Mind the gap: Equipoise calibration of clinical trial design



Fabio Rigat, PhD

Executive Director, AstraZeneca Oncology Biometrics

Opinions presented here do not represent AstraZeneca's positions as a whole

Bayes Pharma conference – October 2026

Overview

- This work is based on a Pharm Stat paper, a JCB paper, a preprint and work in progress
- Constructive feedback and questions at the end of this talk will be welcome.



Current trial design is inefficient and insufficient

- **Trial outcomes are mostly defined by frequentist OCs:** e.g.

$$+_N := 1_{\{HR < HR_{CV}(N)\}},\ -_N := (1 - +_N)\ \ s.t. p(+_N |\ H_0) \leq \alpha, p(+_N |\ H_1) \geq 1 - \beta$$

- **This practice is *inefficient* and *insufficient* to deliver practice-changing outcomes** because:
  1. Frequentist OCs do not quantify the probability of clinical hypotheses,
  2. Statistically positive trials are not associated to clinically relevant CVs by design,
  3. Frequentist OCs are insufficient to defining an optimal N associated with clinically relevant CVs.

- **This talk discusses the following partial solutions**:
  - Designs establishing a target level of *clinical equipoise imbalance* (Bayesian) define adequate power (frequentist) in relation to a clinically meaningful objective
  - *Design utility* provides a formal basis for sample size optimisation based on probability boundaries *and* on the magnitude of detectable effects



Target clinical equipoise imbalance calibrates the sample size

- **Clinical equipoise***[1,2] (CE) is "*uncertainty within the expert medical community about the preferred treatment*", that is a direct statement about the relative probability of clinical hypotheses H_0 vs H_1
  - **Trial design under CE**: "*the trial must be designed to make it reasonable to expect that, if it is successfully concluded, clinical equipoise will be disturbed*"[3], that is, the trial outcome needs reducing uncertainty about the truth of H_0 vs H_1
  - **CE prior**:

$$P(H_0) \sim U(0,1), P(H_1) := 1 - P(H_0) \Rightarrow \frac{P(H_0)}{P(H_1)} \sim BP(1,1)^4 \Rightarrow F_{BP(1,1)}(x) = \frac{x}{1+x} \Rightarrow Q_{50} = 1,$$

A uniform distribution of pre-study confidence in the truth of H_0 within the population of clinical investigators implies clinical equipoise, consisting of 50% pre-study population odds in favour of H_0 vs H_1

- **Sponsor's design prior:** let $\frac{P(H_0)}{P(H_1)} = \mathrm{r}$ be the sponsor's prior odds. Their posterior odds will be:

$$\frac{P(H_1|\mathrm{r},+_N)}{P(H_0|\mathrm{r},+_N)} = \frac{1}{\mathrm{r}} \times \frac{p(+_N|\,H_1)}{p(+_N|\,H_0)}\,;\ \frac{P(H_0|\mathrm{r},-_N)}{P(H_1|\mathrm{r},-_N)} = \mathrm{r} \times \frac{p(-_N|\,H_0)}{p(-_N|\,H_1)}.$$

- **OCs calibration:** Given r , the *frequentist* OCs of the trial outcome need to be calibrated to establishing a pre-specified CE imbalance $\pi \gg 50\%$, quantified as the *population* CE prior percentile (*Bayesian* measure of surprise), e.g.:

$$N^*(\mathrm{r}, \boldsymbol{\alpha}, \pi) := \underset{N}{\mathrm{argmin}} \left\{ min\left(F_{BP(1,1)}\left(\frac{P(H_1|\mathrm{r},+_N)}{P(H_0|\mathrm{r},+_N)}\right), F_{BP(1,1)}\left(\frac{P(H_0|\mathrm{r},-_N)}{P(H_1|\mathrm{r},-_N)}\right)\right) > \pi, s.t. p(+_N|\,H_0) \leq \boldsymbol{\alpha} \right\}$$

