Probability of success in biomedical research

Instructor: Corine Baayen
Corine.Baayen@ferring.com

Helping people live better lives

Program for today/course objectives

Objectives:

- Understand how to obtain a probability of success estimate for a research project
- Ability to discuss various success criteria and their pros and cons
- Understand how to interpret and use a probability of success estimate

Program:

- Mix of lectures and exercises – discussion as we go along is very welcome!
- Two coffee breaks (morning and afternoon) and a lunch break in between

Consider how the material might apply to your own research and share if you like.

FERRING
PHARMACEUTICALS

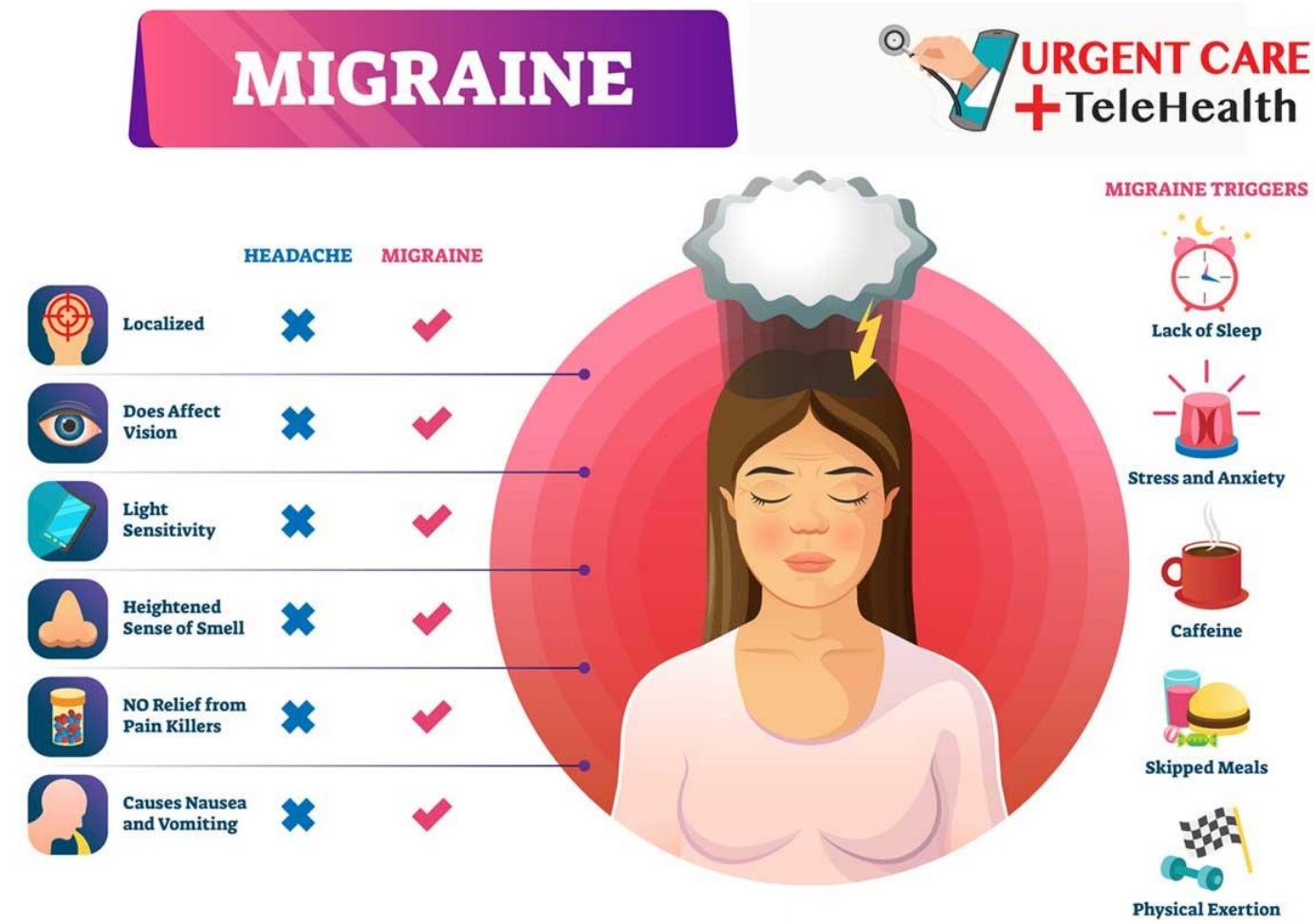
Case study

Loosely based on a real case

A drug is being developed to treat migraine

We have performed a small proof of concept study with positive results

Now we wish to design a large confirmatory trial to confirm that the drug works

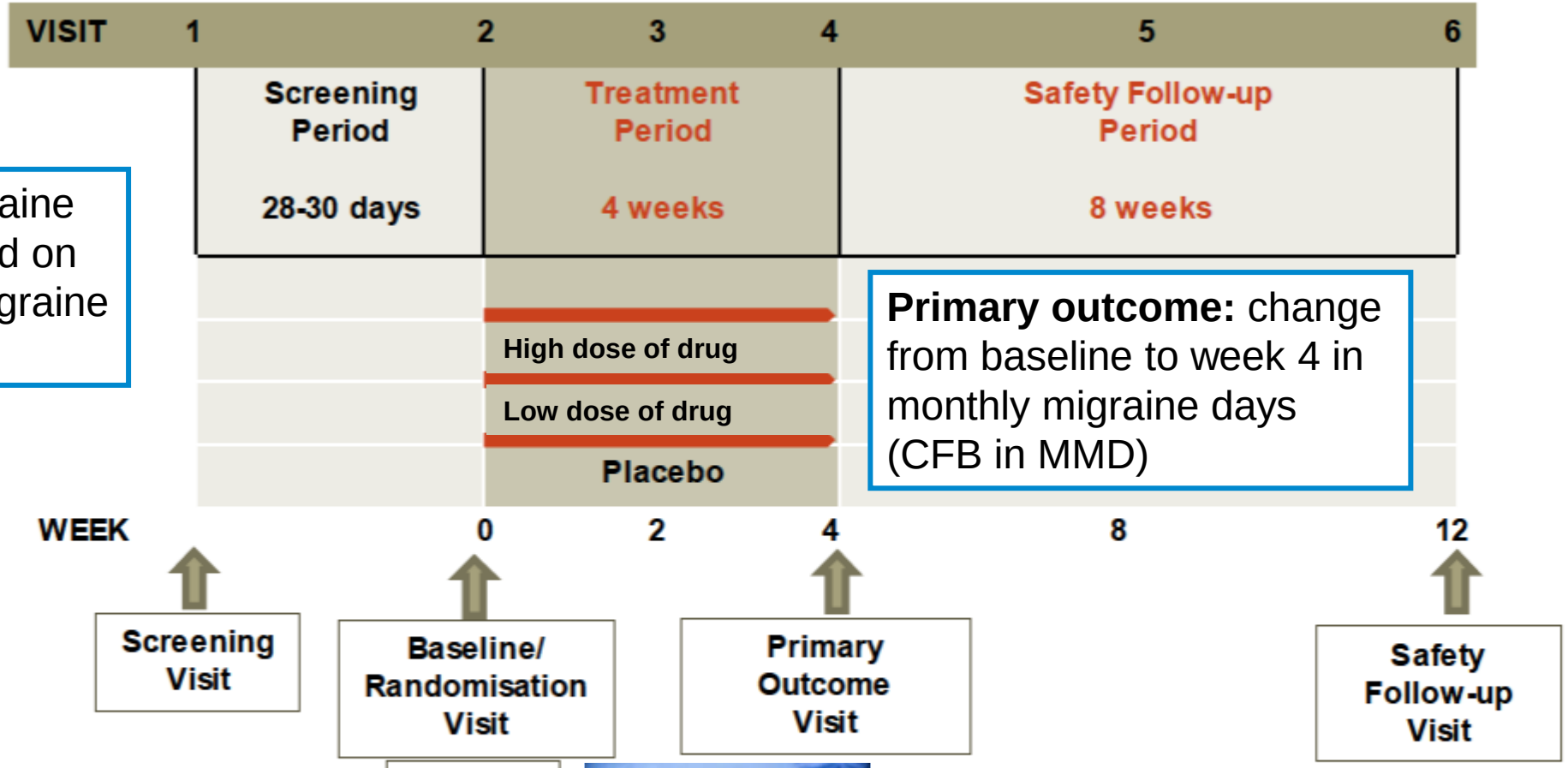


FERRING PHARMACEUTICALS

Design of the proof of concept study

A double-blind, parallel group, placebo controlled trial

Population: migraine patients who failed on 2-4 preventive migraine treatments



Primary outcome: change from baseline to week 4 in monthly migraine days (CFB in MMD)

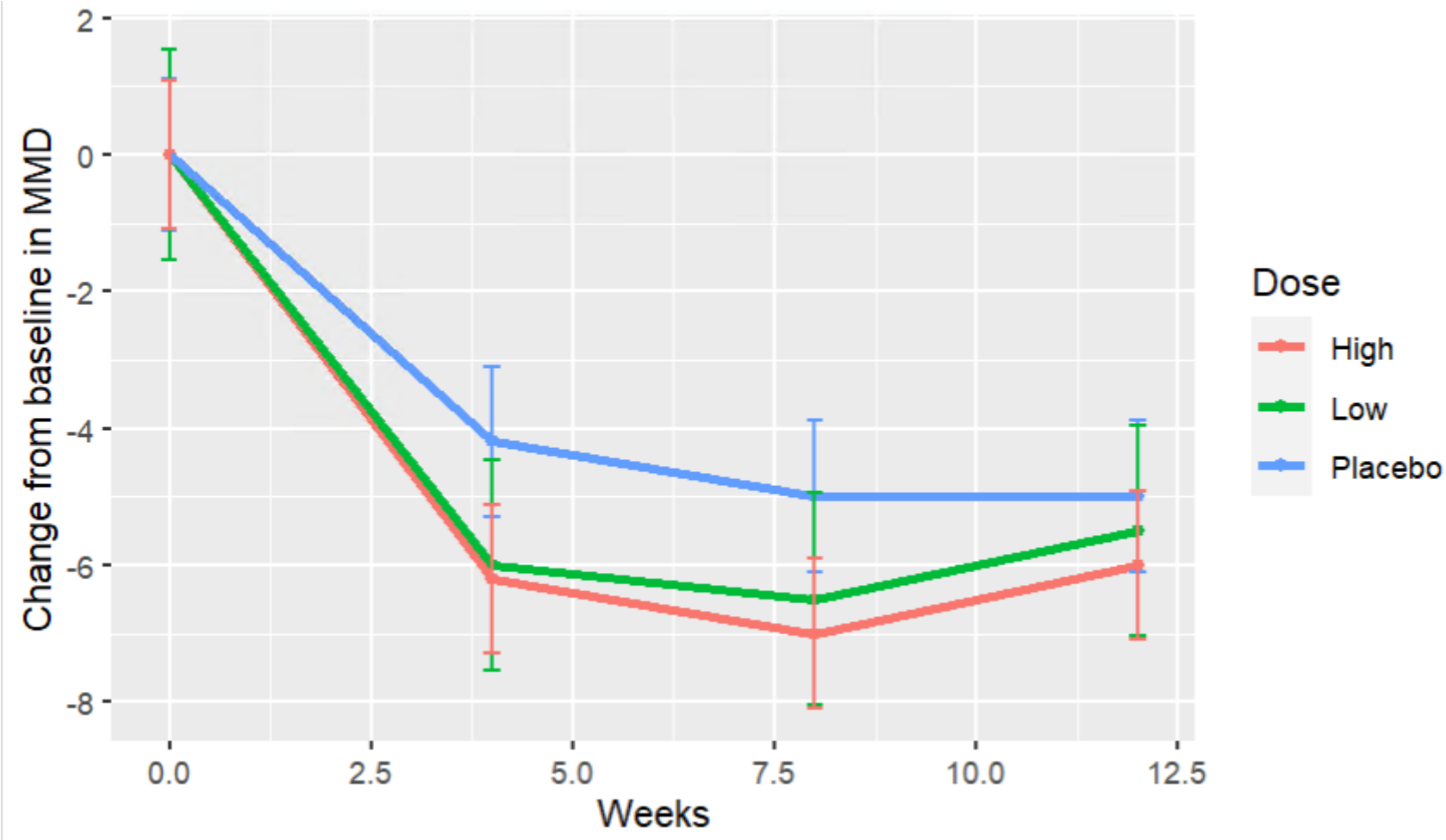
Mode of administration: Single intravenous infusion

IMP = investigational medicinal product.

FERRING PHARMACEUTICALS

Results from the proof of concept study

A positive study



Significant difference at week 4 of -2 (SE=0.89) MMD on high dose compared to pbo (p=0.01)

Standard deviation of CFB in MMD to week 4 of 6.5

Based on Results Posted | A Study With Lu AG09222 in Adults With Migraine Who Have Not Been Helped by Prior Preventive Treatments | ClinicalTrials.gov (data after 4 weeks is fictional)

SE = Standard Error

FERRING PHARMACEUTICALS

Study design for the confirmatory trial

A double-blind, parallel group, placebo controlled trial

VISIT	1	2	3	4	5	6	7
	Screening Period 28-30 days	Treatment Period 12 weeks				Safety Follow-up Period 8 weeks	
		High dose of drug					
		Placebo					
WEEK		0	4	8	12	16	20
	Screening Visit	Baseline/ Randomisation Visit			Primary Outcome Visit		Safety Follow-up Visit
		IMP	IMP	IMP			

Population: migraine patients who failed on 2-4 preventive migraine treatments

Primary outcome: change from baseline to week 12 in monthly migraine days (CFB in MMD)

Mode of administration: Monthly subcutaneous injections

IMP = Investigational Medicinal Product

FERRING PHARMACEUTICALS

Distribution of the difference in means

Suppose patients are randomised to one of two treatments, with n_i patients allocated to Treatment i

Suppose that the *j*th patient receiving Treatment i will yield a continuous response y_{ij} with

$$Y_{ij} \sim N(\mu_i, \sigma_i), \text{ independent.}$$

Then the distribution of the difference in means $\bar{y}_1 - \bar{y}_0 = \frac{1}{n_1}\sum_{i=1}^{n_1} y_{1i} - \frac{1}{n_0}\sum_{i=1}^{n_0} y_{0i}$ is

$$\bar{y}_1 - \bar{y}_0 \sim N(\mu_1 - \mu_0, \sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_0^2}{n_0}})$$

Where $\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_0^2}{n_0}}$ simplifies to $\sigma\sqrt{\frac{2}{n}}$ in case $\sigma_1 = \sigma_0$ and $n_1 = n_0$.

