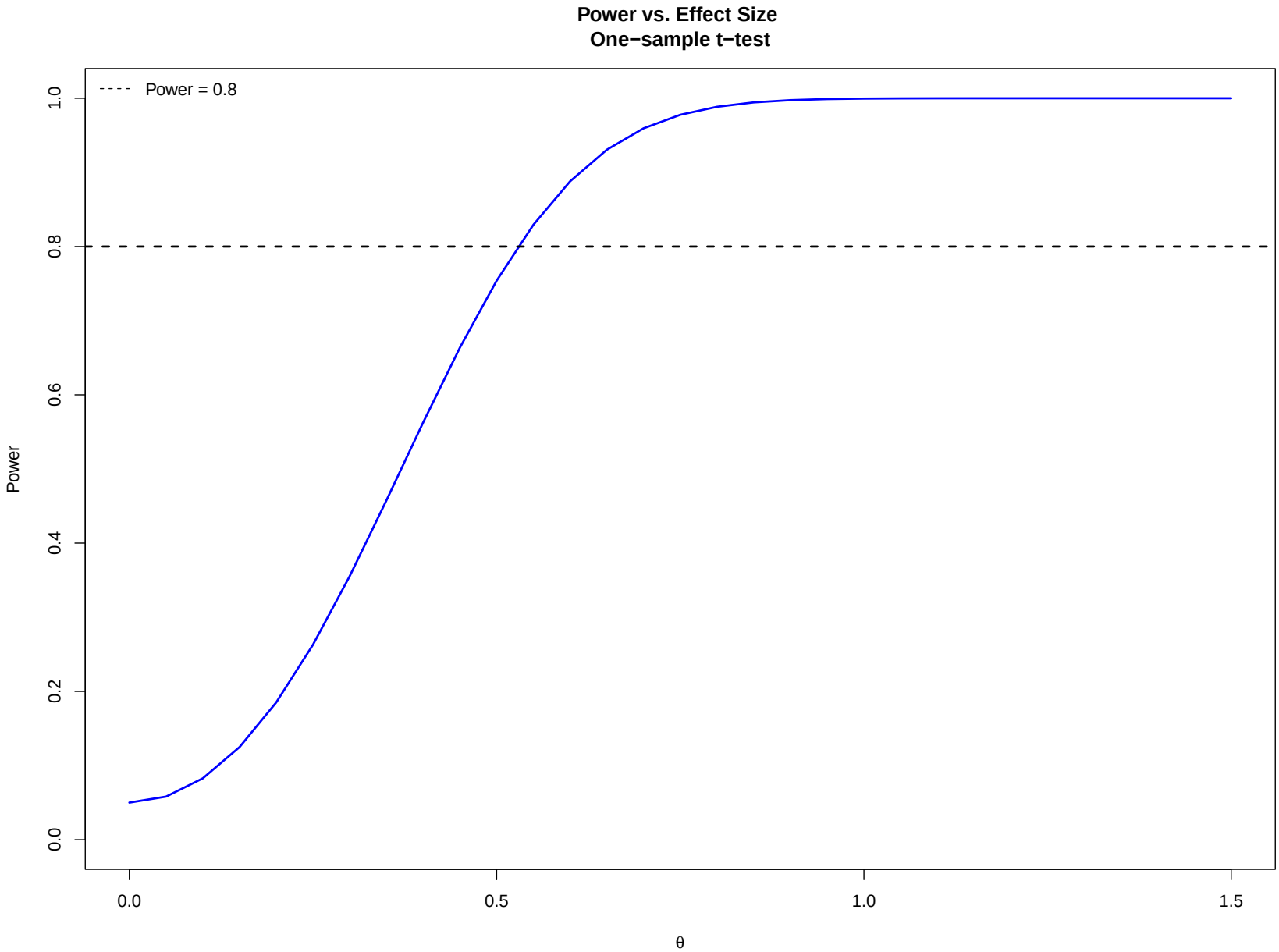
Let the event R denote rejection of the null hypothesis.