* clinical equipoise is also referred to as the uncertainty principle

CE imbalance depends on the pre-specified sponsor's prior

Let $p(+_N|\,H_0) = 5\%,\ p(+_N|\,H_1) = 90\%$ and $\frac{P(H_1|\mathrm{r},+_N)}{P(H_0|\mathrm{r},+_N)} = \frac{1}{\mathrm{r}} \times \frac{p(+_N|\,H_1)}{p(+_N|\,H_0)}$; $\frac{P(H_0|\mathrm{r},-_N)}{P(H_1|\mathrm{r},-_N)} = \mathrm{r} \times \frac{p(-_N|\,H_0)}{p(-_N|\,H_1)}$

1. **Joint equipoise**: *Under sponsor's and population equipoise ($r = 1$), a primary analysis outcome carrying 5% false positive error probability and 90% power establishes* $\geq 90\%$ CE imbalance

$$\frac{P(H_1|r=1,+_N)}{P(H_0|r=1,+_N)} = 18, \frac{P(H_0|r=1,-_N)}{P(H_1|r=1,-_N)} = 9.5 \Rightarrow F_{BP(1,1)}\,(18) = \frac{18}{1+18} = 94.7\%,\ F_{BP(1,1)}\,(9.5) = \frac{9.5}{1+9.5} = \mathbf{90.4}\%,$$

Q: is this what we (implicitly) do now?

2. **Optimistic prior**: *Under $r = 1/10$, a primary analysis outcome carrying 5% false positive error probability and 90% power does not establish CE imbalance if the outcome is negative*

$$\frac{P(H_1|r=0.1,+_N)}{P(H_0|r=0.1,+_N)} = 180, \frac{P(H_0|r=0.1,-_N)}{P(H_1|r=0.1,-_N)} = \mathbf{0.95} \Rightarrow F_{BP(1,1)}\,(180) = \frac{180}{1+180} > 99\%,\ F_{BP(1,1)}\,(0.95) = \frac{0.95}{1+0.95} \approx \mathbf{49}\%,$$

Q: do negative outcomes matter?

3. **Pessimistic prior**: *Under $r = 10$, a primary analysis outcome carrying 5% false positive error probability and 90% power establishes weak CE imbalance if the outcome is positive*

$$\frac{P(H_1|r=10,+_N)}{P(H_0|r=10,+_N)} = 1.8, \frac{P(H_0|r=10,-_N)}{P(H_1|r=10,-_N)} = 95 \Rightarrow F_{BP(1,1)}\,(1.8) = \frac{1.8}{1+1.8} > \mathbf{60}\%,\ F_{BP(1,1)}\,(95) = \frac{95}{1+95} > 98\%,$$

Q: what OCs *should Ph3s have when Ph2s are negative?*

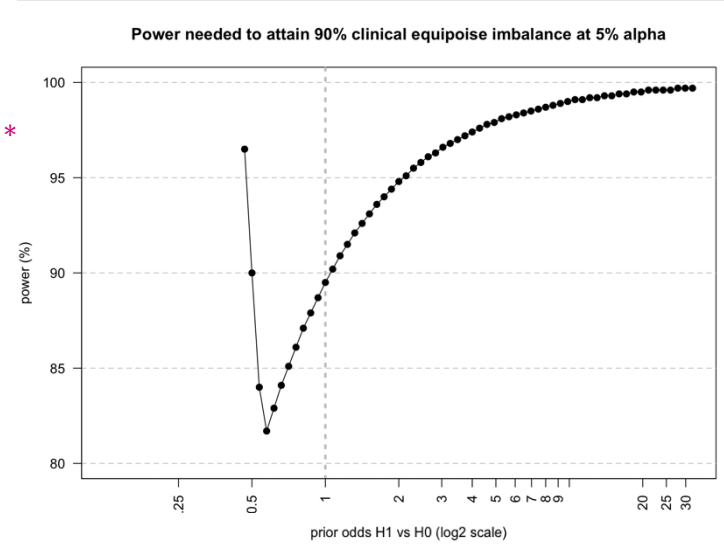
Reverse Bayes: $\pi = 1 - \alpha$ incentivizes the sponsor not to deviate from equipoise