FERRING
PHARMACEUTICALS

Deciding on a sample size

Power calculation (continuous endpoint, assuming known variance)

H_0: null hypothesis - no effect

H_1: alternative hypothesis – the effect equals δ

μ_0: true mean CFB in MMD on placebo

μ_1: true mean CFB in MMD on drug

σ: standard deviation of CFB in MMD

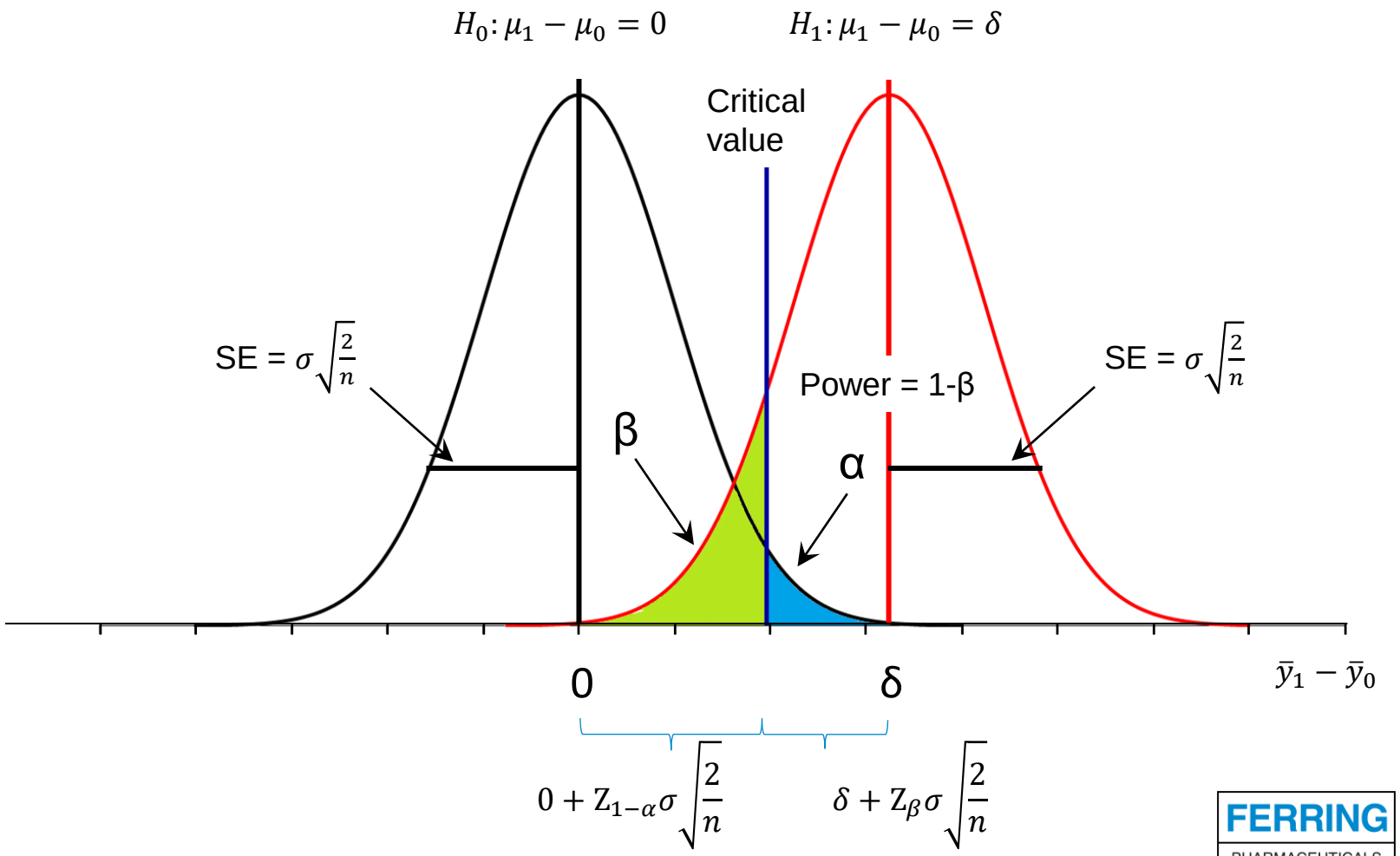
β: level at which to control the false negative rate

α: level at which to control the false positive rate

Z_q: q[th] quantile of the standard normal distribution

The sample size can be calculated by finding *n* such that $0 + Z_{1-\alpha}\sigma\sqrt{\frac{2}{n}} = \delta + Z_{\beta}\sigma\sqrt{\frac{2}{n}}$:

$$n = \frac{2\sigma^2\left(Z_{1-\alpha} + Z_{1-\beta}\right)^2}{\delta^2}$$



SE = Standard Error

FERRING PHARMACEUTICALS

Exercise 1 – propose a sample size for the confirmatory trial

You are very welcome to work together

Requirements/assumptions:

- We would like to have a power of 90%
- We would like to control the false positive rate at a one-sided 2.5% level
- Make appropriate assumptions about:
  - the expected difference δ in mean CFB in MMD at week 12
  - the standard deviation σ for the CFB in MMD at week 12

Hints:

- make use of the results from the proof of concept study
- In R Z_q can be obtained using `qnorm(q)` and x^2 by `x^2`

FERRING PHARMACEUTICALS

The Probability of Success (PoS) of a trial

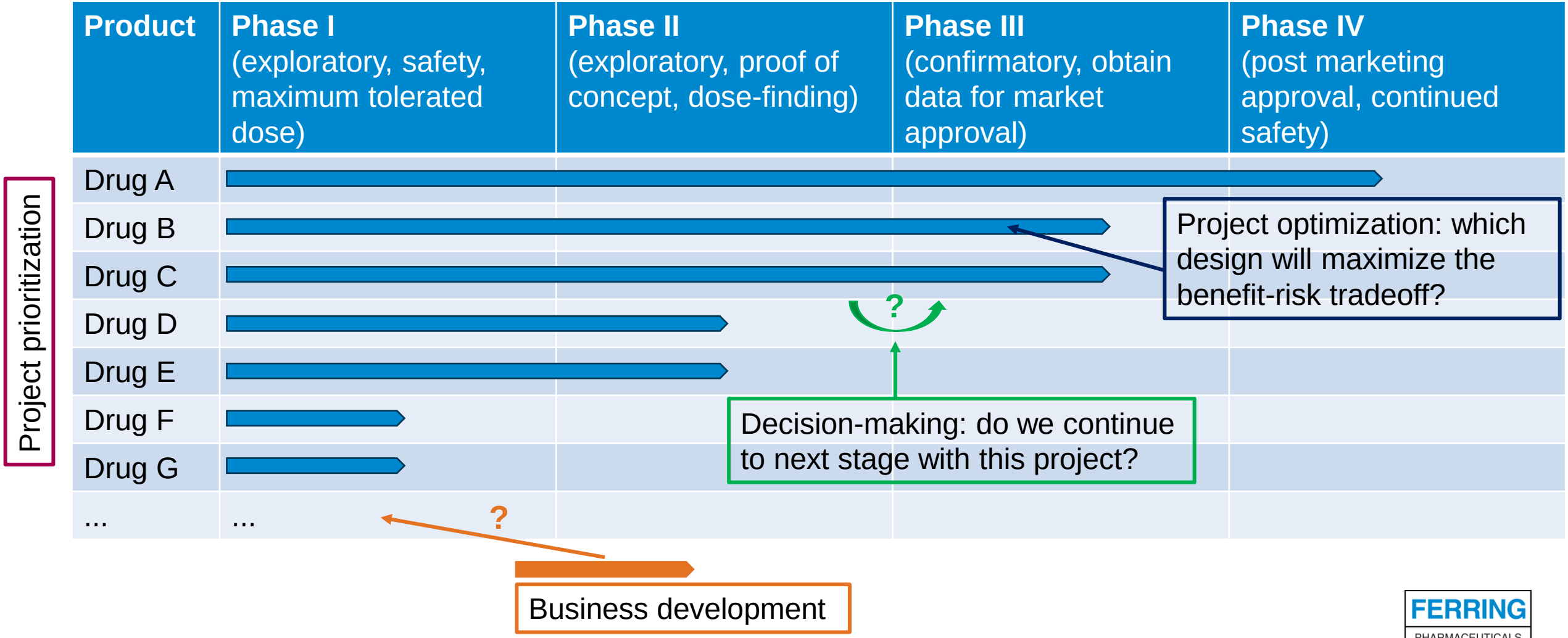
Evaluating the probability of achieving the primary trial objectives is useful to:

- Better understand the risk-benefit trade-off of the trial
- Finetune the study design (e.g. by maximizing the PoS)
- Secure funding (e.g. from sponsor governance boards)

FERRING
PHARMACEUTICALS

PoS in pharma

Project (or activity) prioritization, long term financial planning, decision-making



FERRING PHARMACEUTICALS

In general the PoS is a useful tool to help understand the benefit-risk tradeoff when doing biomedical research

- A low PoS may be acceptable for a project developing a treatment for a high risk disease for which no treatment is available
- A very high PoS may not be acceptable if it implies that the research question can be answered with currently available data – then putting patients at risk in a new trial may not be appropriate

There is no general guideline on an acceptable level of PoS, as it will always depend on project specific factors such as

- the expected benefit
- the costs and resources required
- the risk, e.g. side effects that might be experienced during the trial

The discussions required to obtain the PoS typically bring a lot of clarity to the research team in terms of definition of success, current understanding of the project and the best way forward.

FERRING
PHARMACEUTICALS

Interpreting the power of a trial

For a given

- sample size n
- level at which to control the false positive rate α
- effect size δ
- standard deviation σ

we can calculate the power of a trial with a continuous endpoint as:

$$\text{power} = \Phi\left(Z_\alpha + \frac{\sqrt{n}\delta}{\sqrt{2}\sigma}\right)$$

It gives us the probability of a significant p-value at the end of the trial, if the effect equals δ

Discussion: could power be a good measure for the probability of success of the trial?

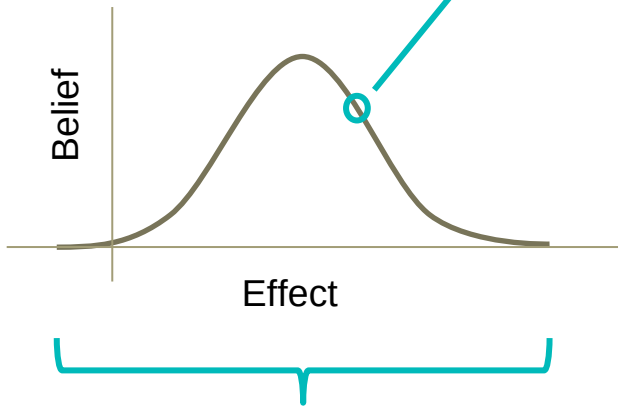
FERRING PHARMACEUTICALS

Assurance to estimate the Probability of clinical Success

Statistical power:

The probability of a significant p-value if the effect of the drug is one specific value

Criterion of success



Clinically relevant effect? Secondary endpoints?

Assurance:

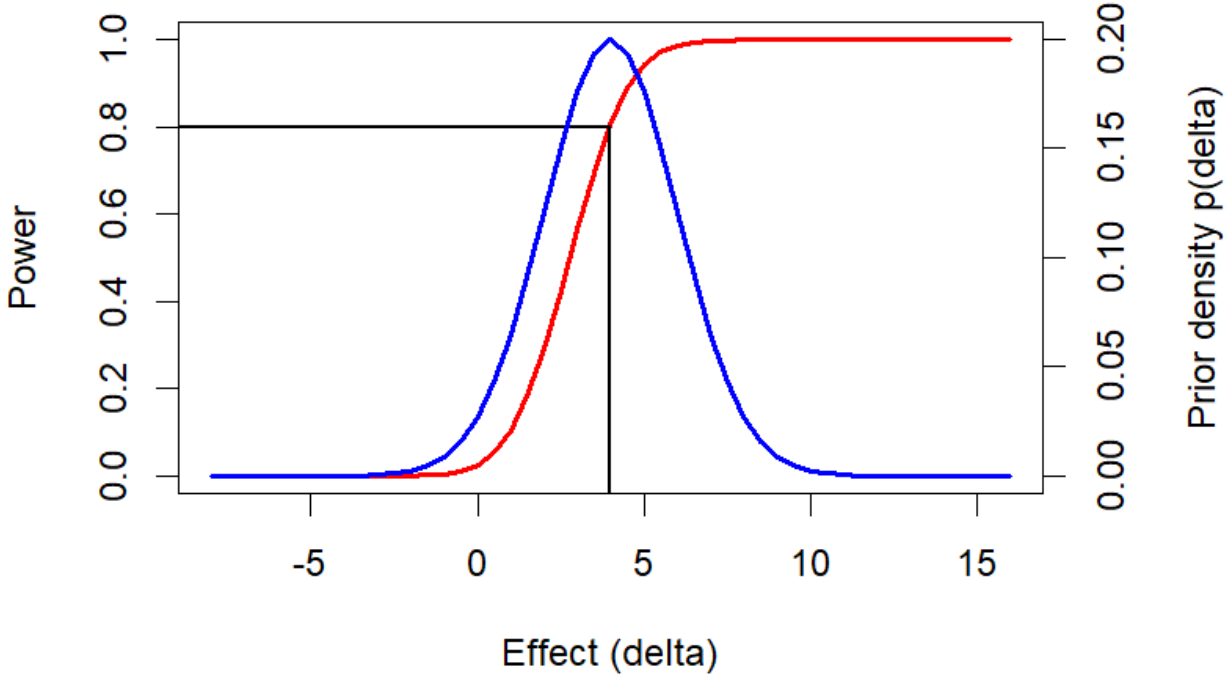
A weighted average power with more elaborate success criteria

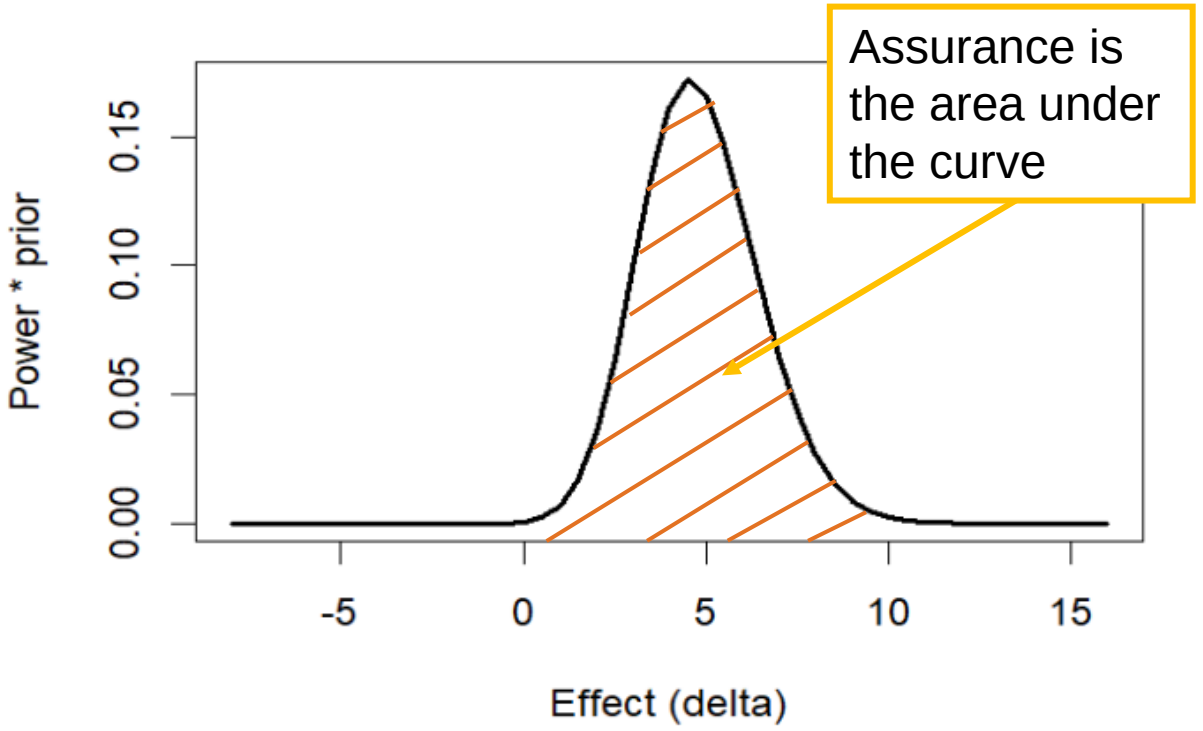
Evaluate "power" for a range of plausible effect sizes and calculate the average while weighing by how much you believe in each effect size

O'Hagan, A., Stevens, J. W., & Campbell, M. J. (2005). Assurance in Clinical Trial Design. *Pharmaceutical Statistics*, *4*(3), 187–201.

FERRING PHARMACEUTICALS

Assurance – simple case with success based on p-value



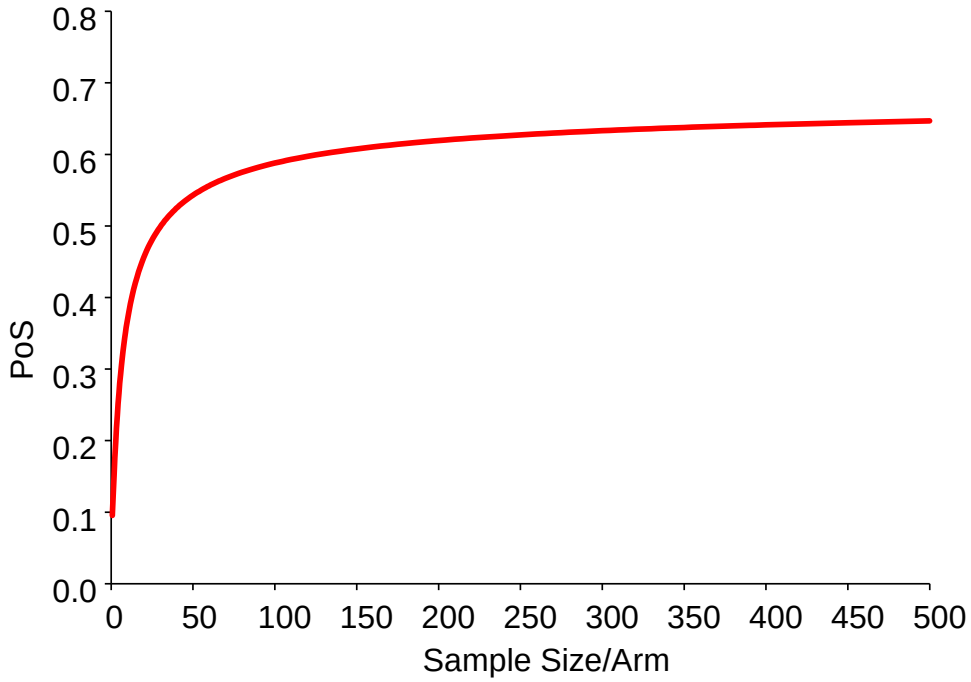


- The power $1 - \beta(\delta)$ is calculated conditional on δ being a specific value
- Assurance is the unconditional power: $\int_{\delta} \big(1 - \beta(\delta)\big) p(\delta) d\delta$ where $p(\delta)$ is the prior distribution

FERRING
PHARMACEUTICALS

Assurance has an upper bound that can be below 1

- Unlike power, assurance will typically reach an upper bound below 1 as sample size increases
- The upper bound is the prior probability of meeting the success criteria before data in the proposed study have been collected.
- This probability should not be “too high”, otherwise it is hard to argue that randomization is ethical



FERRING PHARMACEUTICALS

The prior is a key ingredient to the assurance calculation

Discuss with your neighbour(s) what the prior for the confirmatory study could look like.