The power function is:

$$\pi(\theta) = P(R|\theta)$$

- **Conditional probability of** R, conditional on the unknown true value of the parameters.
- Unknown parameters: true treatment effect and other nuisance parameters.

Power analysis: Power as function of θ



Sample size determination

- The choice of a treatment effect size (θ) is inherently uncertain.
- Other unknown parameters (e.g., sampling variance) are typically fixed using prior estimates.
- Sensitivity analyses assess how assumptions affect sample size.
- Not all parameter values are equally plausible – some are better supported by evidence.
- Use of published data, prior trials, and expert opinion is essential.

Assurance

Assurance[1,2]: An alternative to the concept of power

Since the value of θ is unknown, rather than assigning it a specific value, we assign it a distribution.

[1] O'Hagan, Stevens, and Campbell. "Assurance in clinical trial design." Pharmaceutical Statistics (2005).

[2] Spiegelhalter and Freedman. "A predictive approach to selecting the size of a clinical trial, based on subjective clinical opinion." Statistics in Medicine (1986).

Assurance[1,2]: An alternative to the concept of power

Since the value of θ is unknown, rather than assigning it a specific value, we assign it a distribution.

Assurance is the **unconditional probability** of rejecting H_0, that is:

$$\gamma = P(R) = E(\pi(\theta))$$

[1] O'Hagan, Stevens, and Campbell. "Assurance in clinical trial design." Pharmaceutical Statistics (2005).

[2] Spiegelhalter and Freedman. "A predictive approach to selecting the size of a clinical trial, based on subjective clinical opinion." Statistics in Medicine (1986).

Assurance[1,2]: An alternative to the concept of power

Since the value of θ is unknown, rather than assigning it a specific value, we assign it a distribution.

Assurance is the **unconditional probability** of rejecting H_0, that is:

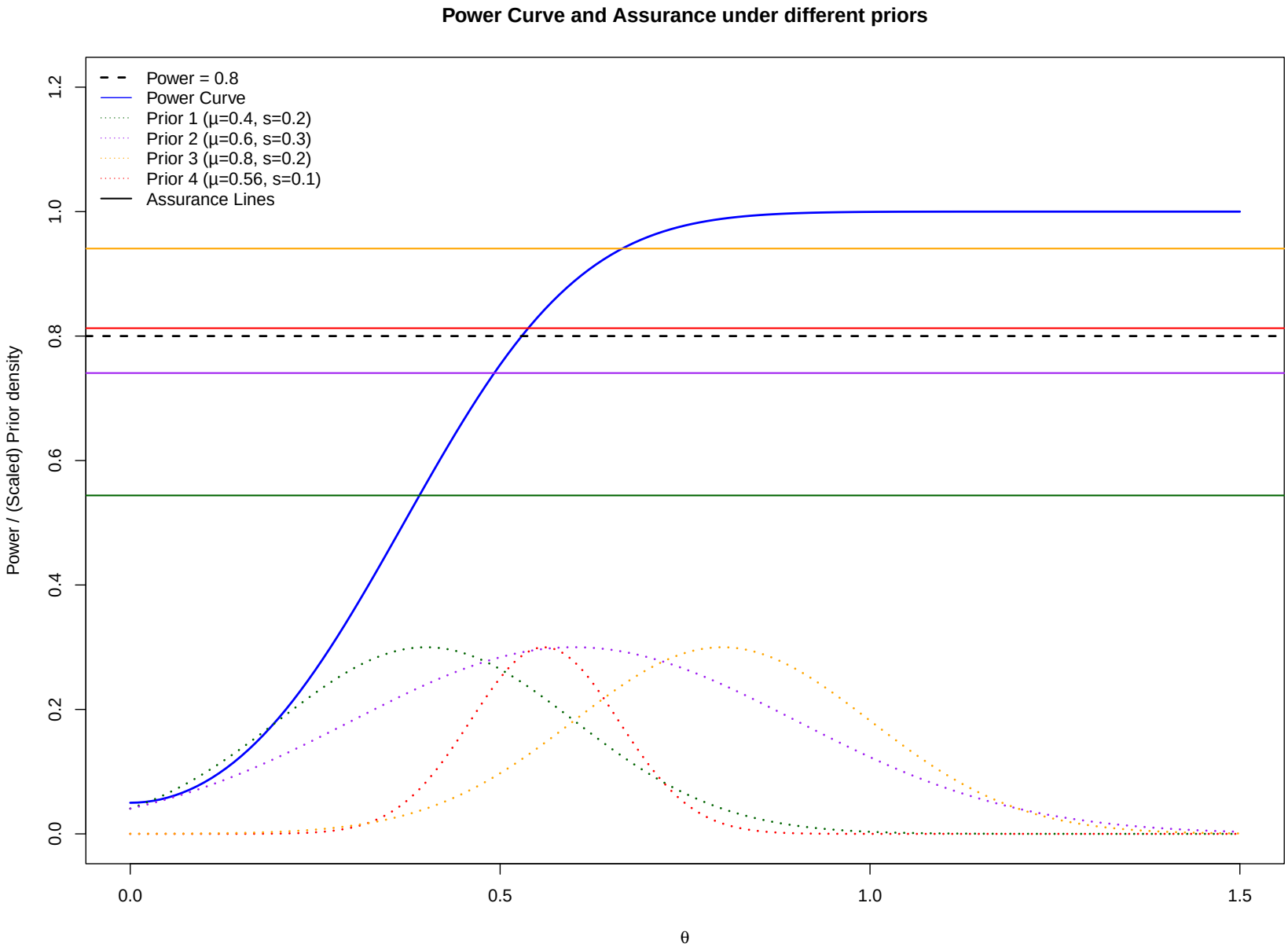
$$\gamma = P(R) = E(\pi(\theta))$$

- Assurance equals the **expected power**, averaged over the prior distribution of θ.
- It differs from power, which is conditional on a fixed value of θ.