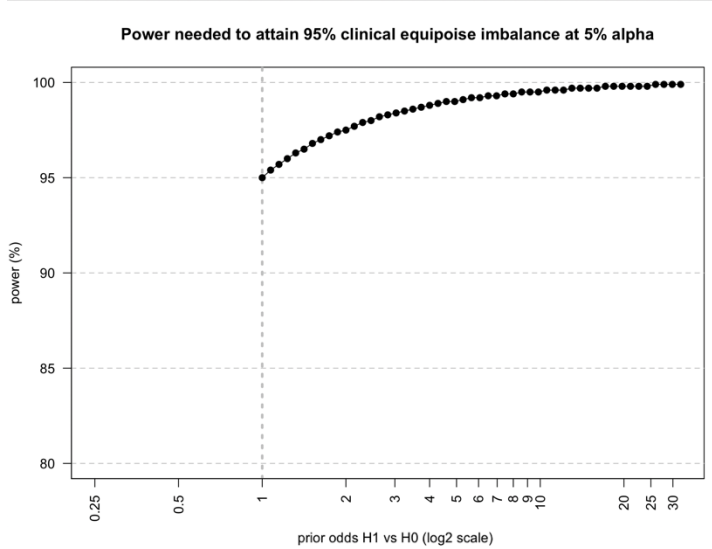
Let $r^* := \underset{N}{\text{argmin}}\{N^*(r, \boldsymbol{\alpha}, \pi), r > 0, given\ \boldsymbol{\alpha}, \pi\}$

$\pi = 90\%, \alpha = 5\%$

Pessimistic r^* minimizes N^*

$r \sim 0.4$ minimize N^* demonstrating 90% CE imbalance (N^* is associated with ~82% power)





$\pi = 95\%, \alpha = 5\%$

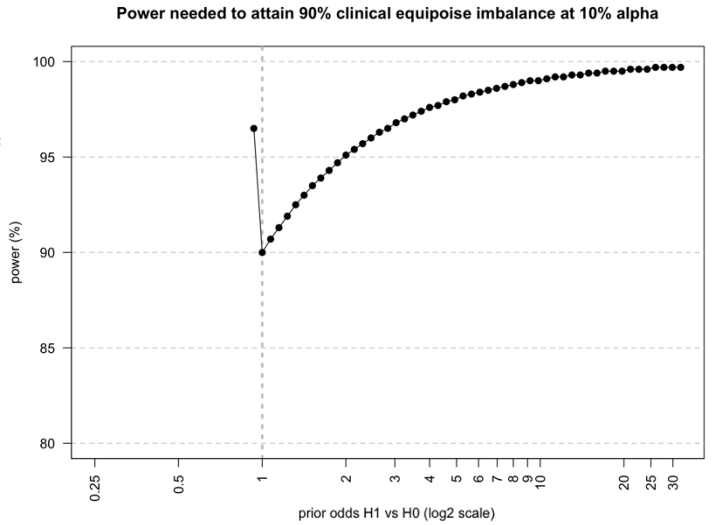
Equipoise r^*minimizes N^*

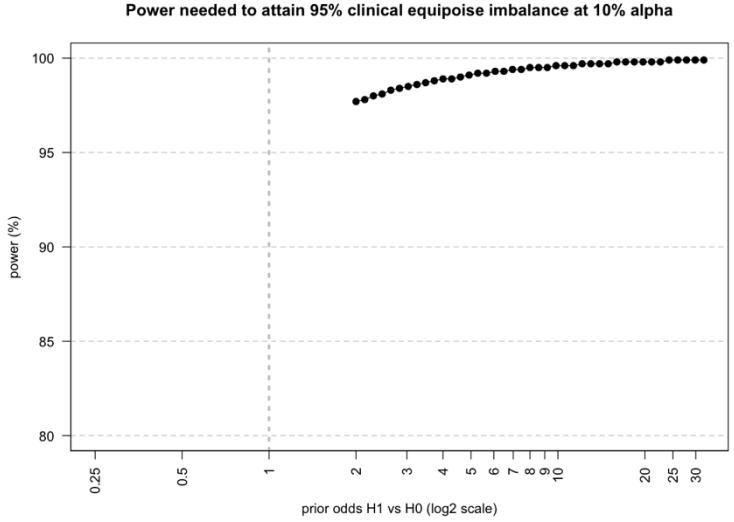
$r = 1$ (1:1 sponsor's prior odds) minimize N^* demonstrating 95% CE imbalance (N^* is associated with 95% power)