How confident are you about your prior? Are there any aspects you would like to understand better before deciding on a final prior?

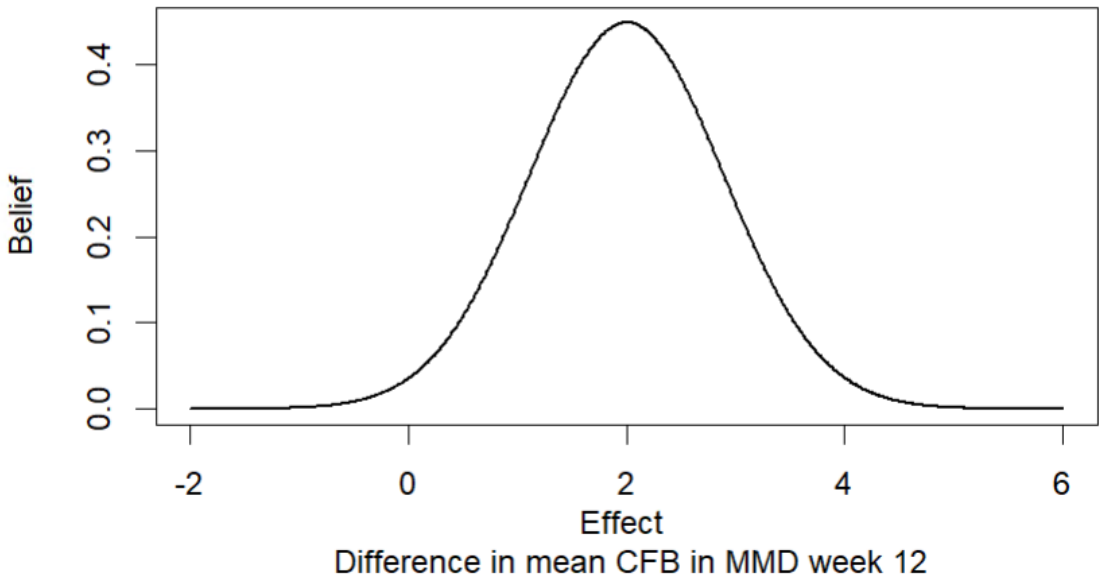
FERRING
PHARMACEUTICALS

A prior based on data alone

If data is available, this is very valuable when defining a prior

The effect estimate and its standard error could be used to define a prior using a normal distribution

For the proof of concept study the effect estimate was 2 MMD, with a SE of 0.89



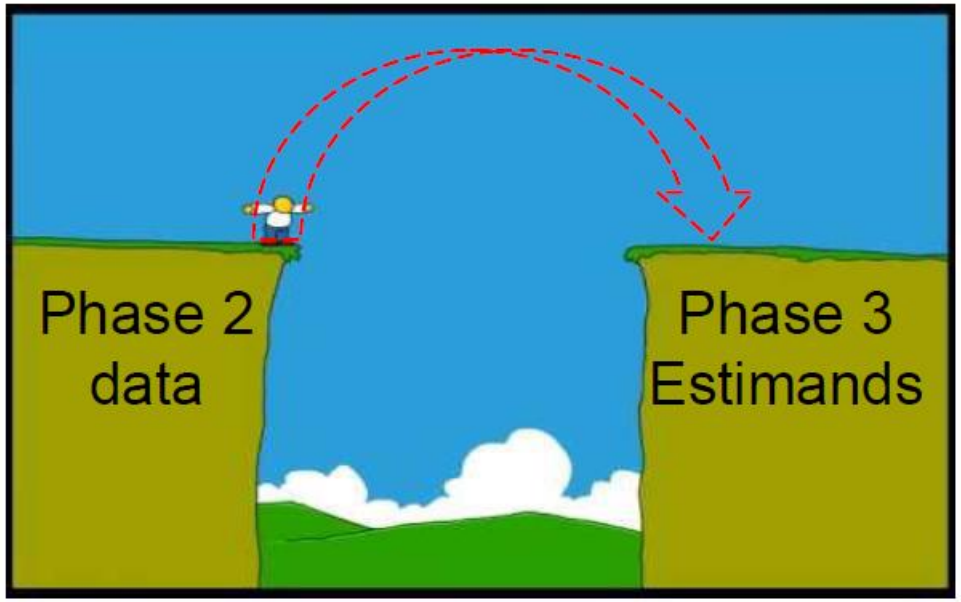
Note that a normal prior assumes that it is equally likely that the effect is smaller or larger than the observed effect from the proof of concept study (symmetric prior)

FERRING
PHARMACEUTICALS

Unfortunately, available data does not always reflect what we need

There are many differences between the proof of concept study and the confirmatory study that makes it difficult to translate the results directly to expectations for the confirmatory study

- Endpoint collected at different timepoints (Week 4 versus Week 12)
- Larger sample size (possibly including more countries/different sites)
- Different mode of administration
- More frequent dosing



Source: Joe Cartoon

Prior elicitation from clinical experts can be used to take into account uncertainties that are not captured by data

- Priors are based on a mixture of internal and external data which typically requires modelling to link the different sources.
  - E.g. internal data from proof of concept study, data from other (internal or external) studies on likely placebo effect, effect at 12 weeks etc.
- In some cases, little or no relevant data are available
- In both situations it can be helpful to draw on expert knowledge to translate available information into a prior distribution, either for the target parameter (effect), or for nonidentifiable parameters in a complex quantitative model
- Formal prior elicitation methods have been developed to derive priors based on input from clinical experts

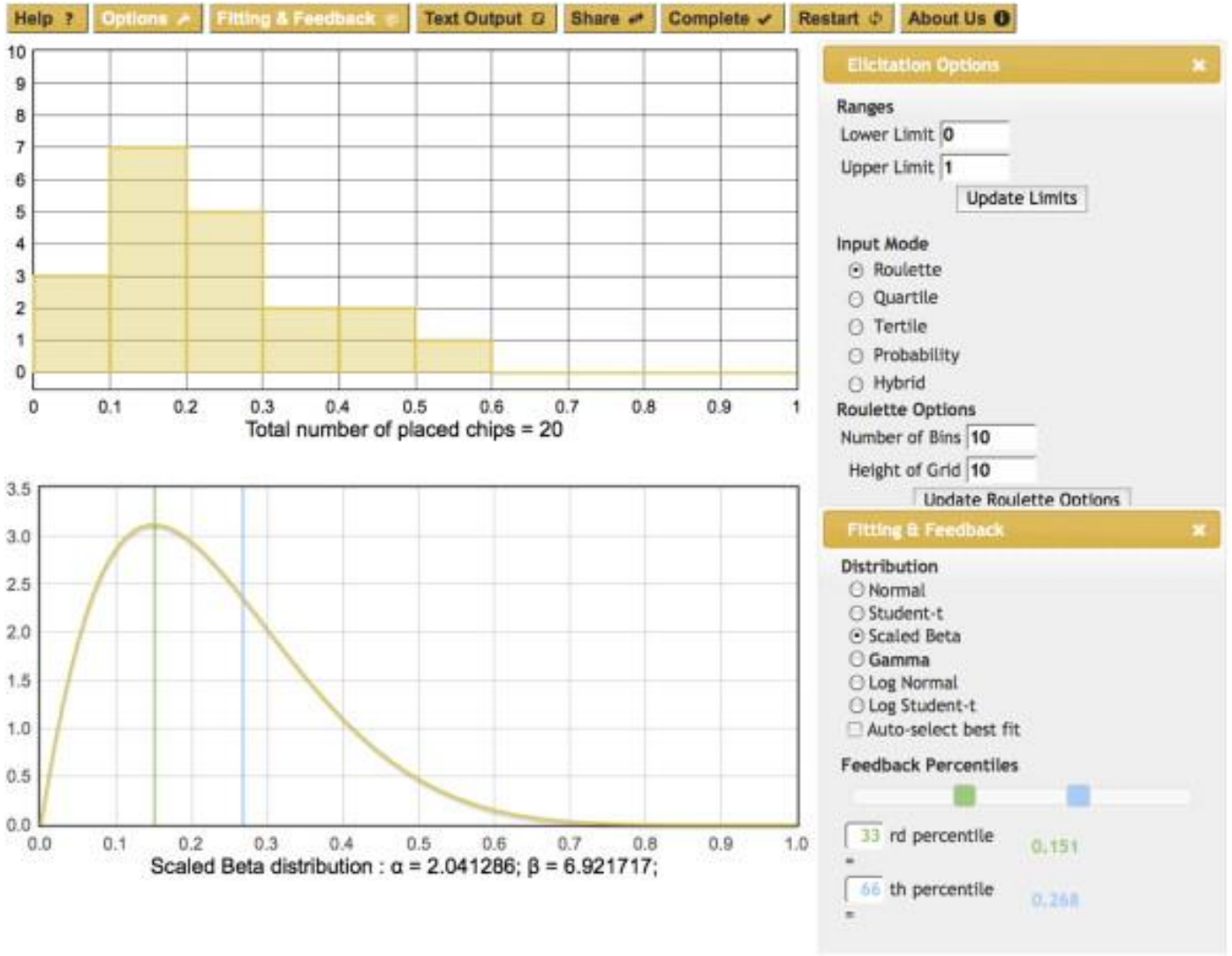
Dallow, N., Best, N., & Montague, T. H. (2018). Better decision making in drug development through adoption of formal prior elicitation, (July 2017), 301–316. https://doi.org/10.1002/pst.1854

FERRING PHARMACEUTICALS

Prior elicitation – Roulette Method

- The expert defines a range of plausible values
- This range is divided into a number of bins
- A fixed number of ‘chips’ is provided to the expert who is to distribute them over the bins (typically 20 works well)
- The proportion of chips in a bin gives the probability that the value of interest lies in this range

Method is often considered intuitive, but risk of too much focus on making distribution look ‘nice’ – training needed



FERRING PHARMACEUTICALS

Prior elicitation – modified PERT distribution

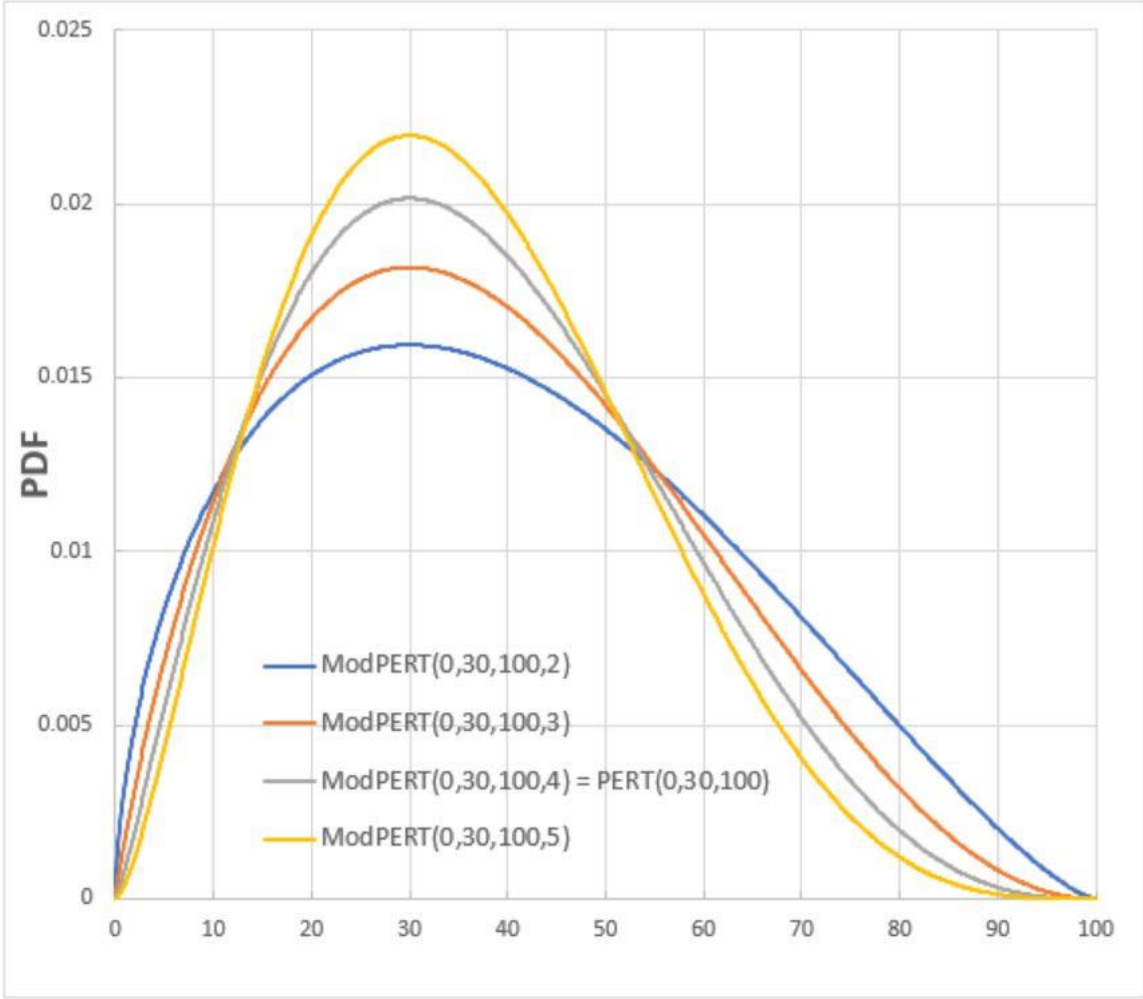
Defined by four parameters that are easy to interpret:

- Minimum possible value (a)
- Maximum possible value (c)
- Most likely value (b)
- Shape parameter (γ, make curve flatter or narrower)

Related to a beta distribution with additional assumption that the mean is:

$$\mu = \frac{a + \gamma b + c}{\gamma + 2}$$

Used for prior elicitation due its easy interpretation of the parameters



FERRING PHARMACEUTICALS

The SHELF elicitation framework discussed in the previous session can be used to obtain a prior from multiple experts

TABLE 1 Main steps in SHELF elicitation process

1. Select experts (and a good moderator)	These can be both internal and external to the company and should involve only those that have a good understanding of the details that need to be elicited.
2. Train experts	Provide experts with an overview of the elicitation process and the use of subjective probabilities and probability distributions
3. Evidence dossier	Prepare and review an evidence dossier that captures all pertinent information that the experts would rely upon to formulate their opinion.
4. Elicit individual priors	Elicit, in a masked fashion, individual priors from each expert (ie, experts are unaware of what other experts believe at this point)
5. Discuss individual priors	Share and review results from individual elicitations including each expert's rationale for their beliefs; discuss differences between experts.
6. Agree consensus prior	Where possible, elicit a "consensus" prior from the experts through discussion of what they collectively agree a "rational independent observer" would determine after having observed the previous conversations.
7. Documentation	Provide a written record of the elicitation session

See also: Dallow, N., Best, N., & Montague, T. H. (2018). Better decision making in drug development through adoption of formal prior elicitation. *Pharmaceutical Statistics*, *17*(4), 301–316.

FERRING PHARMACEUTICALS

Combining priors from different experts

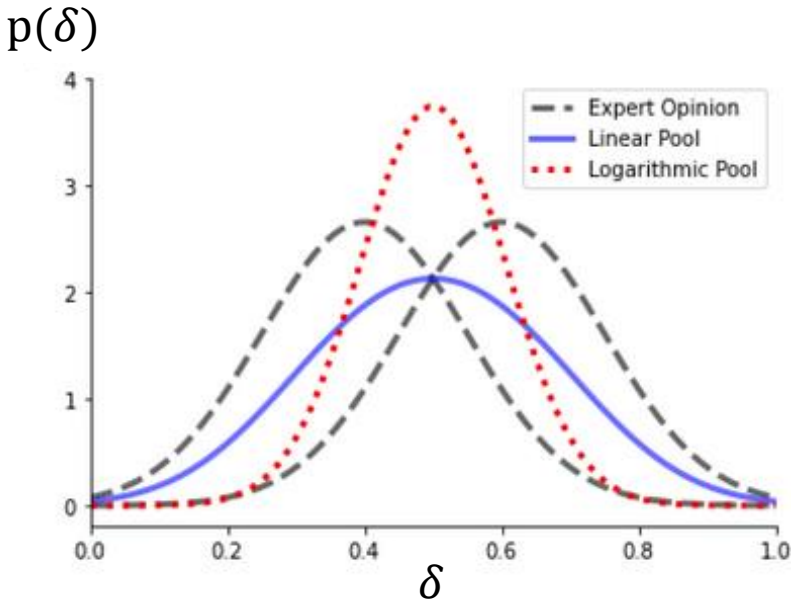
Once priors $p_1(\delta), \ldots, p_k(\delta)$ have been obtained from different experts, they need to be combined into a consensus prior $\mathrm{p}(\delta)$,.

