EVALUATING THE PROBABILITY OF SUCCESS OF A TRIAL OR PROGRAM BASED ON MULTIPLE ENDPOINTS

Corine Baayen - 22 Nov 2022 - Join DSBS/FMS meeting

Why evaluate the probability of success (PoS)?

- Support portfolio management
  E.g. prioritise between projects depending on risk, costs and reward

- Choose between several options for a development program
  E.g. choose between sample sizes, types of study designs

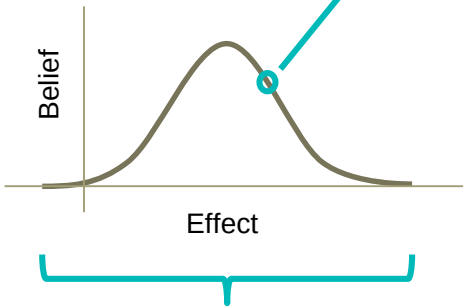
Aim is to support internal decision-making

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?

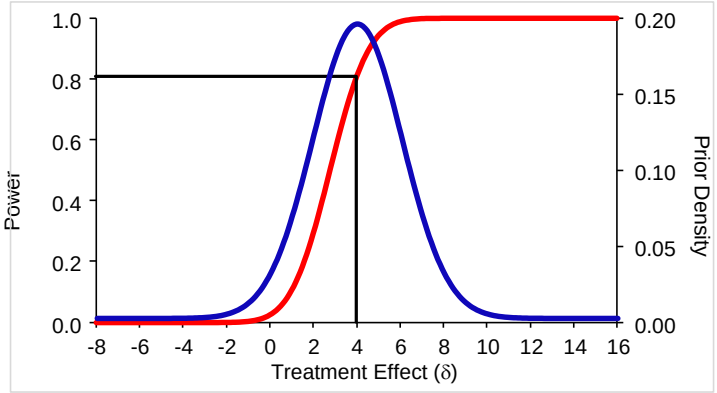
Assurance:

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.

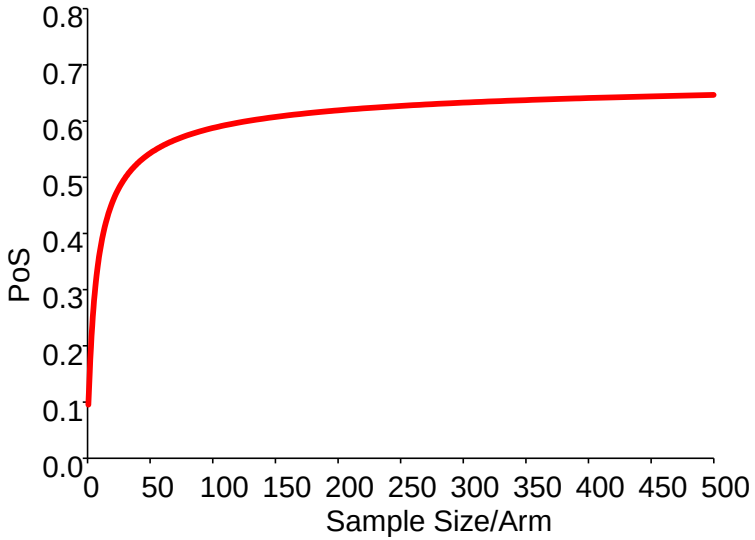
Assurance – simple case with success based on p-value