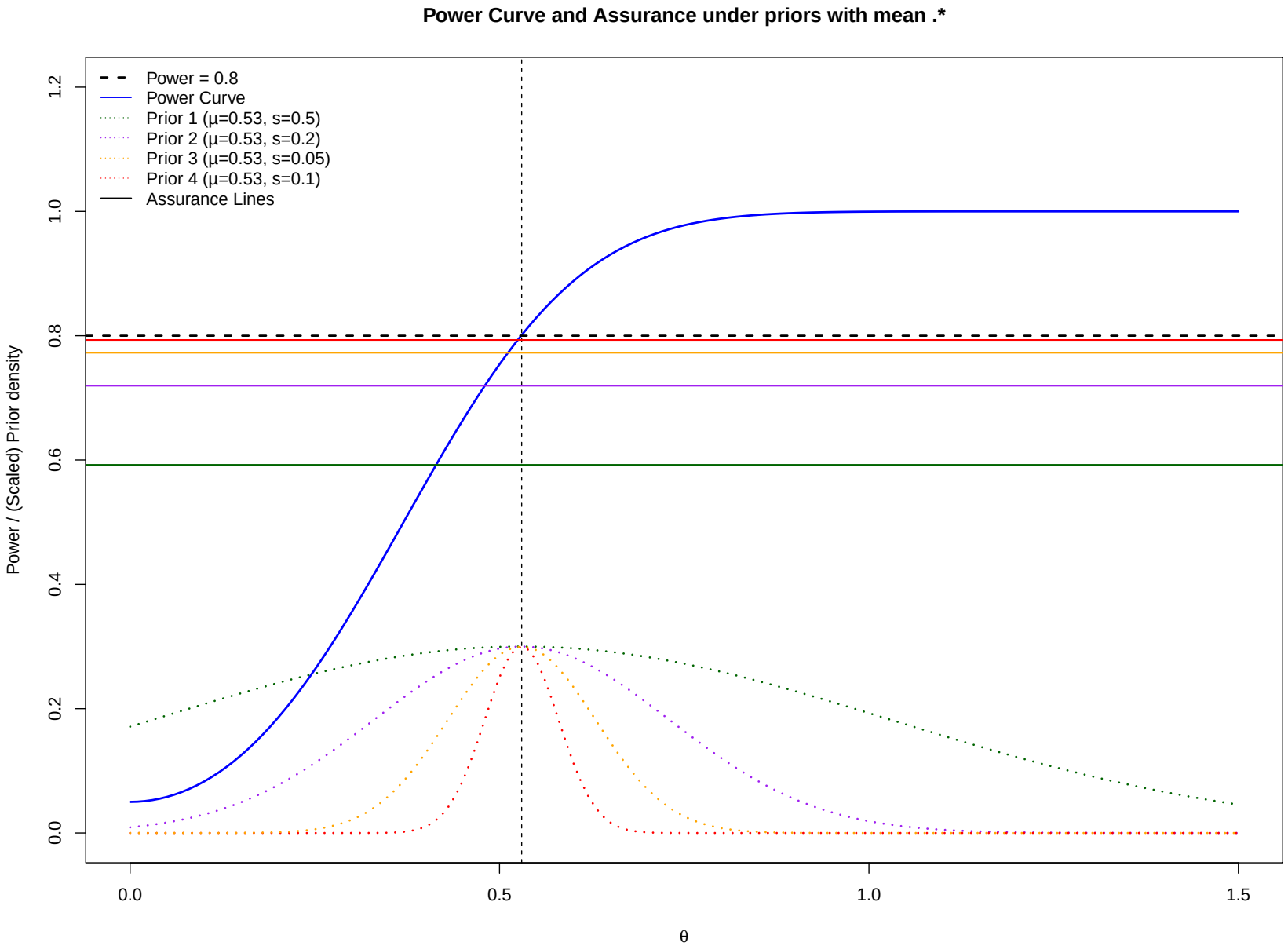
[1] O'Hagan, Stevens, and Campbell. "Assurance in clinical trial design." Pharmaceutical Statistics (2005).

[2] Spiegelhalter and Freedman. "A predictive approach to selecting the size of a clinical trial, based on subjective clinical opinion." Statistics in Medicine (1986).

Power and assurance analysis



Power and assurance analysis (cont'd)

