Concerns with using assurance to optimize trial design

Counter intuitive that a prior on σ does not affect our PoS:

- An under-powered study should result in a greater loss than an over-powered study
- Intuitively, uncertainty in σ should lead us to design a larger trial to have the same 'confidence' in the study design's ability to address the study objectives

Even if we don't require a prior on σ, using assurance to optimize trial design can be problematic

- Especially early in drug development assurance tends to be small
- As the trial design improves, assurance will not continue to increase but simply tend to a relatively low value – difficult to distinguish between design options (normalizing assurance might help, i.e. dividing assurance by its upper bound)

Walley, R. J., Smith, C. L., Gale, J. D., & Woodward, P. (2015). Advantages of a wholly Bayesian approach to assessing efficacy in early drug development: A case study. *Pharmaceutical Statistics*, *14*(3), 205–215. https://doi.org/10.1002/pst.1675

FERRING PHARMACEUTICALS

Posterior conditional success and failure distributions

A better tool to select between study designs?

The posterior conditional success distribution is the distribution for the effect δ assuming that the study will be a success, but without yet having observed any data:

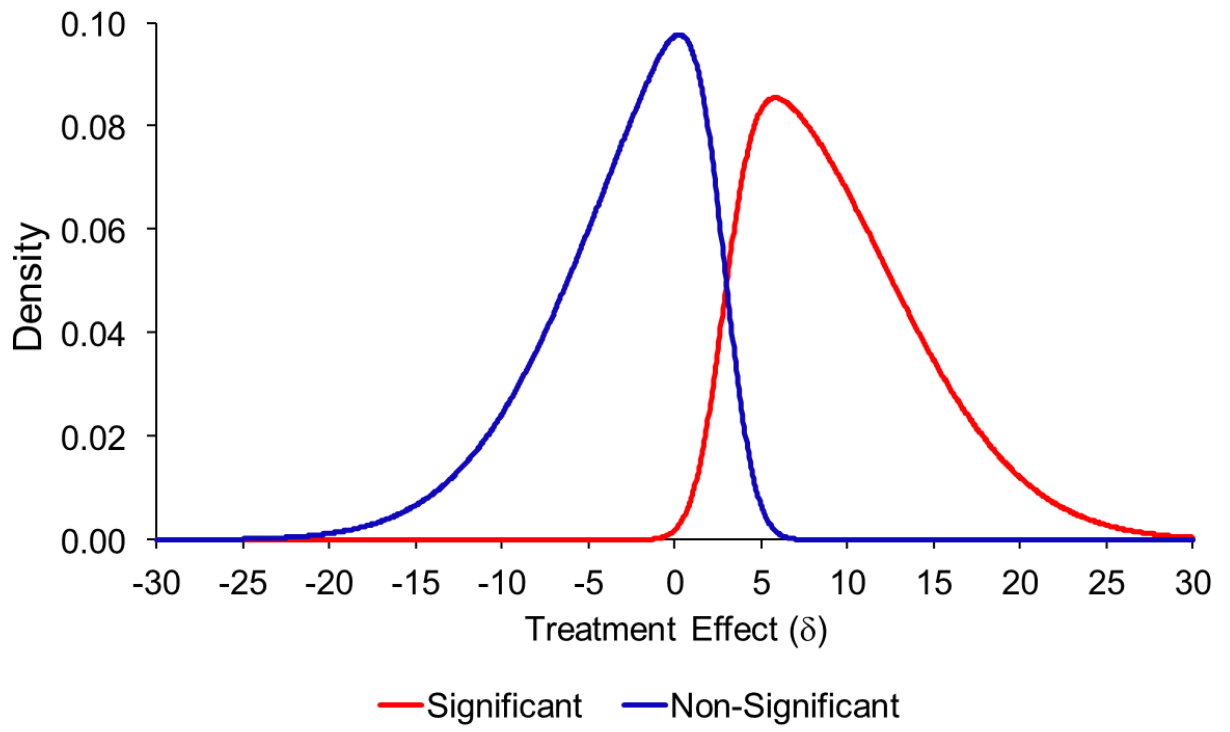
$$P[\delta|success] = P[\delta|\ \bar{y}_1 - \bar{y}_0 > \mathrm{Z}_{1-\alpha}\tau\ ]$$

Following Bayes' theorem, it can be obtained by

$$\frac{p(\delta)P[\bar{y}_1 - \bar{y}_0 > \mathrm{Z}_{1-\alpha}\tau|\delta]}{\int_\delta\ P[\bar{y}_1 - \bar{y}_0 > \mathrm{Z}_{1-\alpha}\tau|\delta]p(\delta)d\delta}$$

Which is simply the prior multiplied by the power function divided by the PoS (assurance)

The posterior conditional failure distribution can be similarly defined



These distributions can be used to assess the ability of the design to separate 'active' and 'inactive' compounds.