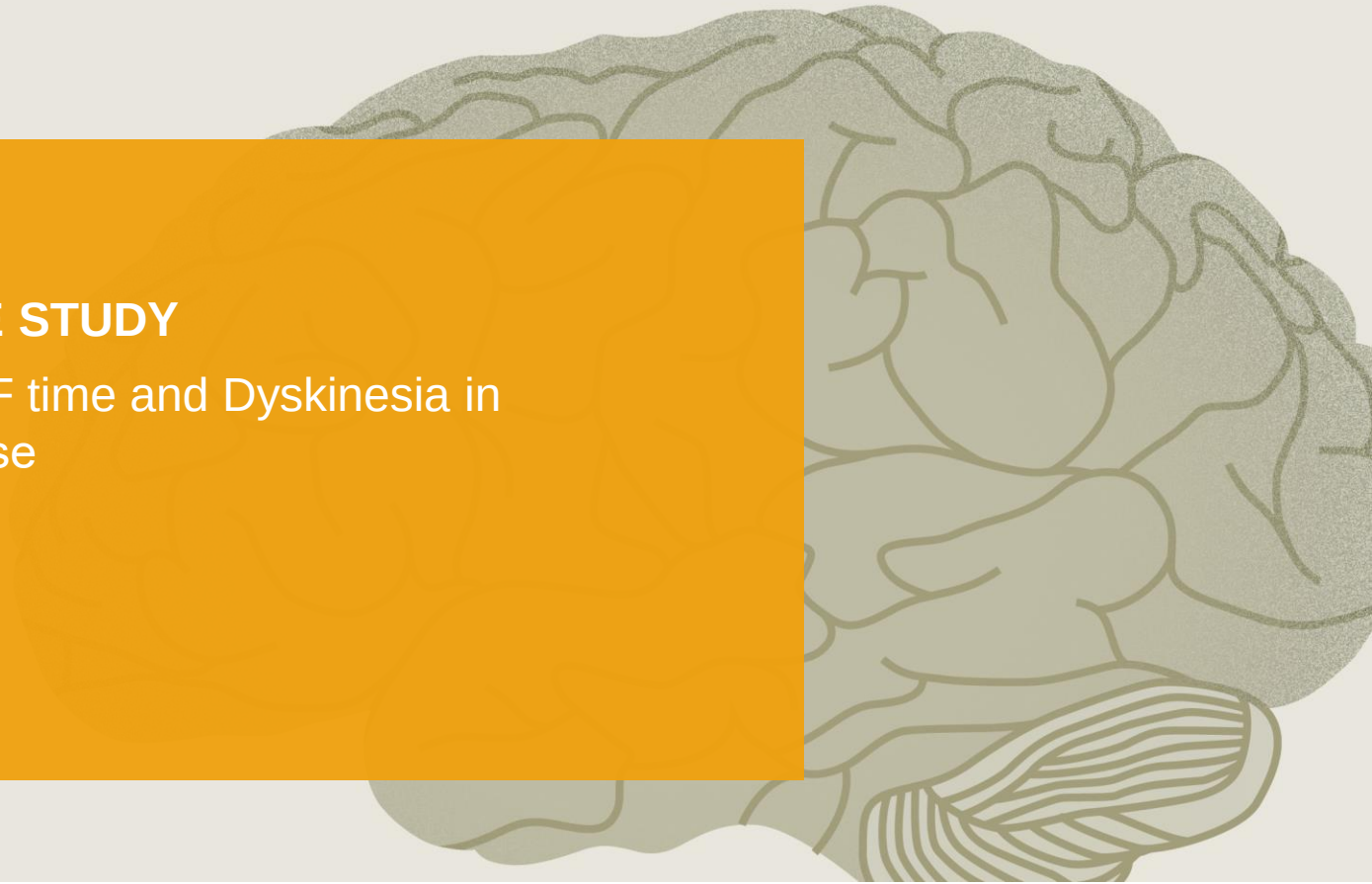


- The power $1 - \beta(\delta)$ is calculated conditional on δ being a specific value
- Assurance is the unconditional power: $\int_{\delta} \big(1 - \beta(\delta)\big) p(\delta) d\delta$ where $p(\delta)$ is the prior distribution