1. Consensus through discussion – experts reveal their priors and agree on a common prior through discussion
2. Linear pooling

$$p(\delta) = \sum_{i=1}^{k} w_i p_i(\delta)$$

3. Logarithmic pooling

$$p(\delta) = c \prod_{i=1}^{k} p_i(\delta)^{w_i}$$



Where w_i $(i = 1, \ldots, k)$ are weights summing to one that can be used to give more weight to specific experts.

Figure from Dong, Jiayuan & Liao, Jiankan & Huan, Xun & Cooper, Daniel. (2023). Expert elicitation and data noise learning for material flow analysis using Bayesian inference. Journal of Industrial Ecology. 27. 10.1111/jiec.13399.

FERRING PHARMACEUTICALS

Be aware of cognitive bias when designing/eliciting a prior

How biased are you? ☺

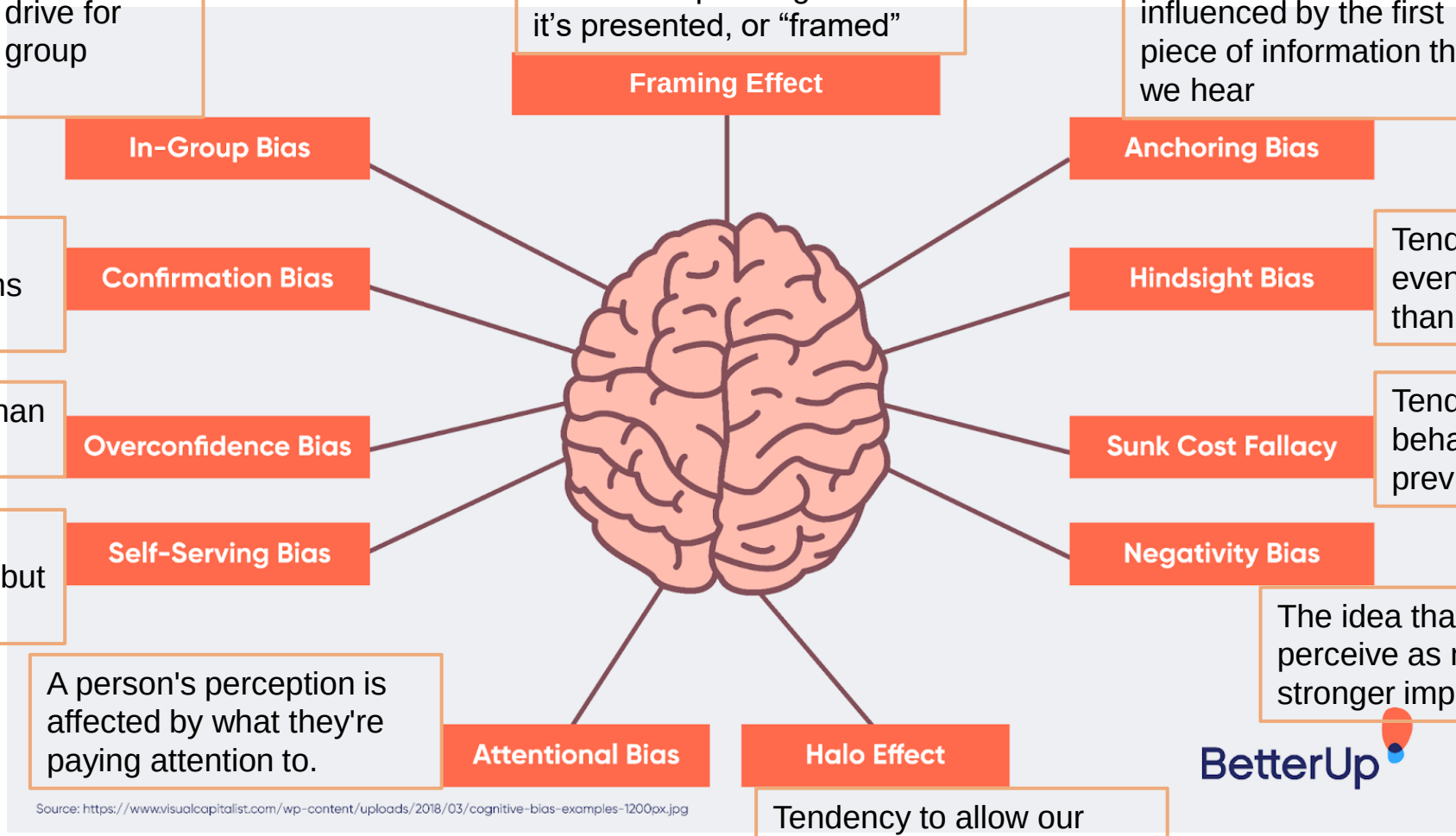
Anders is from Denmark and loves playing the trumpet. Which occupation is Anders more likely to have:

A. Anders is a musician in the national orchestra
B. Anders is a farmer

FERRING
PHARMACEUTICALS

Cognitive bias - to be aware of when designing a prior

And in general when doing research ;)



BetterUp

This is a non-exhaustive list.

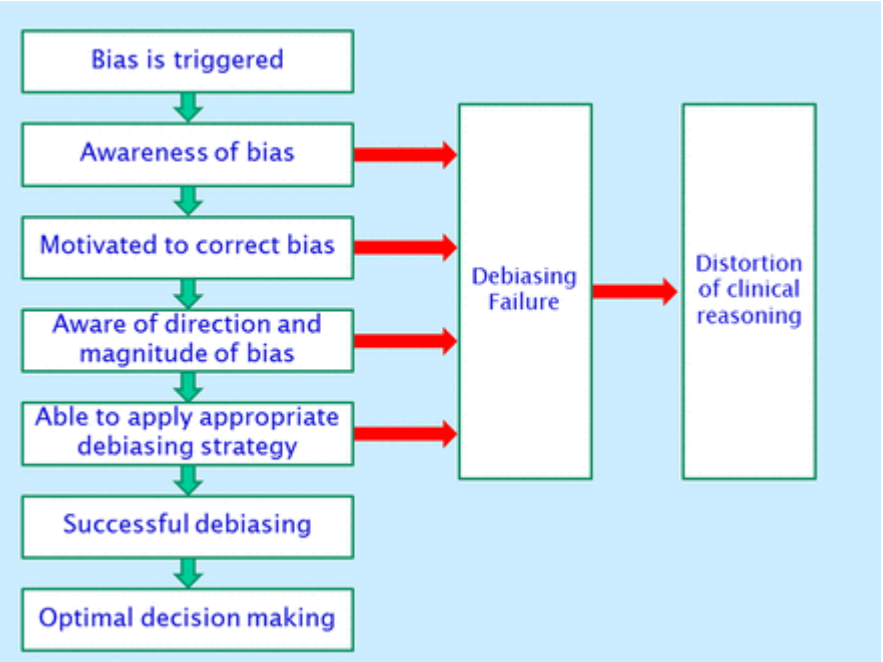
Image from: [Cognitive and Unconscious Bias: What It Is and How to Overcome It (betterup.com)]

FERRING PHARMACEUTICALS

Discuss with your neighbour(s)

Discuss which types of bias can be reduced using the SHELF framework.

Do you have any other suggestions to avoid cognitive bias?



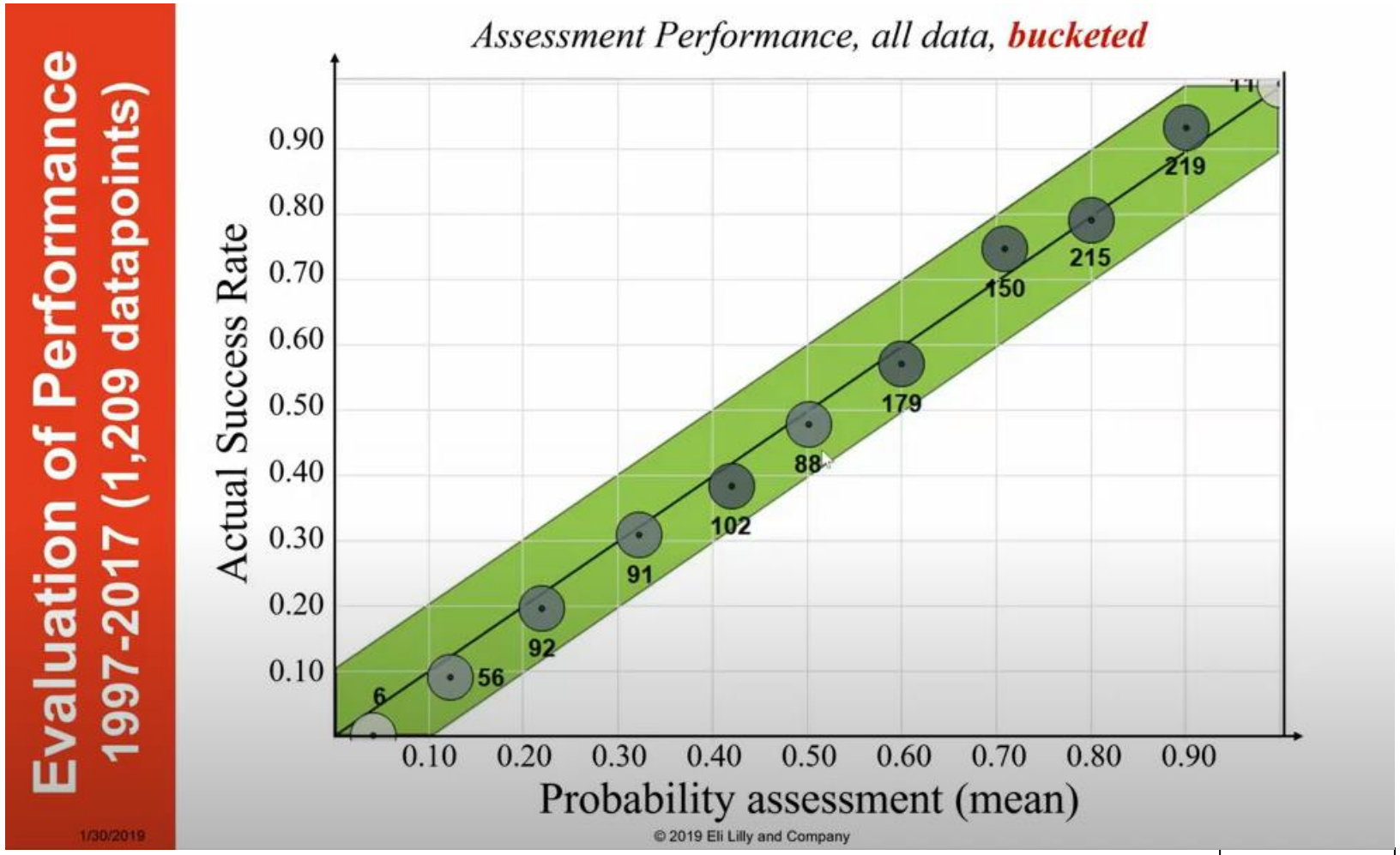
FERRING
PHARMACEUTICALS

Probability of success evaluation based on expert opinions

It can be done!

At Eli Lilly, probabilities of success (PoS) of drug development projects were elicited from a trained assessment group in collaboration with project teams

20 years of data shows that they were quite accurate



[Toward Superforecasting®: Lessons from the Eli Lilly Probability Assessment Panel | SDG – YouTube]

Further reading on avoiding bias: Superforecasting: The Art and Science of Prediction by P.E. Tetlock and Dan Gardner

FERRING PHARMACEUTICALS

Once a prior has been derived, assurance can be calculated in different ways

- Analytic calculation for simpler cases
- Simulation approach for more complex cases

FERRING
PHARMACEUTICALS

Assurance – calculation for simple case

The prior predictive distribution

Suppose that the prior distribution for the effect $\delta = \mu_1 - \mu_0$ has the conjugate normal form $\delta \sim N(\mu_\delta, \sigma_\delta)$

We previously discussed that $\bar{y}_1 - \bar{y}_0 \,|\delta, \tau \sim N(\delta, \tau)$, with $\tau = \sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_0^2}{n_0}}$

Equivalently $\bar{y}_1 - \bar{y}_0 = \delta + \tau Z$ with $Z \sim N(0,1)$

The unconditional (prior predictive) distribution for $\bar{y}_1 - \bar{y}_0$ has mean

$$E(\delta + \tau Z) = E(\delta) = \mu_\delta$$

and variance

$$Var(\delta + \tau Z) = Var(\delta) + Var(\tau Z) = \sigma_\delta^2 + \tau^2$$

In short, unconditionally

$$\bar{y}_1 - \bar{y}_0 \sim N(\mu_\delta, \sqrt{\sigma_\delta^2 + \tau^2})$$

See: O'Hagan, A., Stevens, J. W., & Campbell, M. J. (2005). Assurance in clinical trial design. *Pharmaceutical Statistics*, *4*(3), 187–201.

FERRING PHARMACEUTICALS

Assurance – calculation for simple case

One-sided superiority trial with significance as success criterion

Assurance gives the PoS before having collected any data

Success in this simplest case is defined as observing a one-sided p-value below the significance level α, or equivalently, observing $\bar{y}_1 - \bar{y}_0 > \mathrm{Z}_{1-\alpha}\tau$

Assurance can be calculated as follows:

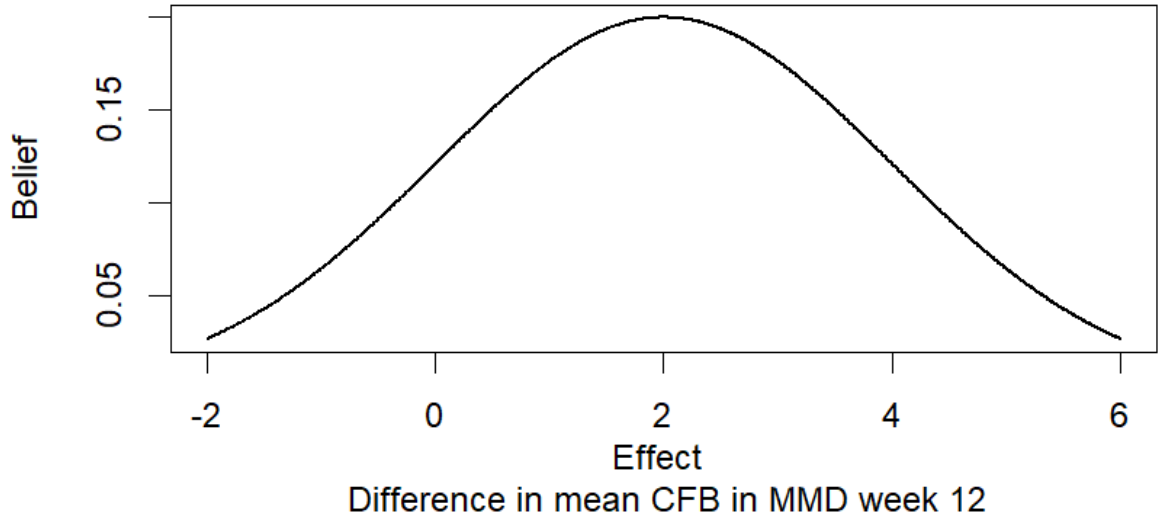
$$P[\bar{y}_1 - \bar{y}_0 > \mathrm{Z}_{1-\alpha}\tau] = \Phi\left(\frac{-\mathrm{Z}_{1-\alpha}\tau + \mu_\delta}{\sqrt{\sigma_\delta^2 + \tau^2}}\right)$$

See: O'Hagan, A., Stevens, J. W., & Campbell, M. J. (2005). Assurance in clinical trial design. *Pharmaceutical Statistics*, *4*(3), 187–201.

FERRING PHARMACEUTICALS

Exercise 2 – calculate assurance for the confirmatory trial

Use a normal distribution with mean 2 and standard deviation 2 as prior



Coding tip: use the apply function to quickly obtain the value of a function for a sequence of input values

```
n <- seq(0,1000,1)
dim(n) <- c(1,length(n))
results <- apply(n,2,yourfunction,...) where you can use ...
to give further arguments to yourfunction
```

1. Create a plot with sample size on the x-axis and PoS on the y-axis. Include one curve displaying power as a function of the sample size and one curve displaying assurance as a function of sample size. **Hint:** *first create R functions that return power and assurance as a function of sample size using formulas from Slides 31 and 13 (if you need help, see last slide in this presentation).*
2. What is the assurance for the sample size you calculated in Exercise 1? Would you choose a different sample size based on assurance?
3. What is the upper bound for assurance for this example? What is the interpretation of this upper bound?