$\pi = 90\%, \alpha = 10\%$

Equipoise r^* minimizes N^*

$r = 1$ minimize N^* demonstrating 90% CE imbalance (N^* is associated with 90% power)





$\pi = 95\%, \alpha = 10\%$

Optimistic r^*minimizes N^*

$r \sim 2$ minimize N^* demonstrating 95% CE imbalance (N^* is associated with >95% power)

CE calibration is insufficient for sample size optimization

- If $p(+_N|\ H_0) = \alpha\ \forall\ N$, as $N \to \infty$, $\frac{P(H_1|r,+_N)}{P(H_0|r,+_N)} = \frac{1}{\mathrm{r}} \times \frac{p(+_N|\ H_1)}{\alpha} \to \frac{1}{r\alpha}$ and $MB_N := 1 - HR_{CV}(N) \to 0$,

  $\Rightarrow N$ maximizing $\frac{P(H_1|r,+_N)}{P(H_0|r,+_N)}$ is ∞, but $MB_\infty = 0 \Rightarrow \frac{P(H_1|r,+_N)}{P(H_0|r,+_N)}$ is insufficient to defining an *optimal* N

- Additional conditions are needed to define an optimal N associated to clinically meaningful CVs
  - A pre-specified minimum MB_N may not be consistently agreed upon, depending on setting/indication
  - *Indication agnostic* criteria are desirable ***as a basis* for N optimization**, e.g. maximise a *design utility** :

$$N^u := argmax_N \left\{ \frac{P(H_1|r,+_N)}{P(H_0|r,+_N)} \times MB_N \right\}.$$

  - If $r = 1$ and $p(+_N|\ H_0) = \alpha \forall\ N$, $N^u := argmax_N\{p(+_N|\ H_1) \times MB_N\}$ maximises <u>*expected*</u> utility

Maximising expected design utility balances the increase in power at larger Ns against the reduction in the minimum clinical benefit associated with a statistically positive trial outcome