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



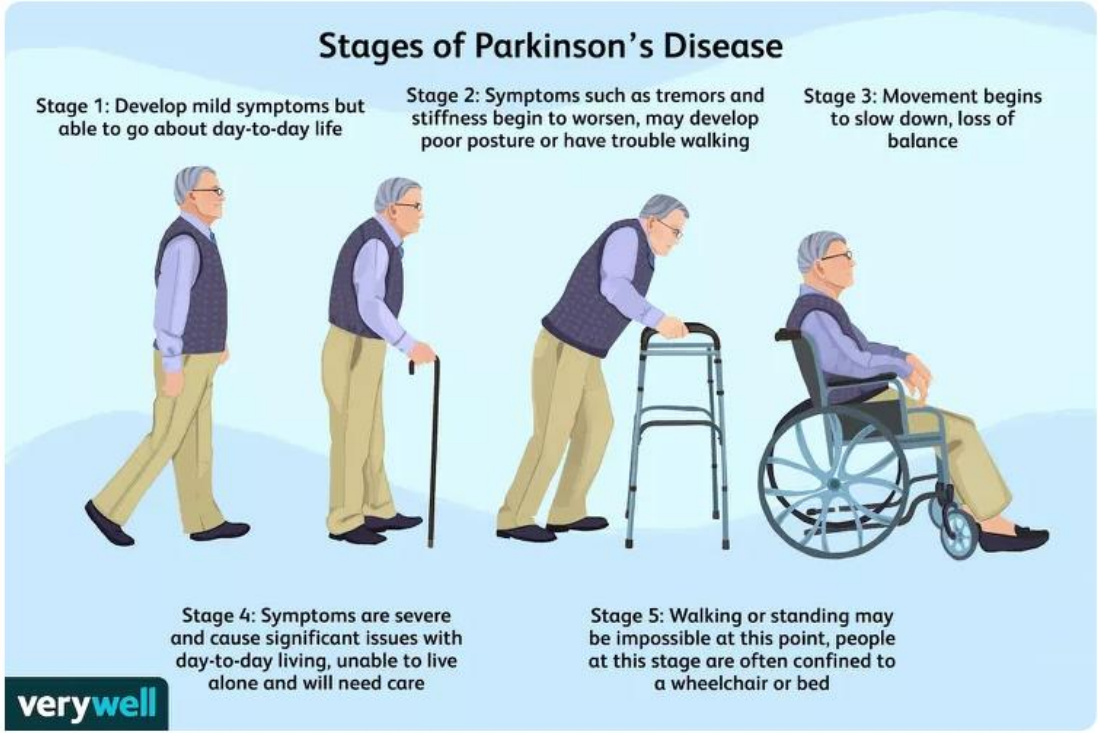


FICTIONAL CASE STUDY

A drug to treat OFF time and Dyskinesia in Parkinson’s Disease

Parkinson’s Disease

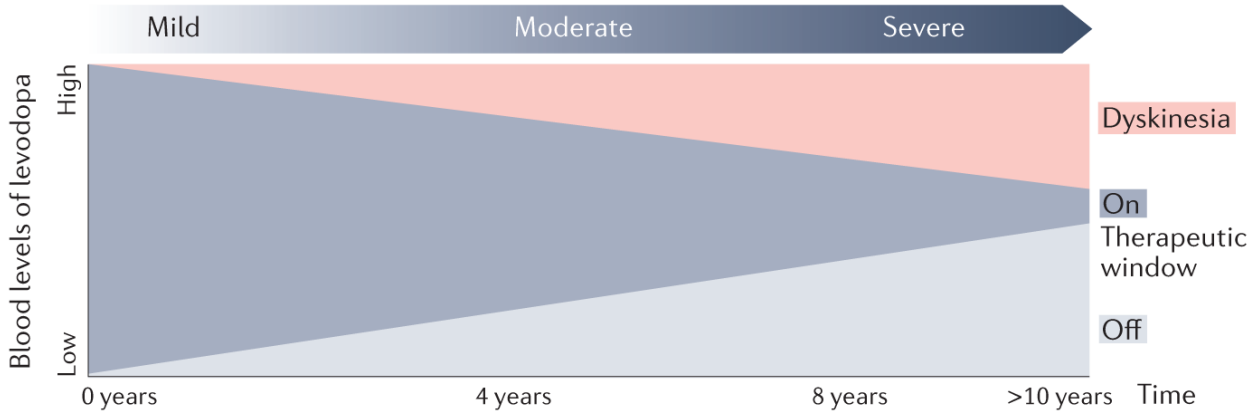
- Neurodegenerative disease
- Parkinson’s disease is strongly associated with the loss of certain nerve cells in the brain that produce dopamine
- Average age of onset is 60 years old. However, it can occur in younger adults between 30 and 40 years old



Verywell / Zoe Hansen

Finding a balance between OFF time and Dyskinesia

- Two major symptoms of PD patients are OFF time and Dyskinesia
- Standard of care Levodopa treatment can reduce OFF time, but increase Dyskinesia