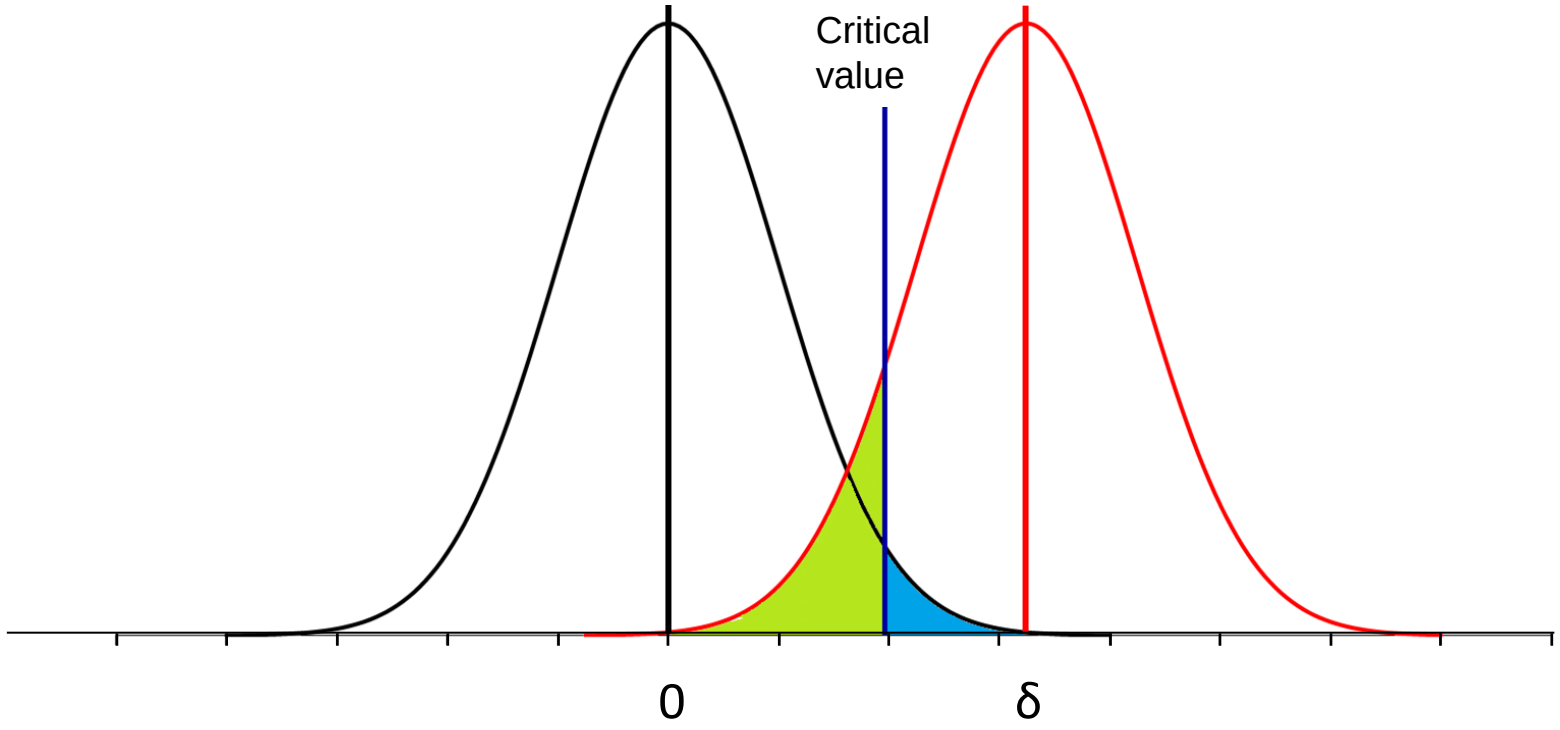
FERRING PHARMACEUTICALS

Success criteria

A p-value alone for a null hypothesis of no effect is rarely sufficient

The null hypothesis will be rejected if p<0.025 or equivalently if $\bar{y}_1 - \bar{y}_0 > Z_{1-\alpha}\tau$

For our confirmatory study, what is the smallest observed effect $\bar{y}_1 - \bar{y}_0$ that would lead us to reject the null hypothesis for a sample size of 222/arm and a common standard deviation of 6.5?



FERRING
PHARMACEUTICALS

More complex success criteria

More complex success criteria can be considered, for example:

- A significant p-value and an observed effect size above a minimum relevant limit
- A *q*% confidence interval of which the lower limit exceeds a minimum relevant limit
- A significant p-value for the null hypothesis that the effect is smaller or equal to the minimum relevant limit
- Criteria based on multiple endpoints
- Bayesian success criteria, e.g. $P[\delta > \Delta] > \pi$
- ...

FERRING
PHARMACEUTICALS

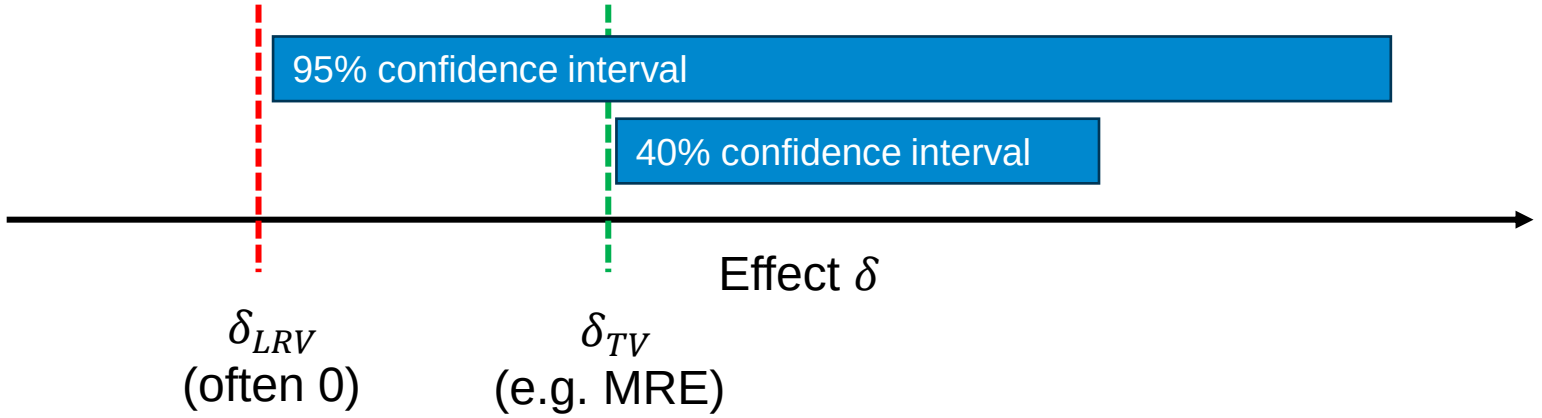
Exercise 3 – adding a requirement on the minimum relevant effect to assurance

1. To the plot from Exercise 2, add a line for assurance where success is declared if the p-value is significant AND the effect estimate is at least 1.5
   - Hint: modify the formula from Slide 31
2. What is the upper bound for this version of assurance?
3. Also add to the plot a line for assurance where success is declared in case of a significant p-value for evaluating the null hypothesis that the effect equals 1.5
   - Hint: modify the formula from Slide 31
4. What is the upper bound for this version of assurance?
5. Which version of assurance do you prefer for the case study?

Bonus: what sample size would we need for the confirmatory trial if we wish to reject a null hypothesis that the effect equals 1.5 and we assume the effect is 2 under the alternative and we require 90% power and wish to control the false positive rate at 2.5% one-sided?

FERRING
PHARMACEUTICALS

Dual success criteria



- **Criterion 1 (minimum requirement)**: at least 97.5% confidence that the effect exceeds δ_{LRV}
- **Criterion 2 (relevance requirement)**: at least 70% confidence that the effect exceeds δ_{TV}

- **Lower reference value (δ_{LRV}):** usually, but not always, the threshold for 'statistical significance', e.g. 0 difference in mean response, odds ratio of 1, hazard ratio of 1 etc.
- **Target value (δ_{TV}):** usually clinically relevant (or commercially viable) value.

Note that we can choose other values than 97.5% and 70% for the levels of confidence

FERRING PHARMACEUTICALS

Declaring success with dual success criteria

Success is declared if:

The minimum requirement is met: $\bar{y}_1 - \bar{y}_0 > \delta_{LRV} + Z_{1-\alpha_0}\tau$

AND

The relevance requirement is met: $\bar{y}_1 - \bar{y}_0 > \delta_{TV} + Z_{1-\alpha_1}\tau$, with for example $\alpha_0 = 0.025$ and $\alpha_1 = 0.3$

		Minimum requirement met?	
		Yes	No
Relevance requirement met?	Yes	Success	Consider
	No	Consider	No Success

FERRING PHARMACEUTICALS

Dual decision criteria – operating characteristics

Both criteria are met (*SUCCESS*) if $\bar{y}_1 - \bar{y}_0 > \max\left(\delta_{LRV} + Z_{1-\alpha_0}\tau,\ \delta_{TV} + Z_{1-\alpha_1}\tau\right) = MAX$

Neither are met (*NO SUCCESS*) if $\bar{y}_1 - \bar{y}_0 < min\left(\delta_{LRV} + Z_{1-\alpha_0}\tau,\ \delta_{TV} + Z_{1-\alpha_1}\tau\right) = MIN$

Decision	Probability of the Decision
SUCCESS	$P_{SUCCESS} = \Phi\left(\frac{-MAX + \mu_\delta}{\sqrt{\tau^2 + \sigma_\delta^2}}\right)$
NO SUCCESS	$P_{NO\ SUCCESS} = \Phi\left(\frac{MIN - \mu_\delta}{\sqrt{\tau^2 + \sigma_\delta^2}}\right)$
CONSIDER	$P_{CONSIDER} = 1 - P_{SUCCESS} - P_{NO\ SUCCESS}$

FERRING PHARMACEUTICALS

Dual decision criteria – upper bound

As $n \to \infty$, $MAX \to \delta_{TV}$ and $MIN \to \delta_{LRV}$ and therefore the upper bounds become

Decision	Probability of the Decision
SUCCESS	$P_{SUCCESS} = \Phi\left(\frac{-\delta_{TV} + \mu_\delta}{\sigma_\delta}\right)$
NO SUCCESS	$P_{NO\ SUCCESS} = \Phi\left(\frac{\delta_{LRV} - \mu_\delta}{\sigma_\delta}\right)$
CONSIDER	$P_{CONSIDER} = 1 - P_{SUCCESS} - P_{NO\ SUCCESS}$

Same upper bound as assurance with success criteria:

1. Significant p-value for $H_0: \mu_1 - \mu_0 = 0$ AND point estimate > δ_{TV}
2. Significant p-value for $H_0: \mu_1 - \mu_0 = \delta_{TV}$

But more rigorous than 1. as it incorporates uncertainty around point estimate and more flexible than 2. allowing for smaller sample sizes.

FERRING PHARMACEUTICALS

Assurance with unknown variance

Let's consider the same framework as discussed so far, but with unknown variance σ^2 that is equal for both groups ($\sigma = \sigma_0 = \sigma_1$)

In this case we typically base our test on the t-distribution rather than the normal distribution, i.e. we reject the null hypothesis if

$$\bar{y}_1 - \bar{y}_0 > t_{1-\alpha,2n-2}\hat{\sigma}\sqrt{\frac{2}{n}},$$

with $\hat{\sigma}^2 = \frac{\sum_{i=0}^{1}\sum_{j=1}^{n_i}(y_{ij}-\bar{y}_i)^2}{2n-2}$ the estimate of the variance and $t_{1-\alpha,2n-2}$ the *1-α* percentile of the *t*-distribution with *2n-2* degrees of freedom.

Since we do now not know σ, we may also wish to define a prior on σ.

O'Hagan, A., Stevens, J. W., & Campbell, M. J. (2005). Assurance in clinical trial design. *Pharmaceutical Statistics*, *4*(3), 187–201.

FERRING PHARMACEUTICALS

For the case of unknown sigma (or complex success criteria) a simulation approach can be used to calculate assurance

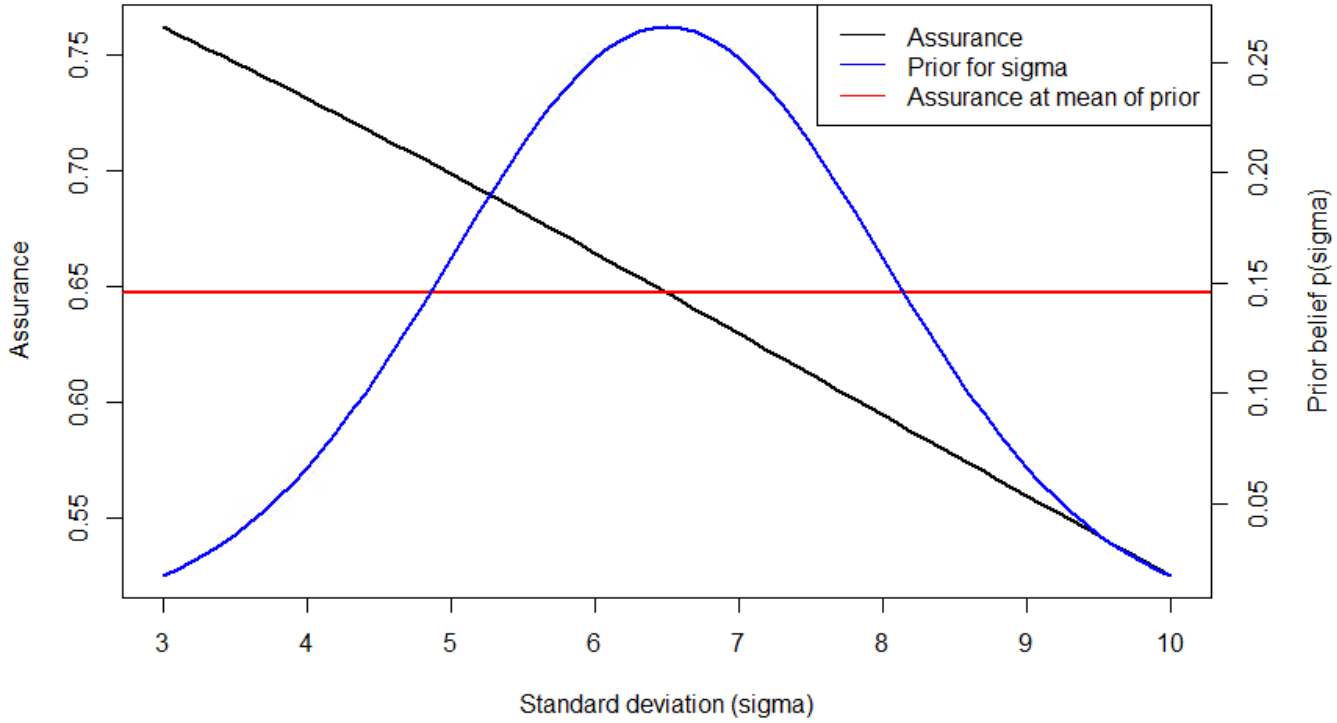
1. Set counters *I* = *P* = *0*. Set required number of simulations *N*.
2. Sample δ and σ from their joint prior distribution.
3. Sample an observed effect $\bar{y}_1 - \bar{y}_0 \sim N(\delta, \tau)$ and an estimated standard deviation using $(2n - 2)\hat{\sigma}^2/\sigma^2 \sim \chi^2_{2n-2}$.
4. Increment *P* if $\bar{y}_1 - \bar{y}_0 > t_{1-\alpha,2n-2}\hat{\sigma}\sqrt{\frac{2}{n}}$.
5. Increment *I*. If *I* < *N*, go to step 2.
6. Estimate assurance by *P/N*

FERRING
PHARMACEUTICALS

A comment on priors on the standard deviation

Both in my own experience and that of Walley et al (2015) a prior on σ does not affect assurance in a relevant manner compared to calculating assurance based on the mean of the prior for σ

This finding is due to the assurance value changing almost linearly over the credible range for σ, which when combined with an approximately symmetrical prior distribution results in the marginal assurance value almost being equal to the assurance at the prior mean for σ.