** multilinear functions fulfill Von Neumann's and Morgenstern utility axioms, e.g. see Keeney and Raiffa 1976*

GS design U-max N is CE calibrated and preserves HR CV < 0.8

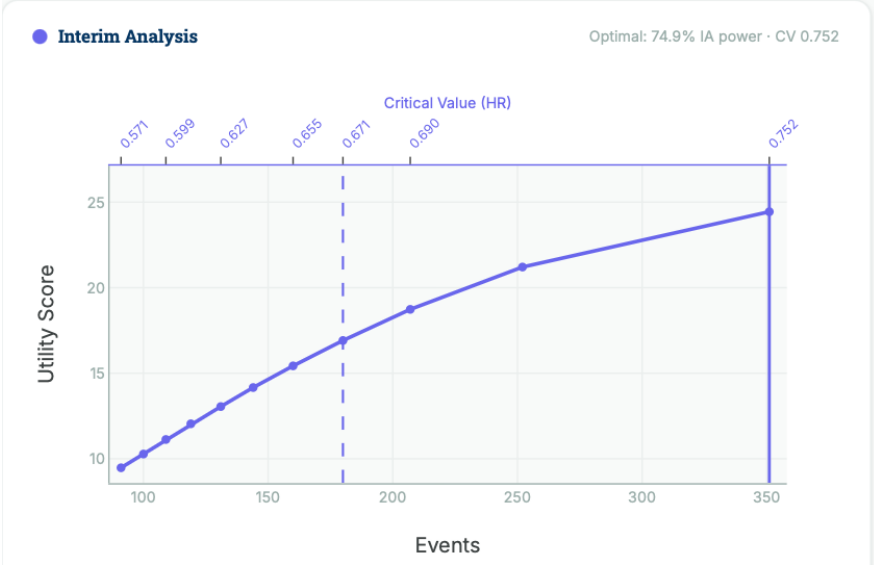
https://gs-intersect.vercel.app/

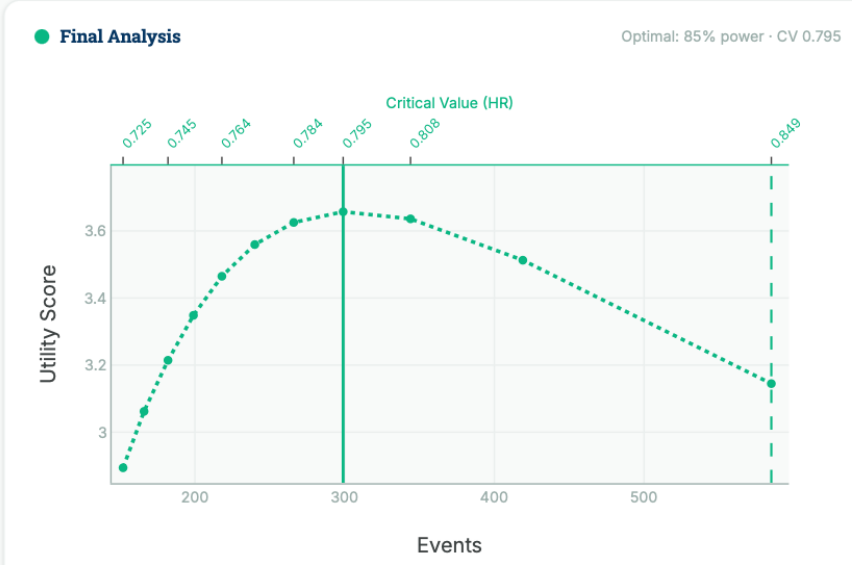
Design assumptions:

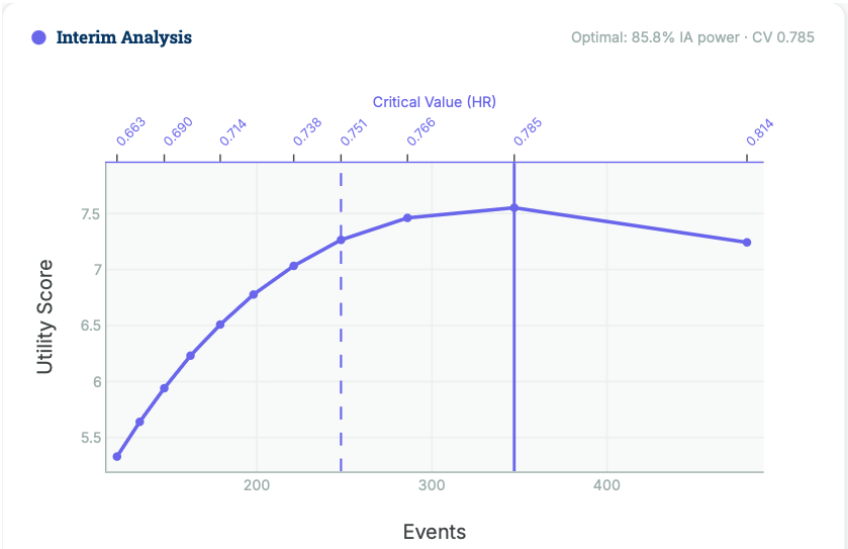
- Median TTE SOC = 12mo
- H_1: HR = 0.7 (median TTE INV = 17mo)
- 18mo enrolment
- Min follow-up 12mo, 5% annual drop-out
- Random right censoring
- Primary analysis: GS (OBF), IA, FA

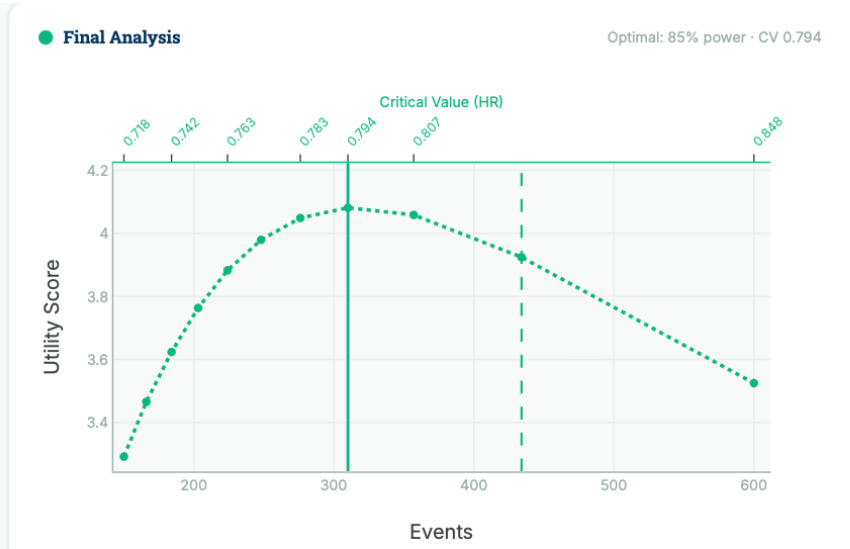
N^u:

1. requires target analysis for optimization: $N^u s$ optimizing IA and FA are different, with N^u for IA at 80% IF $>$ for FA,
2. discourages early looks: IA 50% IF is u-suboptimal, IA 80% IF has a u-maximizing N
3. FA N^u associated with 85% power ,it is CE calibrated at $\pi > 90\%$ under r=1 and it is robust to IA timing.









State of play

- **Is my study adequately powered?** The clinical equipoise prior provides *a non-normative* statistical rationale for calibration of sample size (i.e. power) "at π CE imbalance level"
  - *RCT outcomes at 5% FWER and 90% power provide >90% CE imbalance, under joint equipoise*
- **Sample size optimisation:** posterior odds and minimum detectable effect are building blocks of design utility, which is a formal basis for *efficient* design (Don Berry's[5] challenge).
- **Open questions**:
  - What data support $r \neq 1$ for Ph3 trial design?
  - What π should we target? Should π be the same level for positive/negative outcomes?
    - $\pi \neq 1 - \alpha$ may incentivise $r \neq 1$
  - Is the world ready for design utility as a basis for sample size optimisation?

Hartelijk dank voor jullie aandacht

References

1. https://linkinghub.elsevier.com/retrieve/pii/S2274580725003346
2. https://pubmed.ncbi.nlm.nih.gov/3600702/
3. https://pubmed.ncbi.nlm.nih.gov/3600702/
4. https://en.wikipedia.org/wiki/Beta_prime_distribution
5. https://www.sciencedirect.com/science/article/pii/S2274580725003346

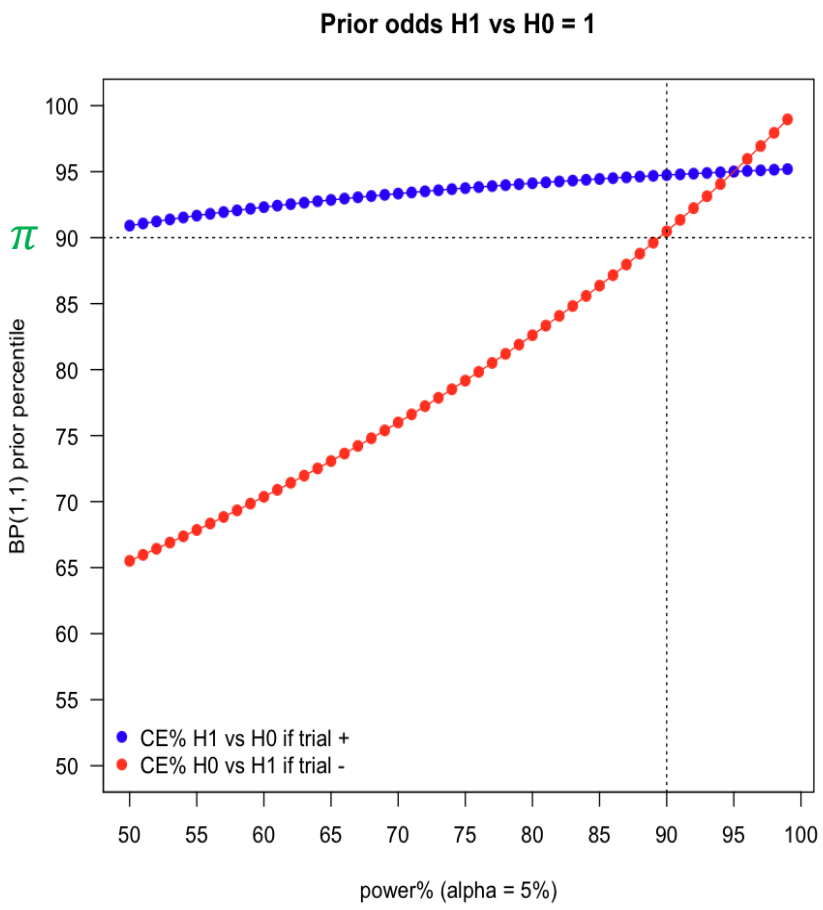
"Statistical methods in clinical research tend to become entrenched. Innovations threaten the status quo. ***The "right way" becomes frozen in lore.*** *This is so even when the "right way" is not best. "Statistical significance" and the* ***associated requirement of "high power" is an example. This attitude is an impediment to efficient design.***"