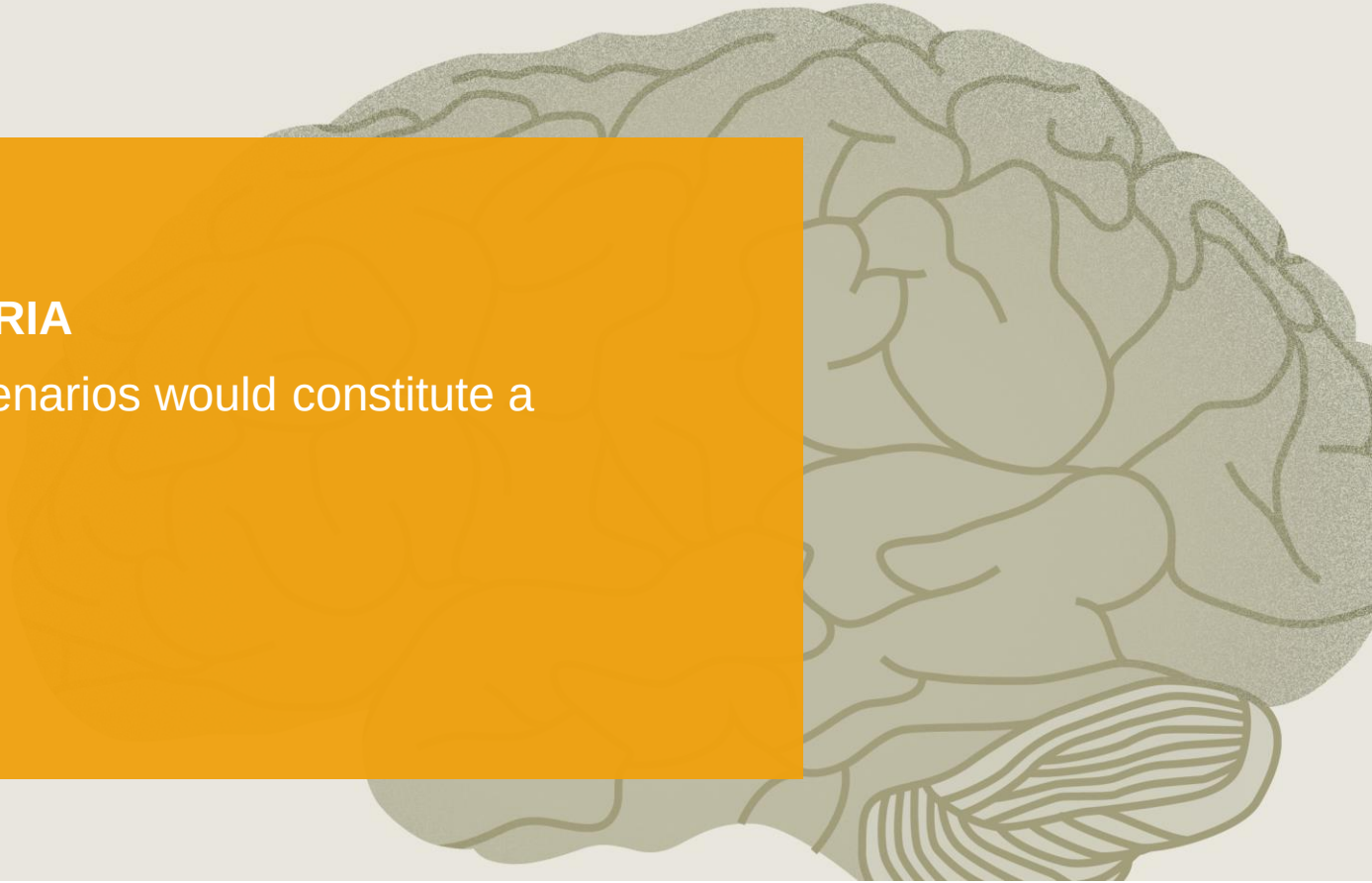


Case study

- 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

SUCCESS CRITERIA

Which possible scenarios would constitute a success?