Exercise 4

The posterior conditional failure and success distributions for the case study

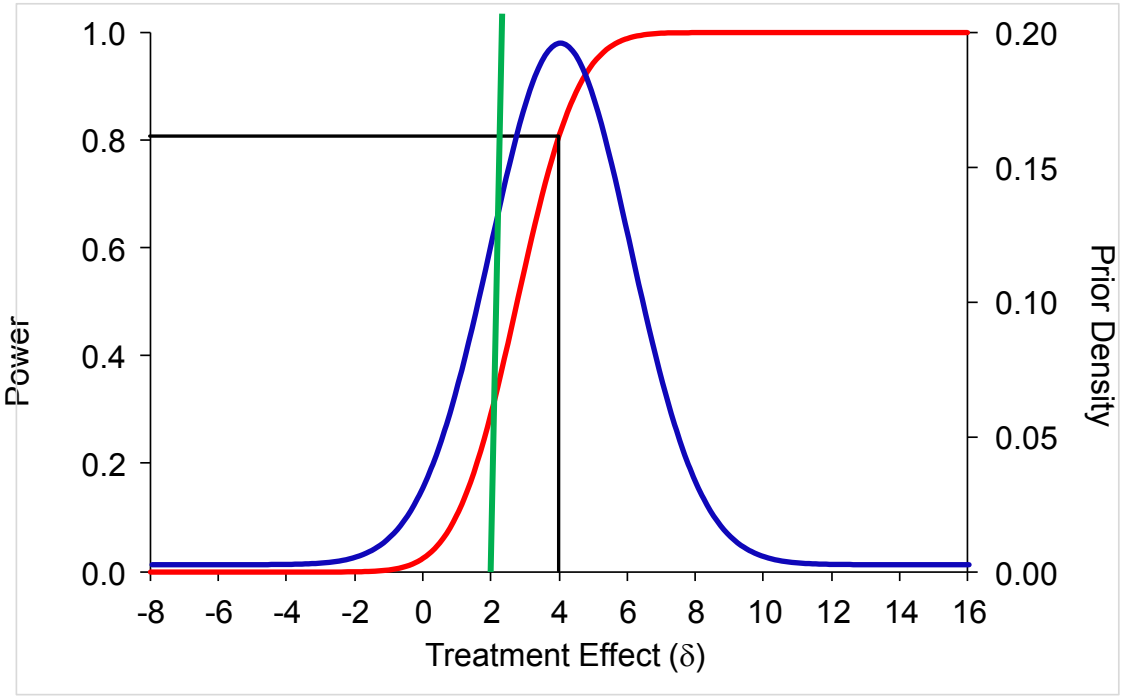
1. Write an R function that returns $p(\delta)$ for a given δ for the prior from Exercise 2

   Hint: use the R function `dnorm`

2. Write an R function that returns $P[\bar{y}_1 - \bar{y}_0 > Z_{1-\alpha}\tau|\delta]$ (i.e. the power) as a function of δ for the confirmatory trial with a sample size of 222 patients per arm and a standard deviation of 6.5.
3. Calculate the assurance for the confirmatory trial in case a sample size of 222 per arm is used and the success criterion is a significant p-value (you can re-use the result from Exercise 2 if you did it with a sample size of 222 per arm)
4. Use the results from steps 1-3 to create a plot of the posterior conditional success distribution
5. Similarly, derive the posterior conditional failure distribution and add it to the plot
6. Are you satisfied with the proposed design in its ability to distinguish between a drug that works and a drug that doesn't work?

FERRING PHARMACEUTICALS

Decomposition of assurance



Suppose the minimum clinically relevant difference is 2 units in the example on the left.

In calculating PoS we are averaging over regions which are not of interest to us – are not a success.