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

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:

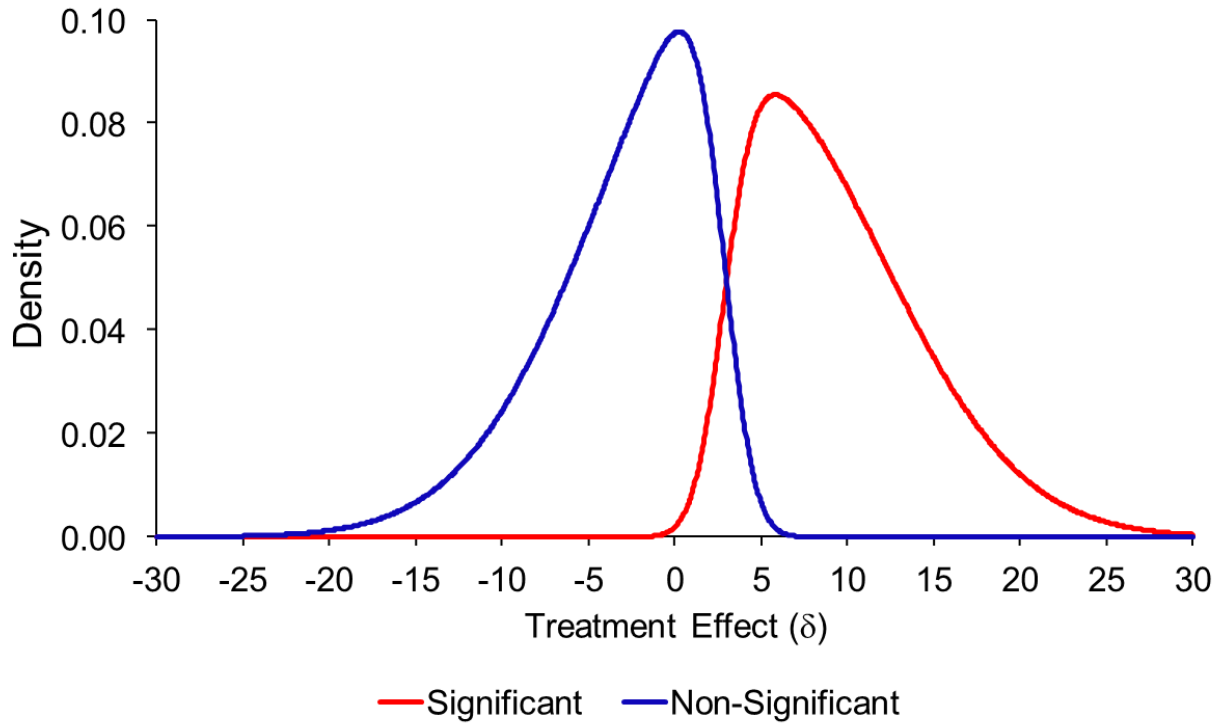
$$P[\delta|success] = P[\delta|\ \bar{y}_1 - \bar{y}_0 > \mathrm{Z}_{1-\alpha}\tau\ ]$$

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.

FERRING PHARMACEUTICALS

Exercise 4

The posterior conditional failure and success distributions for the case study

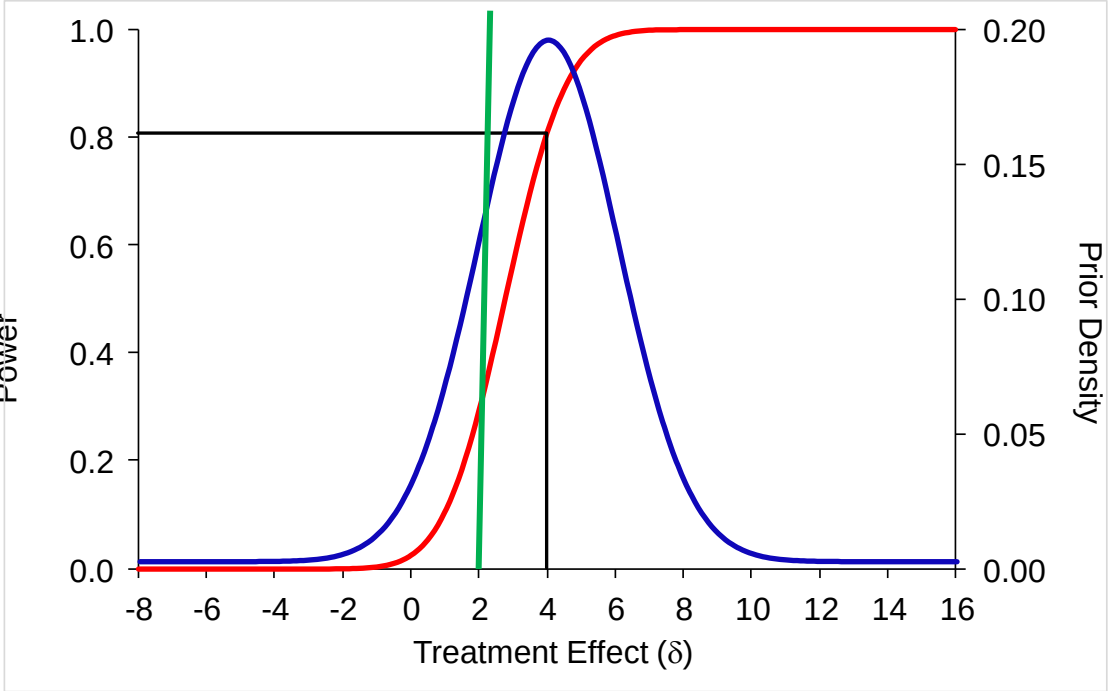
1. Write an R function that returns $p(\delta)$ for a given δ for the prior from Exercise 2

   Hint: use the R function `dnorm`

2. Write an R function that returns $P[\bar{y}_1 - \bar{y}_0 > \mathrm{Z}_{1-\alpha}\tau|\delta]$ (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.

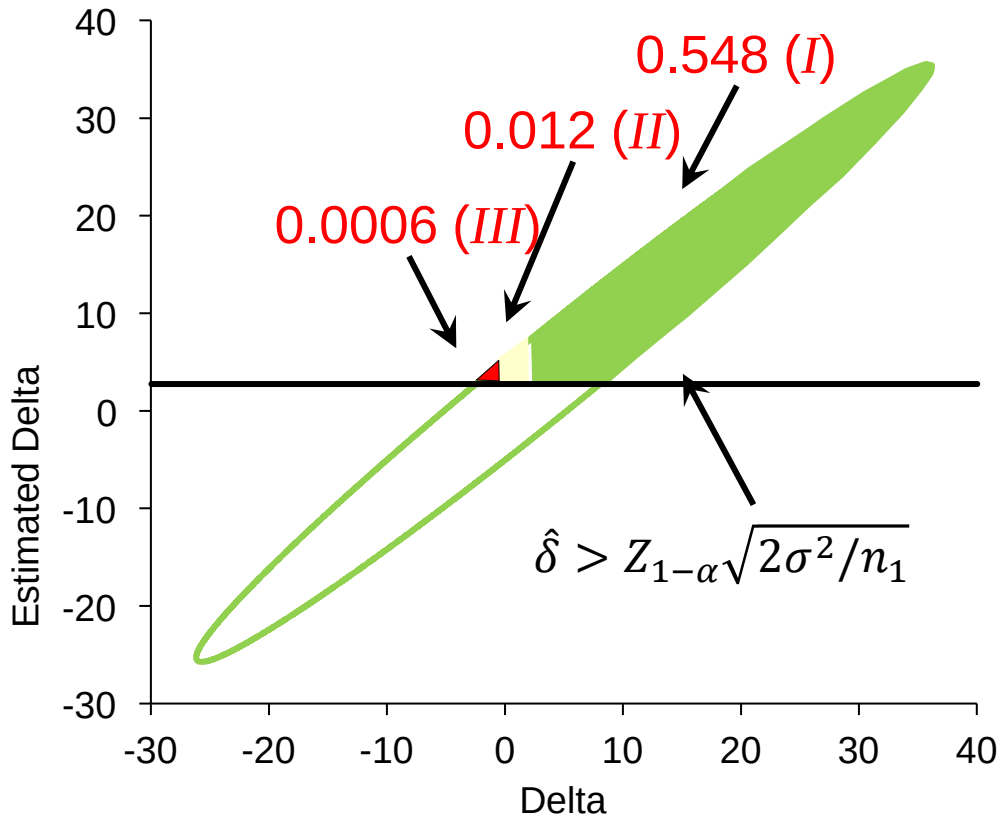
FERRING
PHARMACEUTICALS

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)



FERRING PHARMACEUTICALS

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

FERRING
PHARMACEUTICALS

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

A dose-finding example

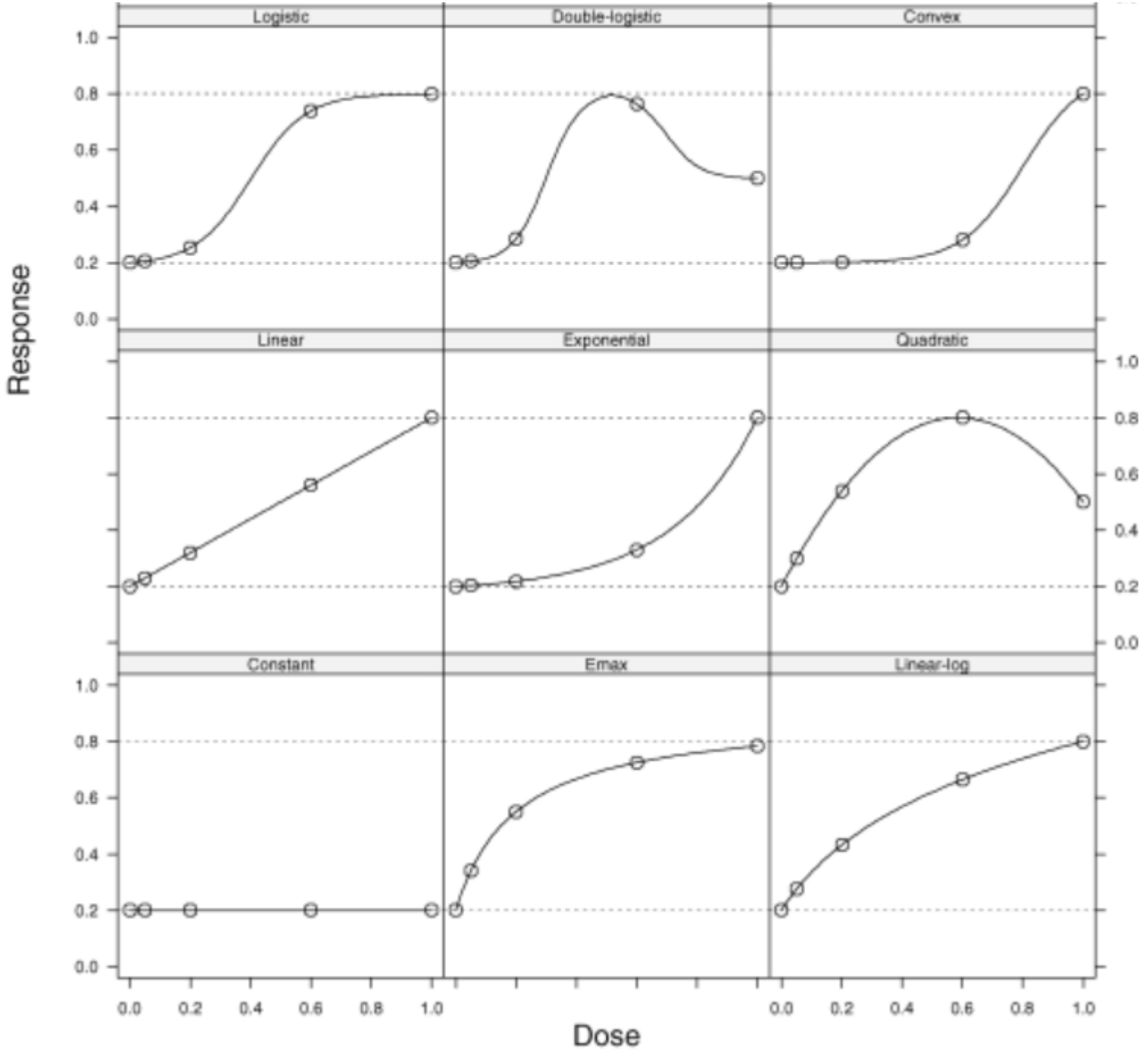
Imagine a dose-finding study where we aim to compare multiple doses against placebo to determine the optimal dose

Plausible dose-response shapes can be elicited from experts

The optimal choice of analysis method depends on whether or not the dose-response curve can be non-monotone

If asked, a clinician may not be willing to exclude a non-monotone curve as an option

A formal prior elicitation will help to quantify the likelihood of a non-monotone curve and impact on PoS for chosen method



PoS of a series of studies/conditional PoS in drug development

- Suppose we are at the end of Phase 1, we could determine
  - PoS (2), PoS (3(1)), PoS (3(2))
  - PoS (3(1) and 3(2)) – successful Phase 3 program
- We might also be interested in conditional success, e.g. to see by how much the phase II trial de-risks the program
  - PoS (3) | success in 2
  - PoS (3(1) and 3(2)) | success in 2

Results from the trial become dependent through the common prior. Hence their joint distribution needs to be considered.

For details see [Hybrid Frequentist/Bayesian Power and Bayesian Power in Planning Clini (taylorfrancis.com)]

Success can mean more than establishing efficacy

Example from drug development

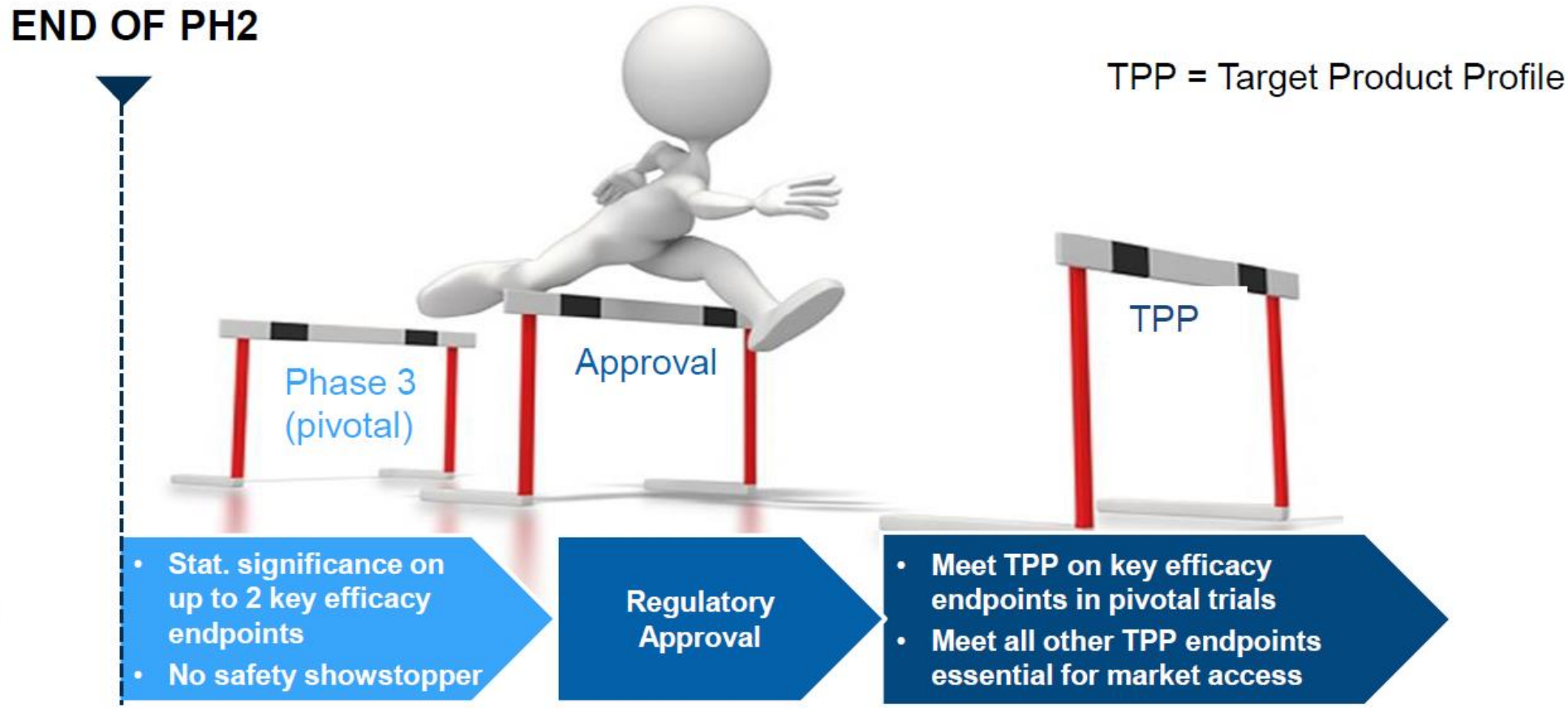


Image from Novartis slide deck presented at joint DSBS/FMS meeting (https://dsbs.dk/wp-content/uploads/2022/12/FMS-DSBS-2022.zip)