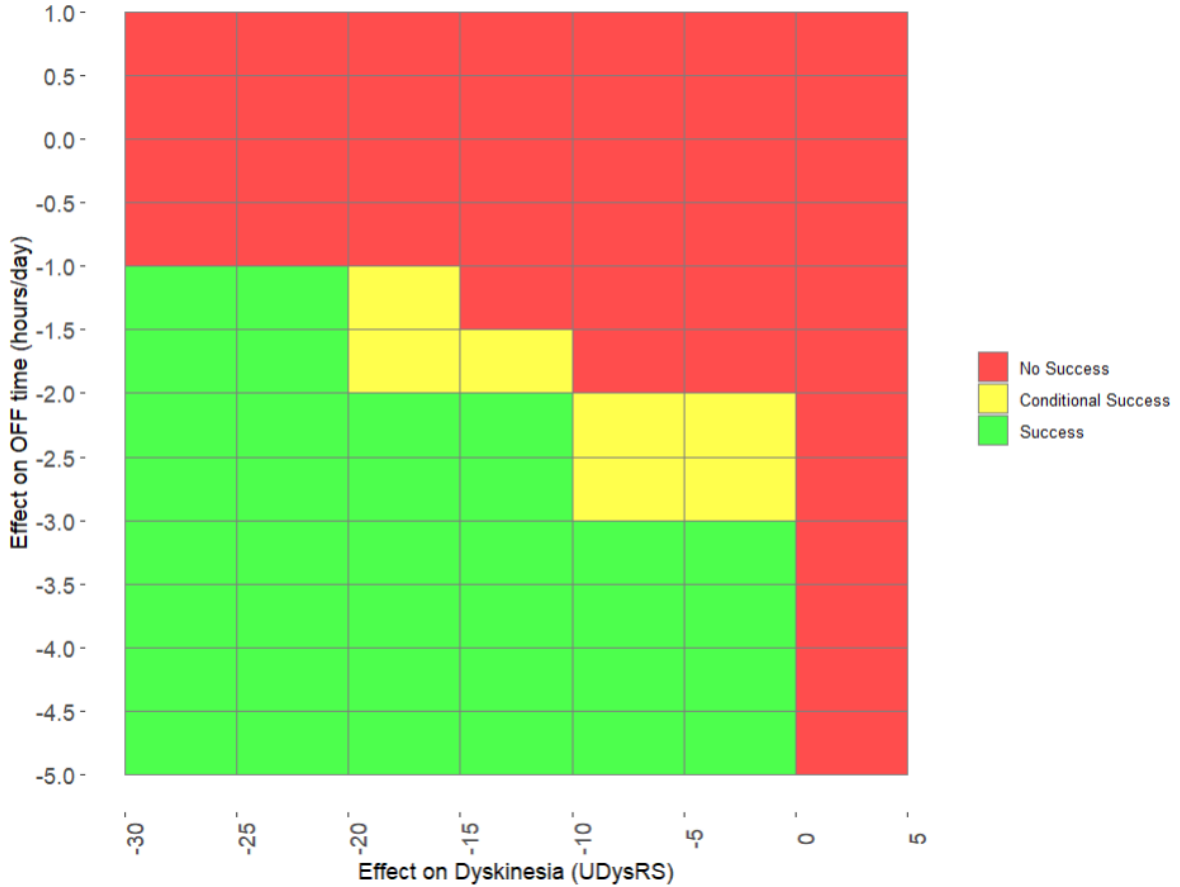
Success grid

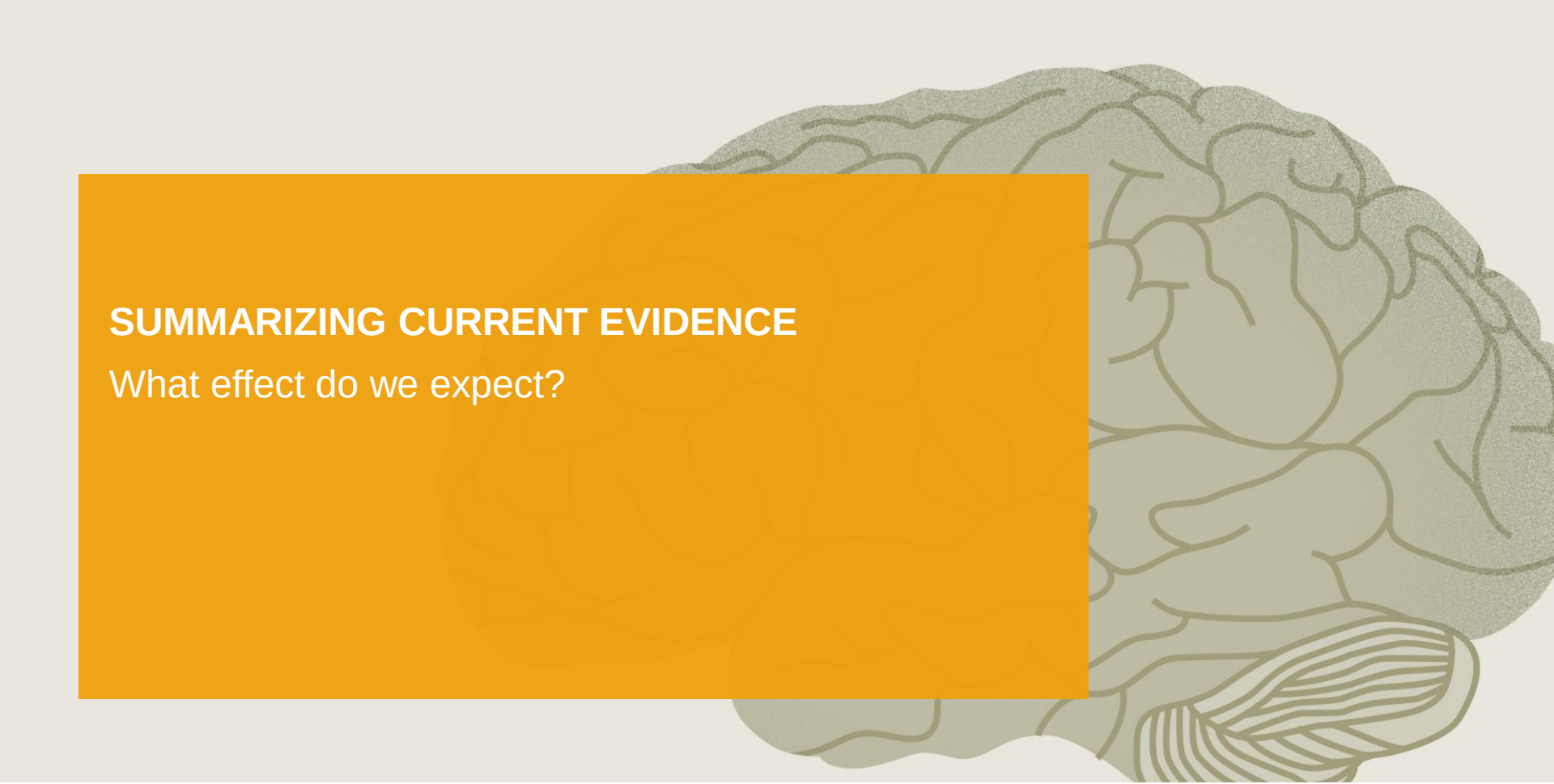
Effect needed on OFF time:

- 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



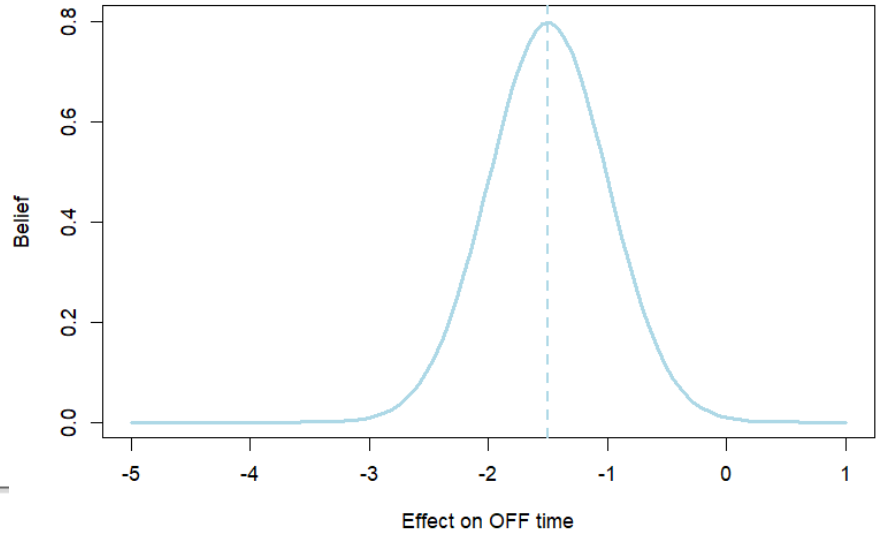


SUMMARIZING CURRENT EVIDENCE

What effect do we expect?