Back-ups



Sample size follows from pre-specified π, r

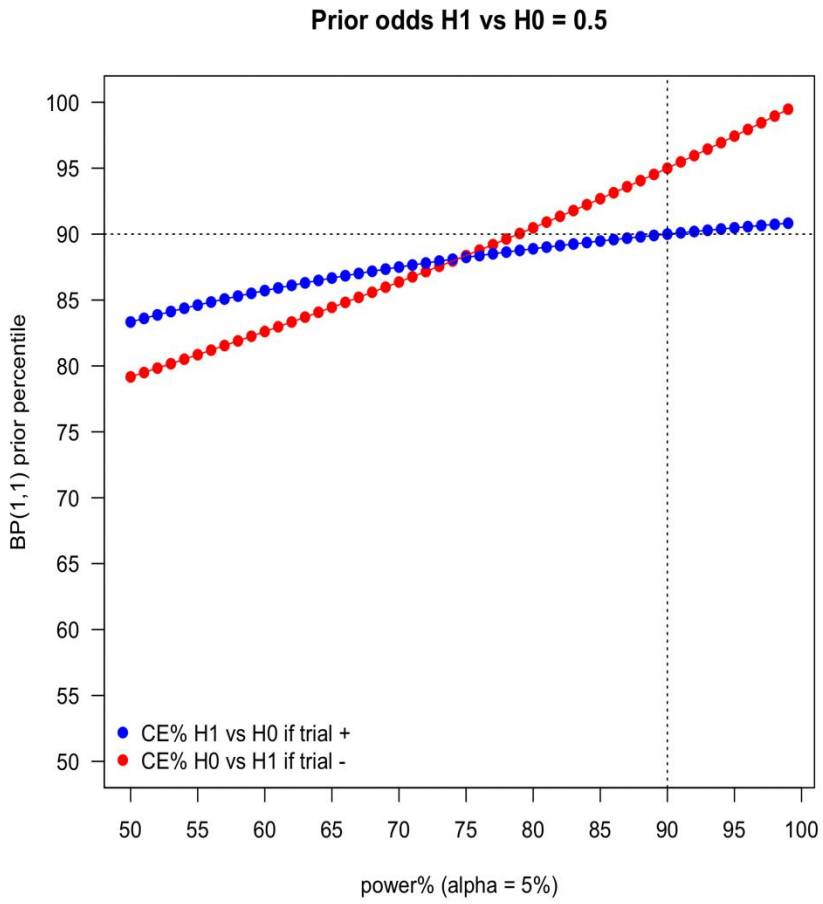
$\alpha = 0.05$



π

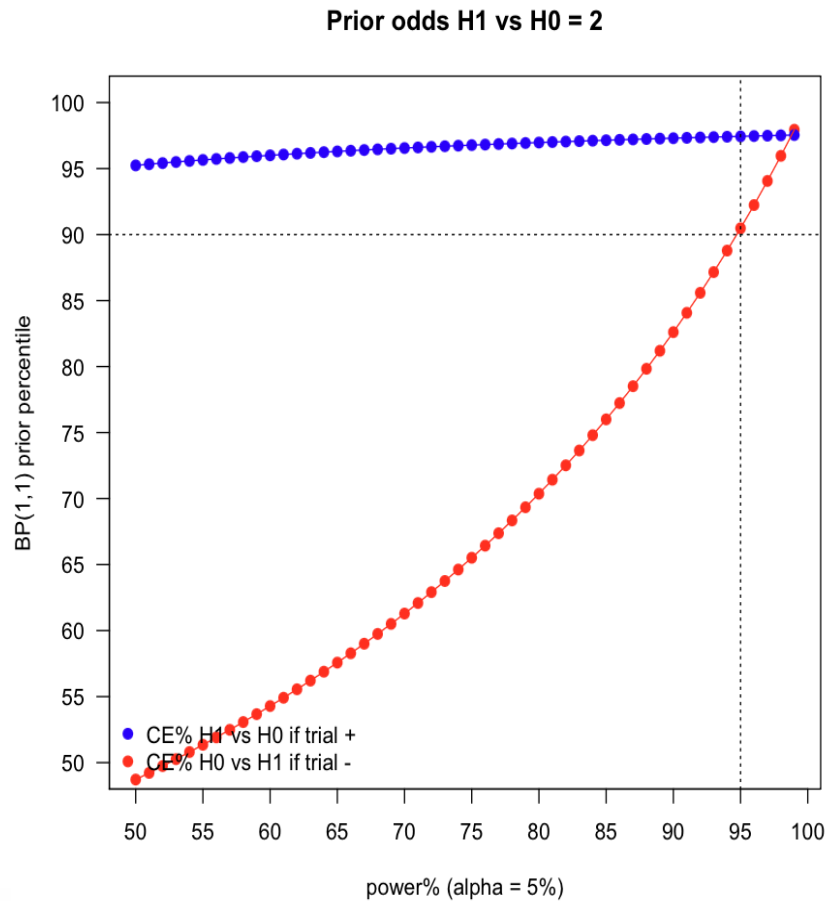
Equipoise prior

r = 1



Pessimistic prior

r = 2



Optimistic prior

r = 1/2

Under the CE prior, (population) surprise = (sponsor's) probability

- Under sponsor's equipoise $p(H_1) = p(H_0)$, so that (e.g.) when $\mathrm{p}(+|\mathrm{H}_0) \approx 5\%$ and $\mathrm{p}(+|\mathrm{H}_1) \approx 90\%$

$$p(H_1|+) = \frac{p(H_1)\times 0.9}{p(H_1)\times 0.9 + p(H_0)\times 0.05} = \frac{0.9}{0.9+0.05} \approx 94.7\%$$

$$p(H_0|-) = \frac{p(H_0)\times 0.95}{p(H_0)\times 0.95 + p(H_1)\times 0.1} = \frac{0.95}{0.95+0.1} \approx 90.4\%$$

$\Rightarrow$ when $\mathrm{p}(+|\mathrm{H}_0) \approx 5\%$ and $\mathrm{p}(+|\mathrm{H}_1) \approx 90\%$ and H_1 are H_0 are equally likely for the sponsor,