FERRING PHARMACEUTICALS

An (almost) real case study in Parkinson’s disease

FERRING
PHARMACEUTICALS

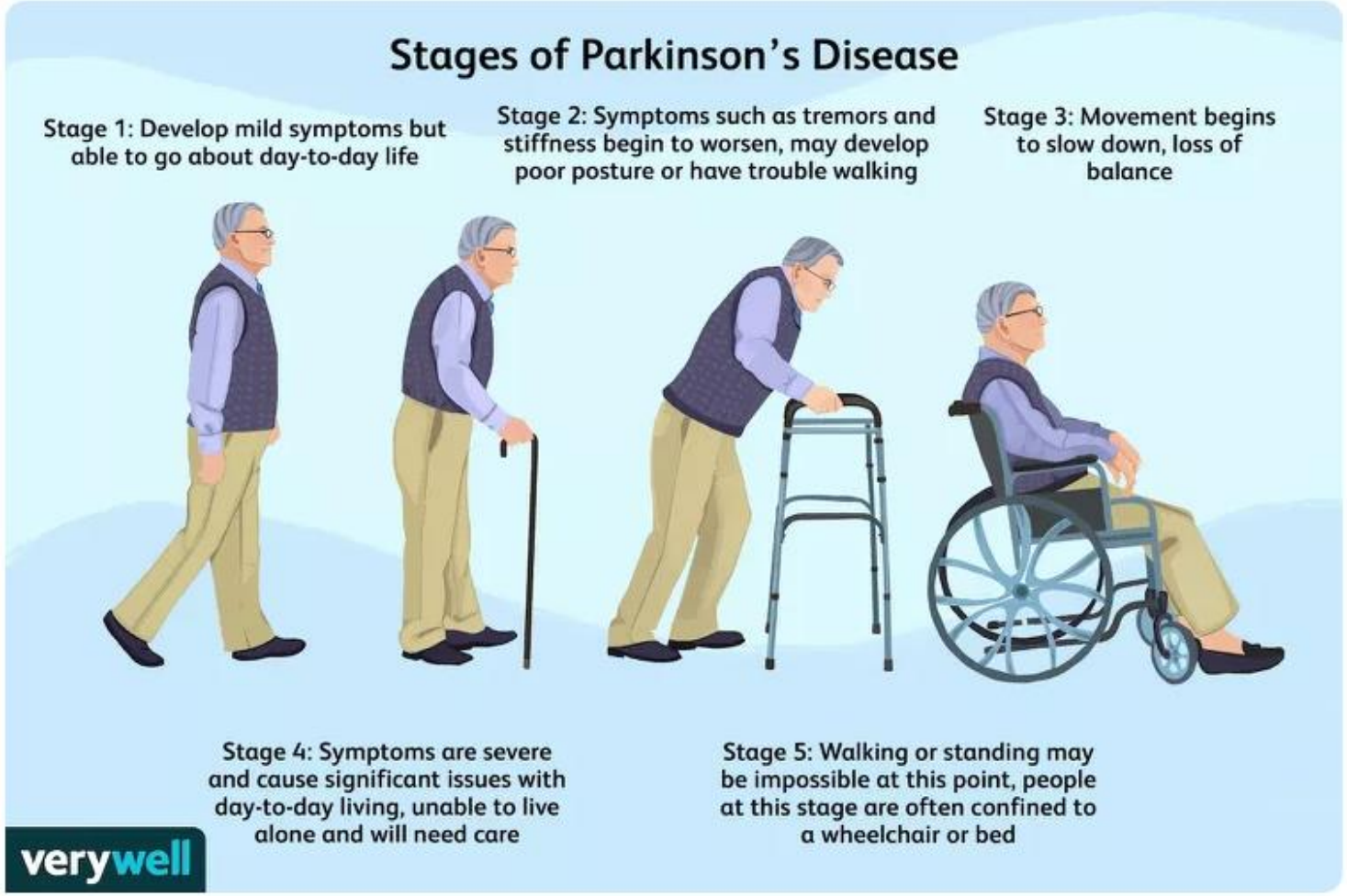
Parkinson’s Disease

Neurodegenerative disease

Parkinson’s disease is strongly associated with the loss of certain nerve cells in the brain that produce dopamine

OFF time: time during which patient experiences stiffness

Dyskinesia: involuntary, erratic, writhing movements of the face, arms, legs, or trunk



Verywell / Zoe Hansen

FERRING PHARMACEUTICALS

Case study

- A fictional drug (Drug L) that treats PD
- Aim to have an indication for the treatment of both
  - OFF time (measured in hours/day based on patient diary)
  - Dyskinesia (measured by Unified Dyskinesia Rating Scale (UDysRS), max score 104)
- It is a “me too” drug: it has a similar mode of action to an approved competitor drug (Drug C), but is expected to have certain advantages, e.g. higher efficacy due to better absorption and a better safety profile
- We will calculate the PoS for a single Phase II trial that aims to establish proof of concept

FERRING
PHARMACEUTICALS

Which scenarios constitute a success?

FERRING
PHARMACEUTICALS

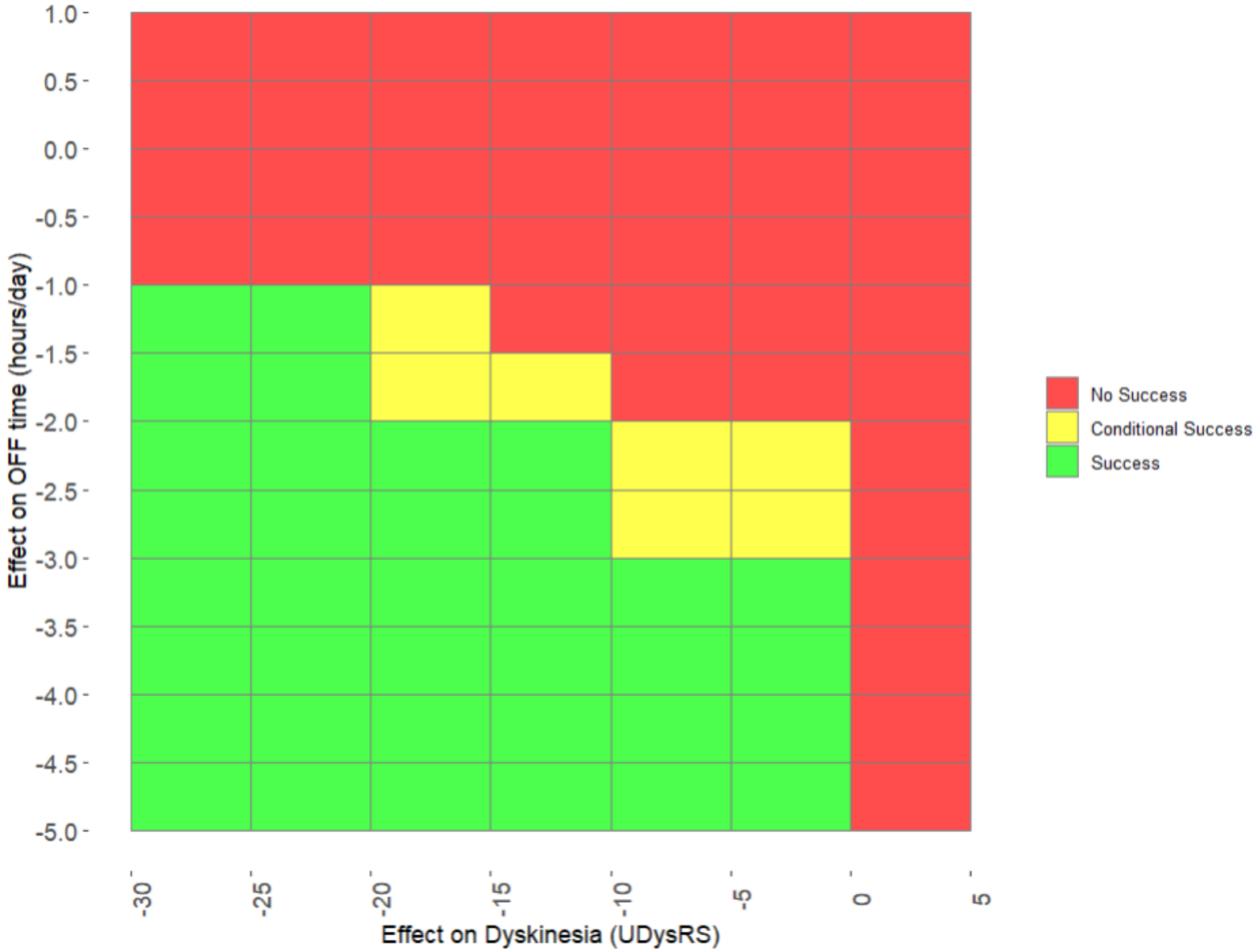
Success grid

Effect needed on OFF time:

- Drug C (and several other PD drugs) can reduce OFF time by 1 hour/day
- Minimum clinically relevant change on OFF time is 1 hour/day

Effect needed on Dyskinesia (UDysRS)

- Drug C reduces Dyskinesia by 15 points on UDysRS (no other drugs available)
- Minimum clinically relevant change on UDysRS is 10 points



FERRING PHARMACEUTICALS

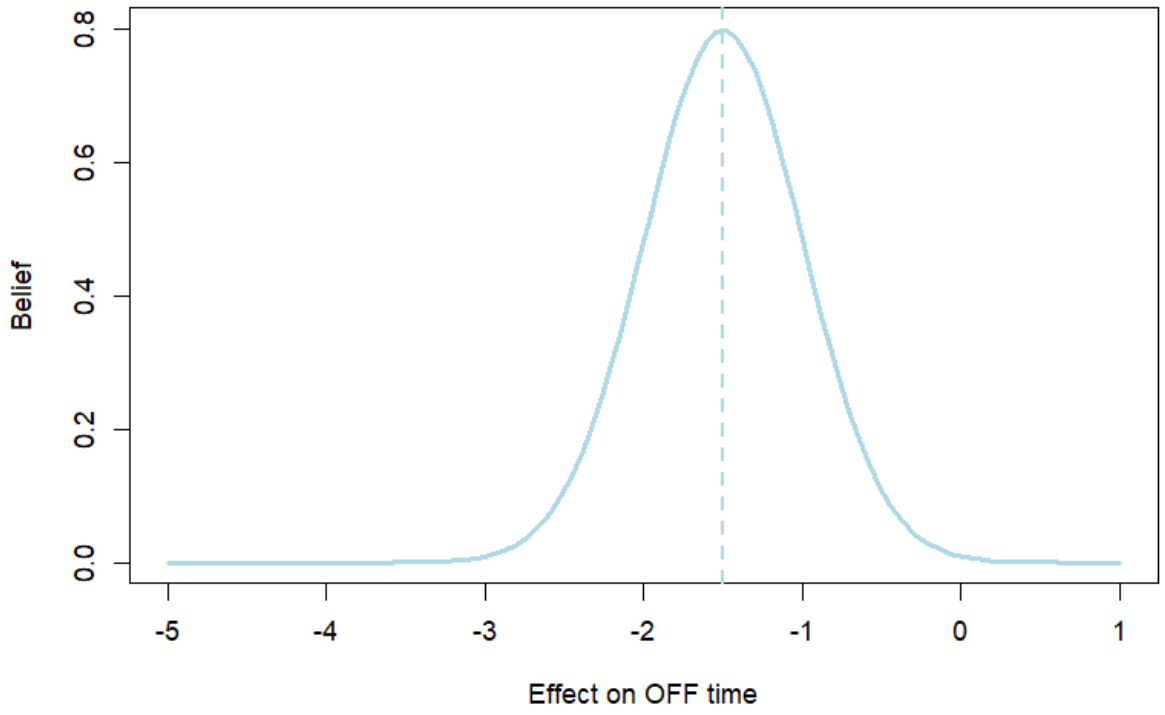
Summarizing current evidence

What effect do we expect?

FERRING
PHARMACEUTICALS

Prior for OFF time

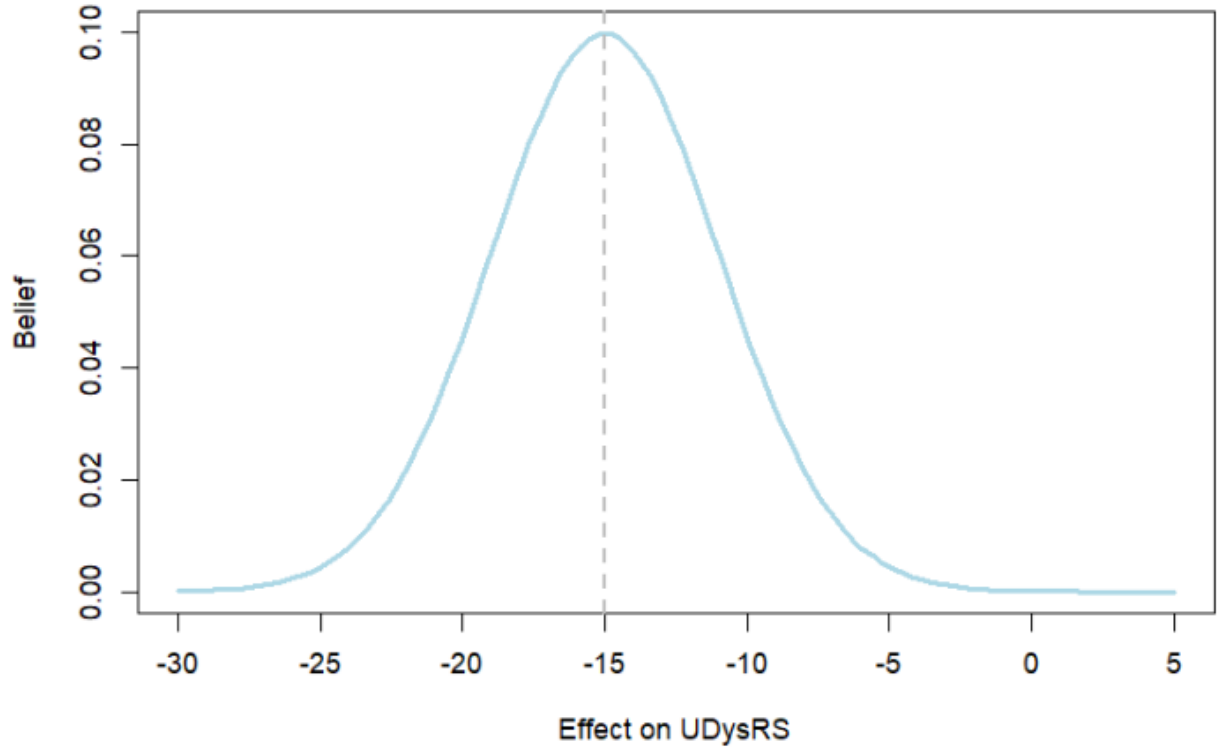
- Based on a meta-analysis of existing trials of Drug C the estimate of the effect on OFF time equals -1 (SE=0.5)
- Based on internal knowledge of Drug L, it is expected that the absorption of the drug will be slightly better than Drug C, resulting in a higher effect



FERRING
PHARMACEUTICALS

Prior for Dyskinesia (UDysRS)

- Based on a meta-analysis of existing trials of Drug C the estimate of the effect on the UDysRS equals -15 (SE=4)
- Based on internal knowledge of Drug L, it is believed that the effect on dyskinesia will be similar to Drug C