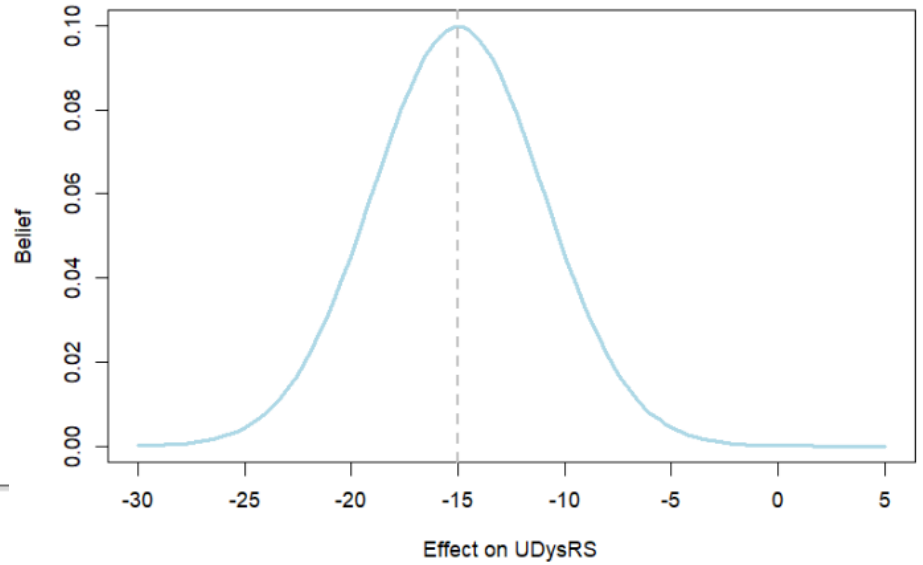
Prior for OFF time

- 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



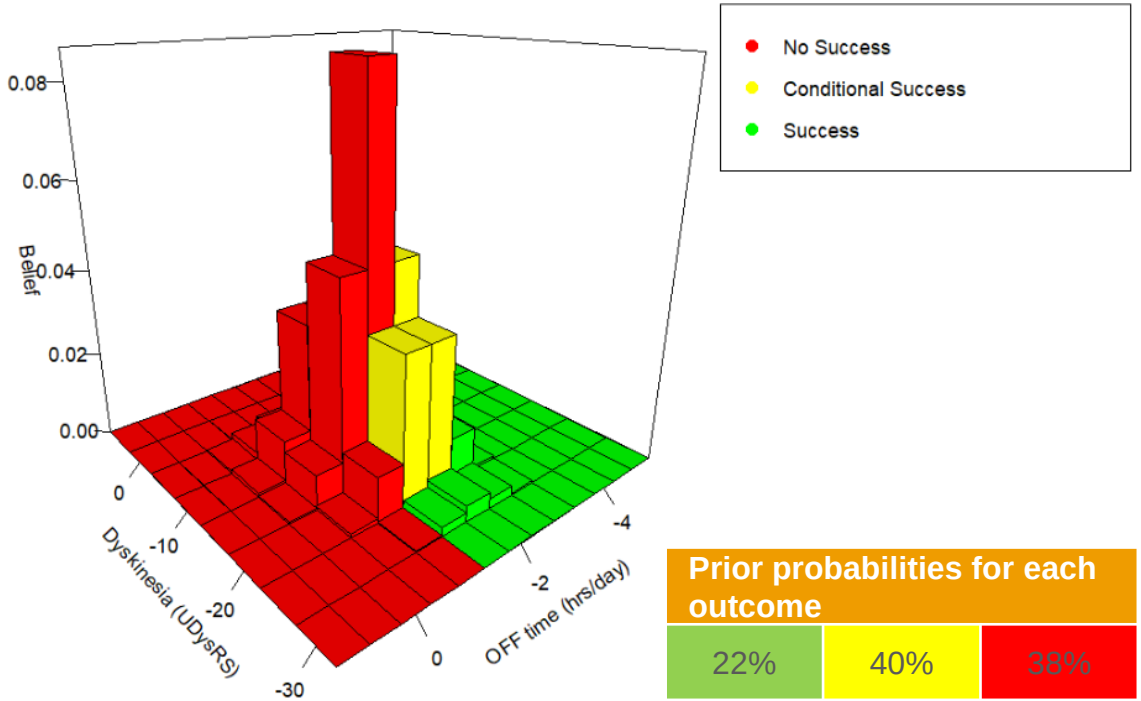
Prior for Dyskinesia (UDysRS)

- 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



Joint prior and prior PoS

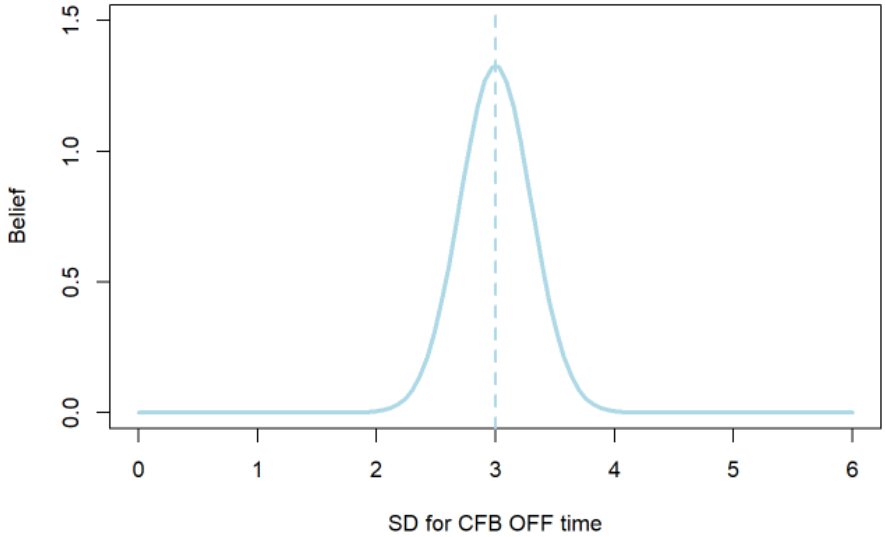
- Based on internal data, the correlation between the changes from baseline in OFF time and UDysRS was estimated to be 0.4
- Higher beliefs indicate that the values are more likely
- The colouring is according to the grid of the success criteria, indicating the success outcome under each of the possible scenarios