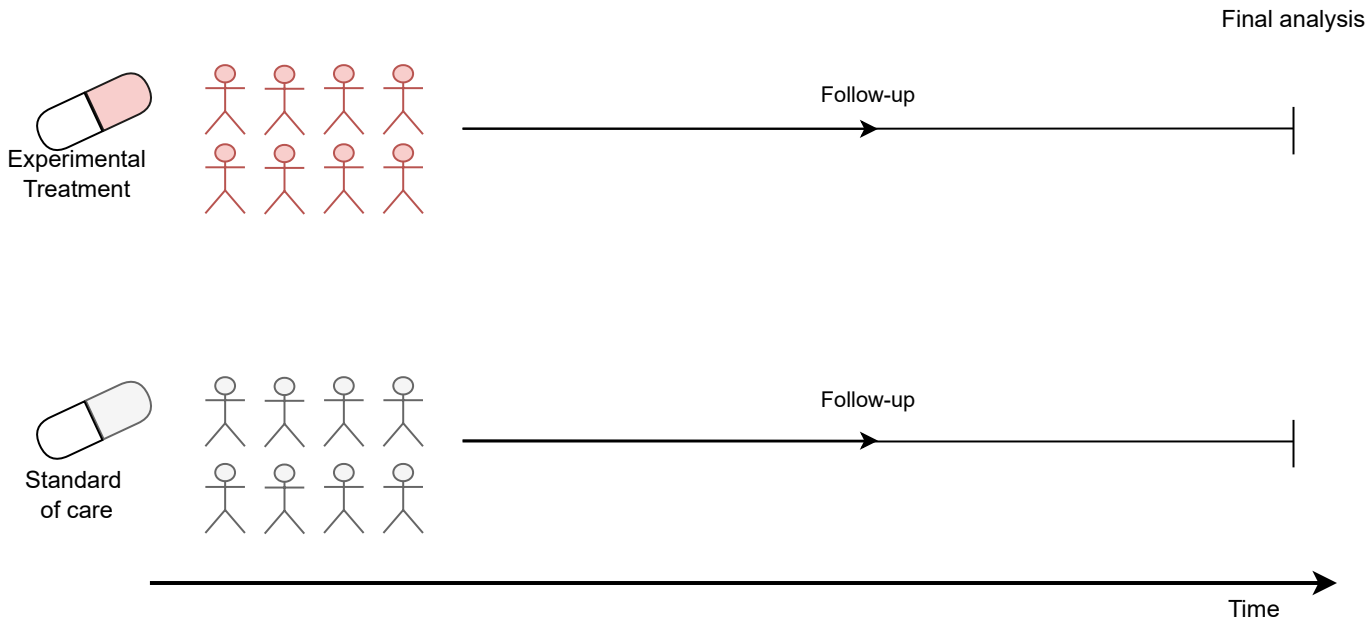


Group-sequential designs

Classical design

Patients are randomised into groups and followed over time until the study concludes.

The **final analysis** is then performed to draw conclusions.