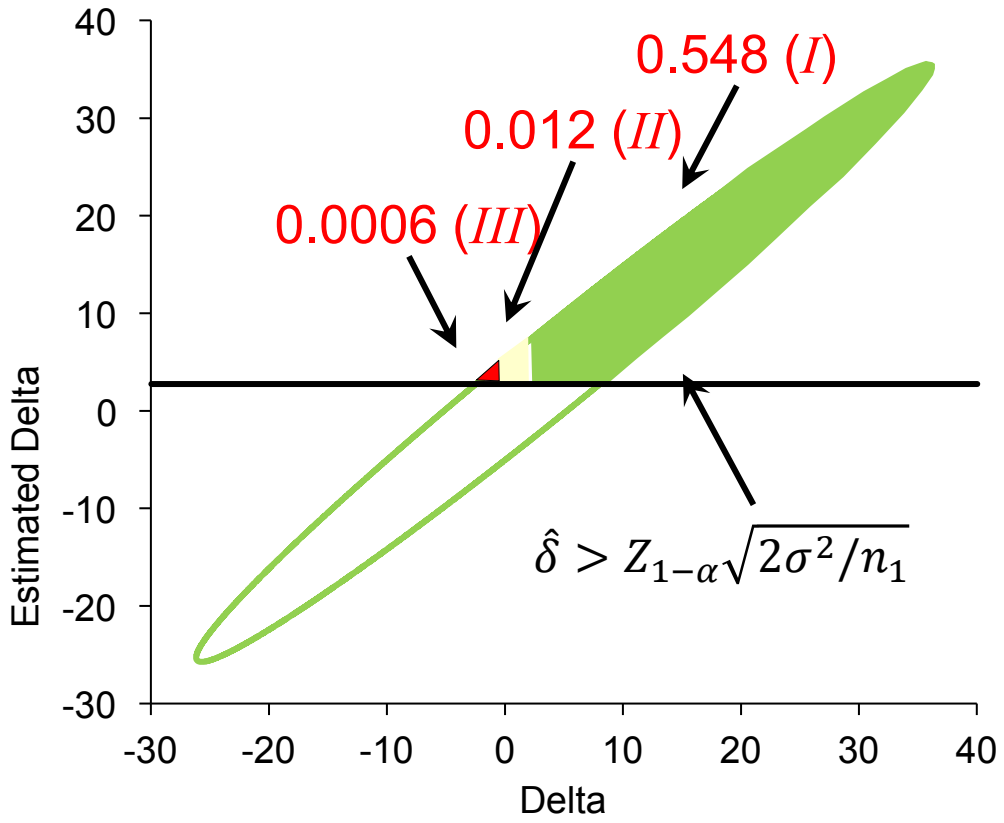
More extremely, values of $\delta < 0$ are contributing to the PoS in a region in which control is outperforming the test treatment.

Decomposition of assurance

Assurance is the probability of observing a success

This includes:

- False positive results in situations where the control treatment is better (III)
- False positive results in which the active treatment is better, but not by a relevant amount (II)
- True positive results (I)



The decomposition debate

The probability of success ought to be the **probability of a true success**

- We are interested in developing drugs that have clinical value and not in designing trials that clear a purely statistical hurdle
- Most appropriate for communicating the risk associated with the trial, e.g. for portfolio management
- Given high focus on type I error control, it seems strange to implicitly include type I errors as successes
- If success requires that the effect exceeds a certain threshold, the difference between the two approaches can be larger

The **probability of success may include false successes**

- If the success criterion is the p-value only, both versions of PoS are very similar as the PoS is inflated at most by the probability of a type I error (low impact as prior mass typically low for effects below 0)

My opinion: computing both can be very insightful, if we see a difference, it can be a sign that something is wrong in how we defined our success criterion

If we only compute the probability of a true success, we may not realize that our success criterion gives us too many observed successes

Conversely, we may end up with a too high PoS if including false positives

See also: A Review of Bayesian Perspectives on Sample Size Derivation for Confirmatory Trials - PMC (nih.gov)

Calculating the probability of a true success

We are interested in $P[\bar{y}_1 - \bar{y}_0 > \mathrm{Z}_{1-\alpha}\tau \; AND \; \delta > 0]$

- Can easily be done using simulations by only counting successful outcomes when the data generating effect size indeed constituted a success
- An analytical approach may also be possible. Here we illustrate a simple case:

The joint distribution of $\bar{y}_1 - \bar{y}_0$ and δ is a multivariate normal distribution with covariance

$$Cov(\bar{y}_1 - \bar{y}_0, \delta) = Cov\left(\delta + \sqrt{\sigma_\delta^2 + \tau^2}Z, \delta\right) = Cov(\delta, \delta) = Var(\delta) = \sigma_\delta^2,$$

with $Z \sim N(0,1)$. The marginal distributions remain as before.

To get the probability of a true success, we can use the bivariate normal distribution function to compute the probability that $P[\bar{y}_1 - \bar{y}_0 > \mathrm{Z}_{1-\alpha}\tau \; AND \; \delta > 0]$.

Exercise 5

1. Compute the probability of a true success for the confirmatory trial where a significant p-value is considered the criterion for success (for the null hypothesis of no effect)

   Hint: use the function `pmvnorm` from the package `mvtnorm`

2. Compare this probability to the assurance you calculated earlier.
3. Compute the probability of a true success for the confirmatory trial where a significant p-value as well as a point estimate above 1.5 is considered the criterion for success.
4. Compare your result from step 3 to the same version of assurance where we do not require a true success.

Benefits of using assurance

- Transparent evaluation of the risk of a program or study (considering both sampling variability and uncertainty about the drug effect)
- Foster and drive cross-functional exchanges/discussions (R&D and commercial functions)
- Triggers good discussions about expectations and facilitates alignment of expectations
- Enhance discussions through an analytical approach / data- or fact-based discussions

FERRING PHARMACEUTICALS