  - a positive readout shows the sponsor that H_1 was *correct* with 94.7% posterior probability
  - a negative readout shows the sponsor that H_0 was *correct* with 90.4% posterior probability

- Under joint equipoise, $p(H_1|+) = F_{BP(1,1)}\left(\frac{P(H_1|\mathrm{r}=\mathbf{1},+_N)}{P(H_0|\mathrm{r}=\mathbf{1},+_N)}\right), p(H_0|-) = F_{BP(1,1)}\left(\frac{P(H_0|\mathrm{r}=\mathbf{1},-_N)}{P(H_1|\mathrm{r}=\mathbf{1},-_N)}\right)$
  - e.g. when $\mathrm{p}(+|\mathrm{H}_0) \approx 5\%$ and $\mathrm{p}(+|\mathrm{H}_1) \approx 90\%$

$$p(H_1|+) \approx 94.7\% = \frac{18}{19} = F_{BP(1,1)}\left(\frac{P(H_1|\mathrm{r}=\mathbf{1},+_N)}{P(H_0|\mathrm{r}=\mathbf{1},+_N)}\right)$$

$$p(H_0|-) \approx 90.4\% = \frac{9.5}{10.5} = F_{BP(1,1)}\left(\frac{P(H_0|\mathrm{r}=\mathbf{1},-_N)}{P(H_1|\mathrm{r}=\mathbf{1},-_N)}\right)$$