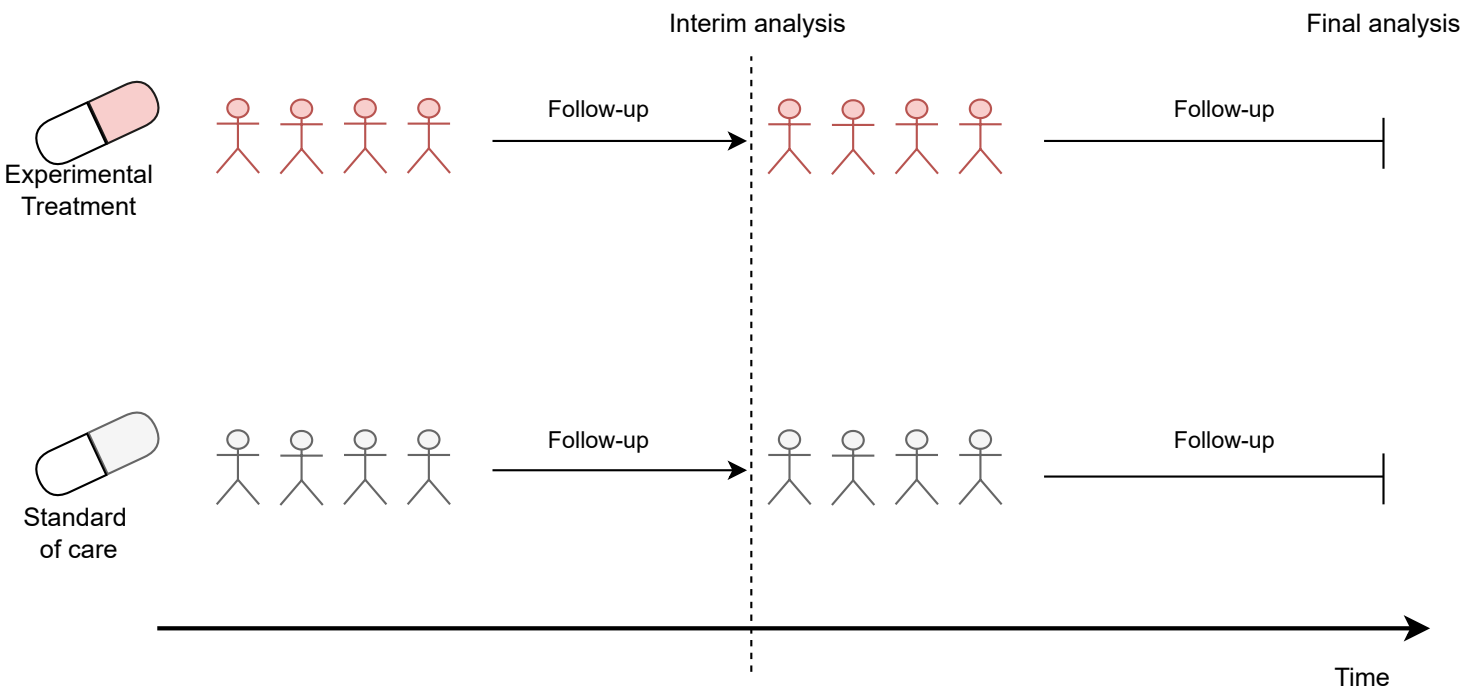


Group-sequential design

Patients are randomised into groups and followed over time until the study concludes.

The study includes an **interim analysis** (analysis based on data collected up to then), based on which an early decision can be made.

If the study continues after the interim analysis, the **final analysis** is then performed to draw conclusions.



Conditional Power

Conditional power is the probability of rejecting H_0 at the final analysis, given the **data** observed up to the interim analysis and given the value of θ.

Thus, it is defined by:

$$\text{CP} = P\left(R \mid D_{\text{int}}, \theta\right)$$

where D_{int} = data observed up to the interim analysis.

Conditional Power

Conditional power is the probability of rejecting H_0 at the final analysis, given the **data** observed up to the interim analysis and given the value of θ.

Thus, it is defined by:

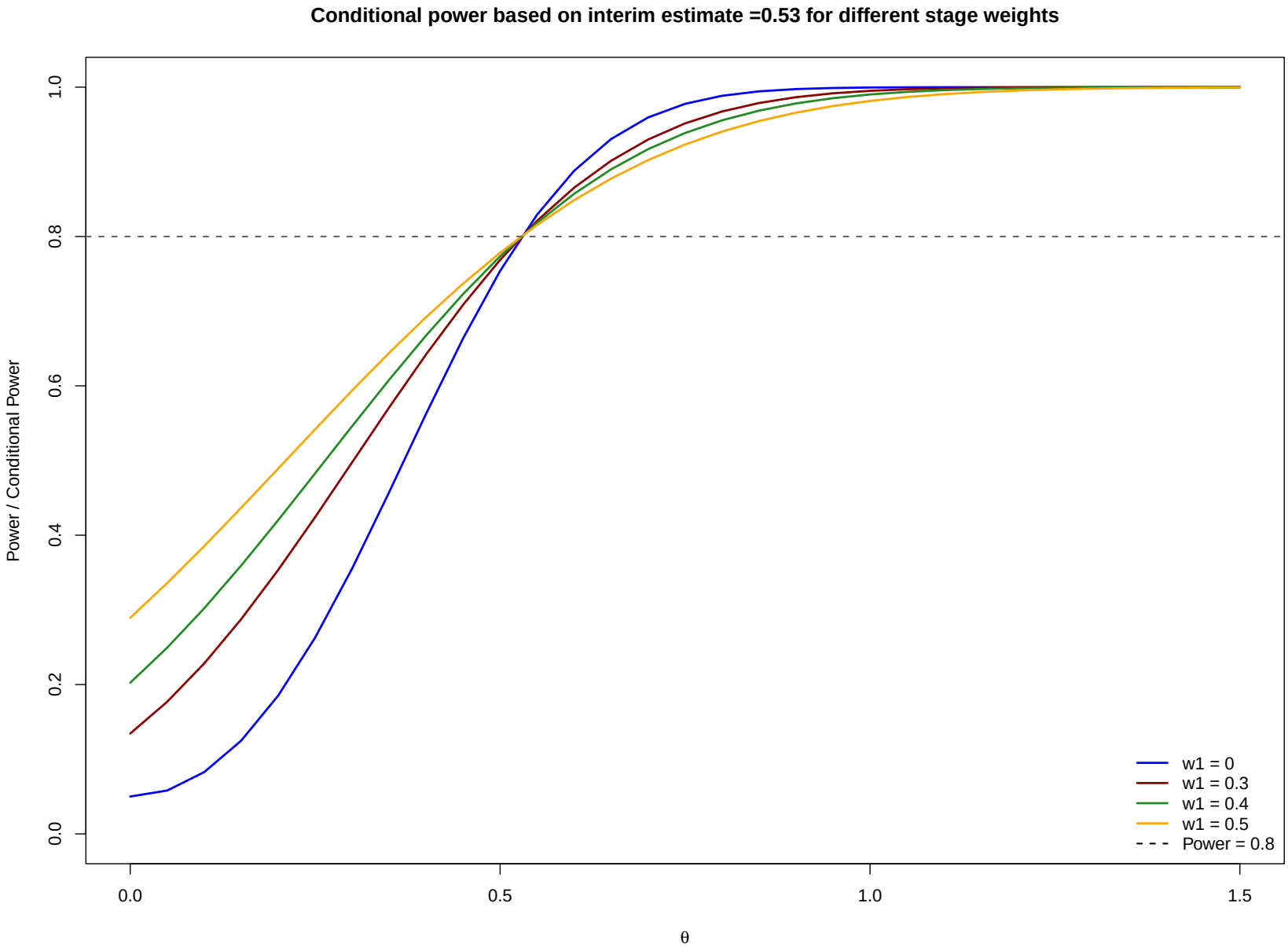
$$\text{CP} = P\left(R \mid D_{\text{int}}, \theta\right)$$

where D_{int} = data observed up to the interim analysis.