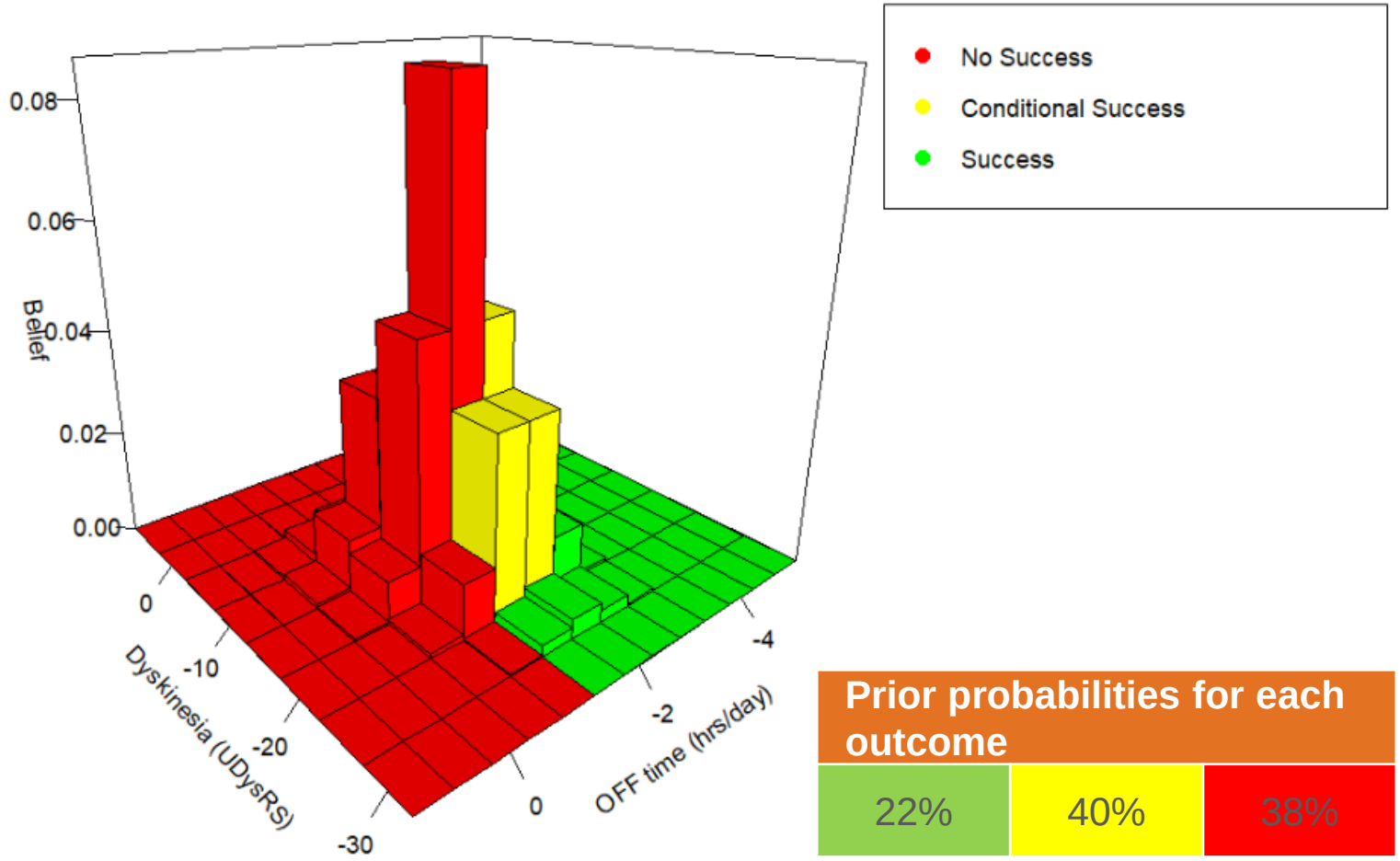


FERRING PHARMACEUTICALS

Joint prior and prior PoS

Based on internal data, the correlation between the changes from baseline in OFF time and UDysRS was estimated to be 0.4

- Higher beliefs indicate that the values are more likely
- The colouring is according to the grid of the success criteria, indicating the success outcome under each of the possible scenarios



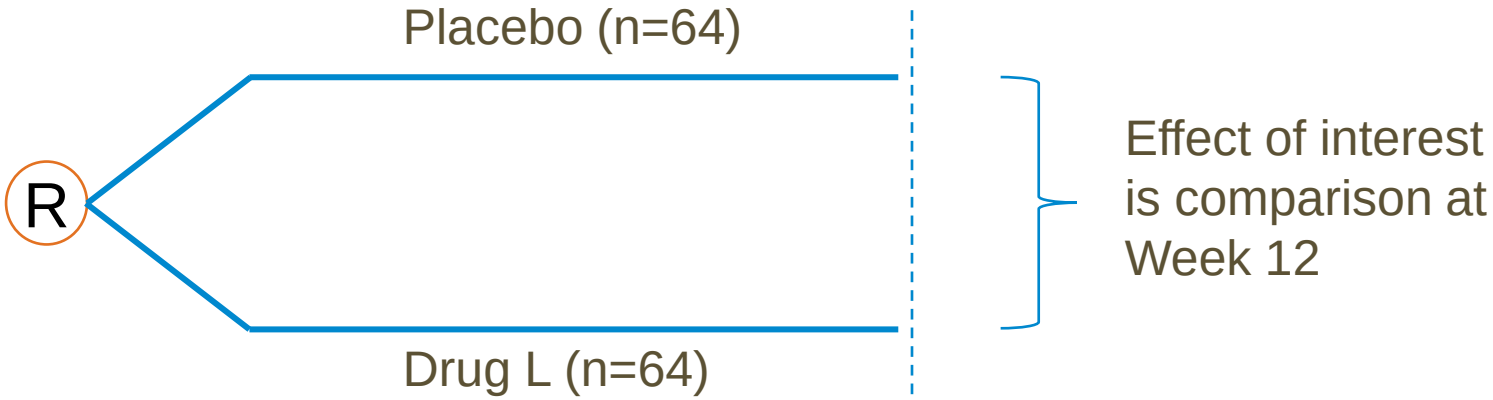
Prior probabilities for each outcome		
22%	40%	38%

Evaluating the PoS for the trial

FERRING
PHARMACEUTICALS

Study Design

- For this example, a simple T-test is used to compare results on OFF time and UDysRS between arms at week 12.
- The study is powered at 80% for the primary endpoint (OFF time), with a two-sided significance level of 5%, a target effect size of -1.5 and an SD of 3
- UDysRS is a more sensitive endpoint than OFF time, so power is OK



Week 12
Primary: change from baseline in OFF time
Secondary: change from baseline in UDysRS

FERRING PHARMACEUTICALS

Results on PoS

Probabilities of success according to different criteria for a successful trial:

Success criteria for the trial	PoS (10000 sims)	PoS (analytical, fixed σ)	Proportion of true success
Significant results on both endpoints	72%	73%	20%
Significant results on both endpoints Estimates in Success scenario	29%	30%	15%

Prior PoS (upper bound for true success)

Prior probabilities for each outcome		
22%	40%	38%

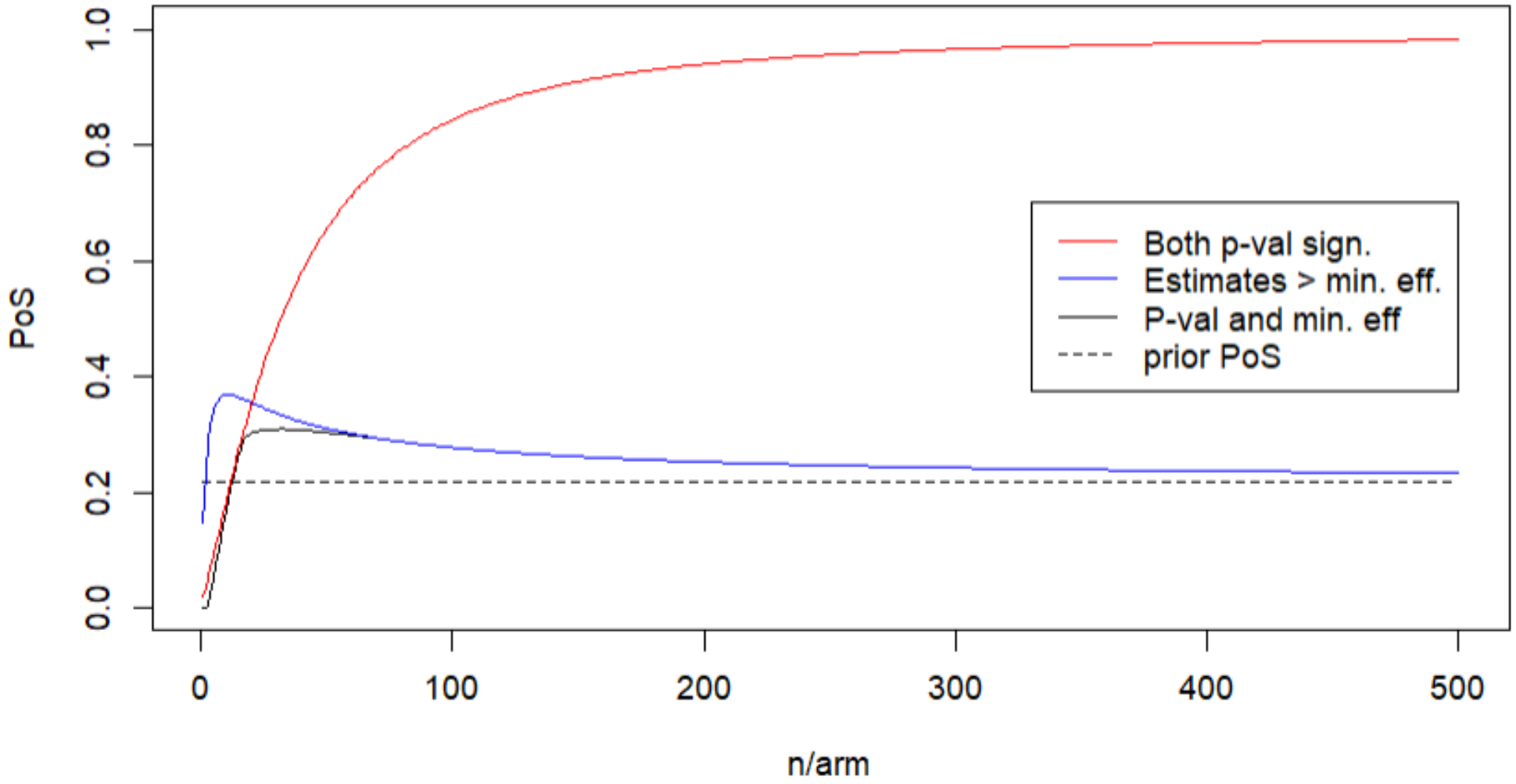
Example where probability of success and probability of **true** success differ!

Illustrates that trial success criteria may need to be refined.

FERRING
PHARMACEUTICALS

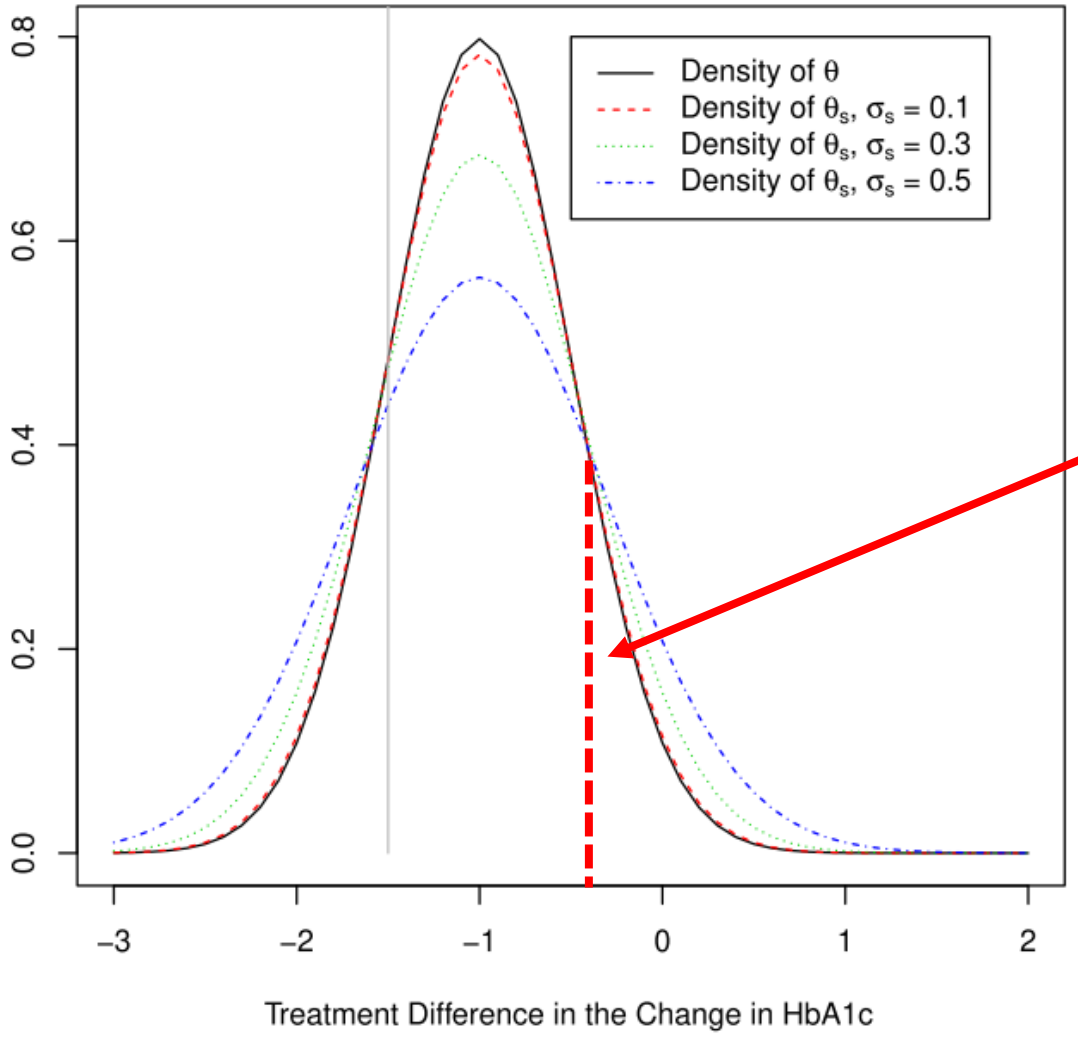
PoS for case study with increasing sample size

The PoS for the trial was higher than the prior PoS – proposed trial success criteria too optimistic



FERRING
PHARMACEUTICALS

The probability of obtaining an effect estimate ≥ X can be larger for smaller sample sizes if the cutoff is extreme



Minimum relevant effect cutoffs beyond this point will result in non-monotone PoS (smallest sample sizes giving higher PoS)

Using a dual criteria approach will reduce (but not resolve) this issue

Issue most relevant for more exploratory trials where the prior might have a lot of weight on scenarios that don't constitute a success

To be aware of when **not** using the probability of a **true** success!

See also: Qu, Y., Du, Y., Zhang, Y., & Shen, L. (2020). Understanding and adjusting for the selection bias from a proof-of-concept study to a more confirmatory study. *Statistics in Medicine*, 1–12.

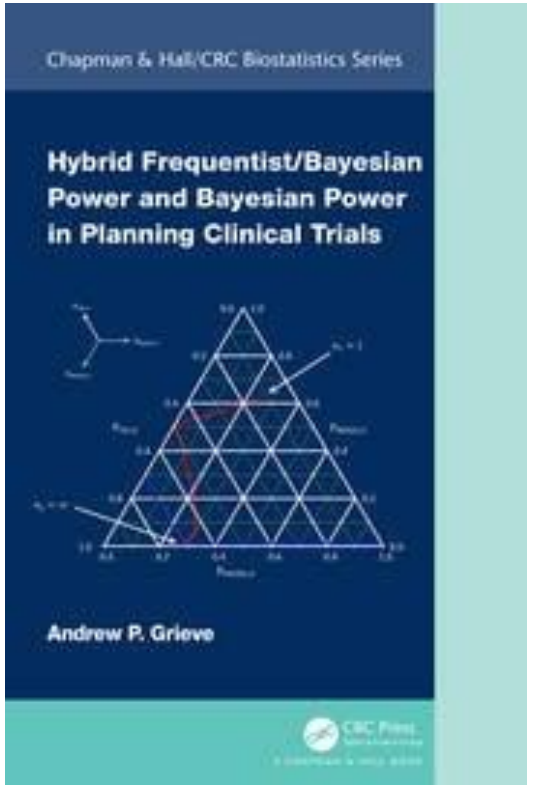
FERRING PHARMACEUTICALS

Further reading

[Hybrid Frequentist/Bayesian Power and Bayesian Power in Planning Clini (taylorfrancis.com)]

By Andrew P. Grieve

Very nice book covering most of the material from this session.



FERRING
PHARMACEUTICALS

Coding aid

This is the code for Exercise 2.1, it will be helpful for remaining exercises as well

```
pwr <- function(sigma,delta,alpha,beta,n){                #function that returns power
  pnorm(qnorm(alpha)+(sqrt(n)*delta)/(sqrt(2)*sigma))
}

assurance <- function(sigma,alpha,n,sigma_prior,mu_prior){  #function that returns assurance
  tau <- sigma*sqrt(2/n)
  pnorm((qnorm(alpha)*tau+mu_prior)/sqrt(sigma_prior^2+tau^2))
}

#creating the plot
n <- seq(0,1000,1)
dim(n) <- c(1,length(n))
pwrs <- apply(n,2,pwr,sigma=6.5,delta=2,alpha=0.025,beta=0.1)
asnc <- apply(n,2,assurance,sigma=6.5,alpha=0.025,sigma_prior=2,mu_prior=2)

plot(n,pwrs,type="l",xlab="Sample size per arm",ylab="PoS",lwd=1.5)
lines(n,asnc,lwd=1.5,col="blue")
legend("bottomright",c("Power","Assurance"),col=c("black","blue"),lty=1,lwd=1.5)
```

FERRING PHARMACEUTICALS