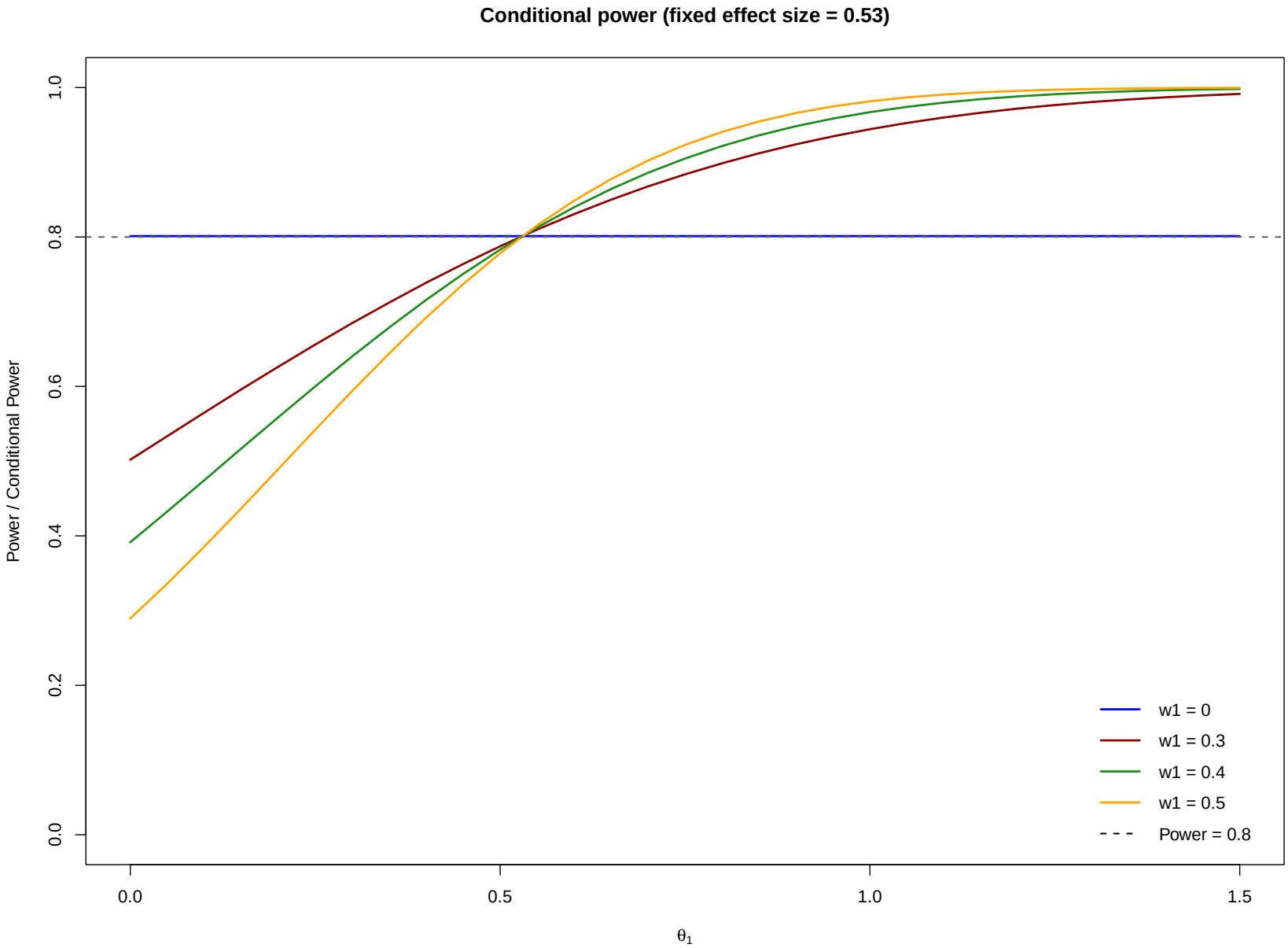
Conditional power then implicitly depends on:

- Denote by $\hat{\theta}_1$ the estimator of θ based on D_{int}.
- Denote by w_1 the proportion of total sample size used at the interim.

Let's get some intuition on what the conditional power is



Let's get some intuition on what the conditional power is



Conditional power and early stopping

- Group-sequential designs allow stopping the trial early due to safety concerns, overwhelming evidence of benefit (*efficacy*), or lack of benefit (*futility*).
- Conditional power quantifies the chance of success based on the data observed at the interim.
- By evaluating the likelihood of eventual success, conditional power provides a quantitative basis for deciding whether to continue or stop a trial early for futility.
- Low conditional power signals a low likelihood of success, supporting early stopping to save resources.

Conclusions and future directions

Conclusions

- There are several criteria to define the probability of rejecting the null hypothesis (or alike), each relying on different known information and underlying assumptions.
- Using assurance and incorporating prior distributions to elicit the treatment effect can be valuable during the design phase, especially when the expected effect size and/or nuisance parameters are uncertain.
- From the sponsor's perspective, this approach helps quantify uncertainty in the probability of success and supports more robust sample size determination.

So what's next? Open questions

- How can we update assurance at interim to assess the chance of success?
- Is it possible to define futility stopping rules based on conditional assurance at interim?
- How do we determine sample size when using assurance instead of traditional power?
- Use a time-to-event study as example to demonstrate how assurance captures uncertainty in critical trial factors –such as recruitment rates and number of events– that affect both the probability of success and trial power.

That's all folks!

UNIVERSITAT POLITÈCNICA
DE CATALUNYA
BARCELONATECH