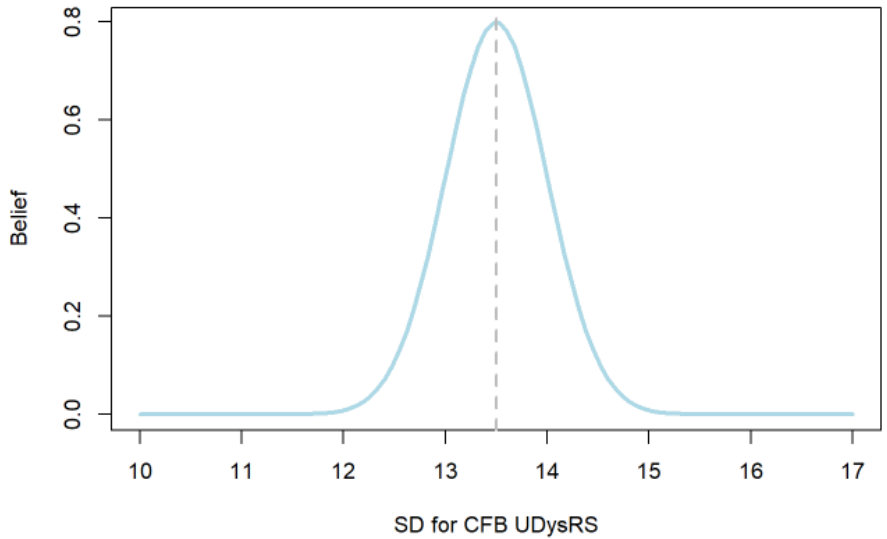


Prior probabilities for each outcome		
22%	40%	38%

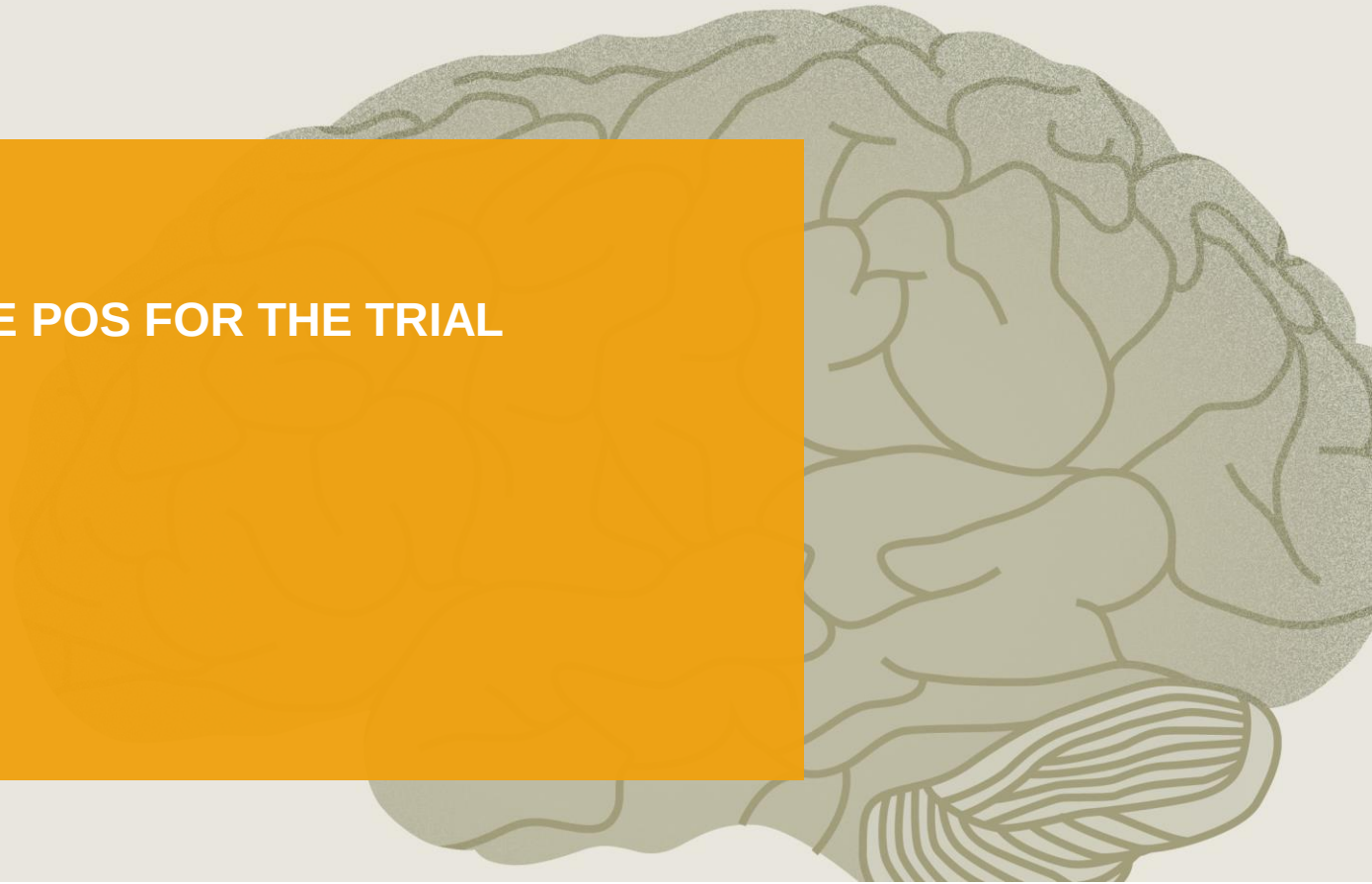
Priors for SDs for study endpoints

- The SD for the changes from baseline in OFF time and the UDysRS to Week 12 varies from study to study, therefore, a prior was also used for these parameters



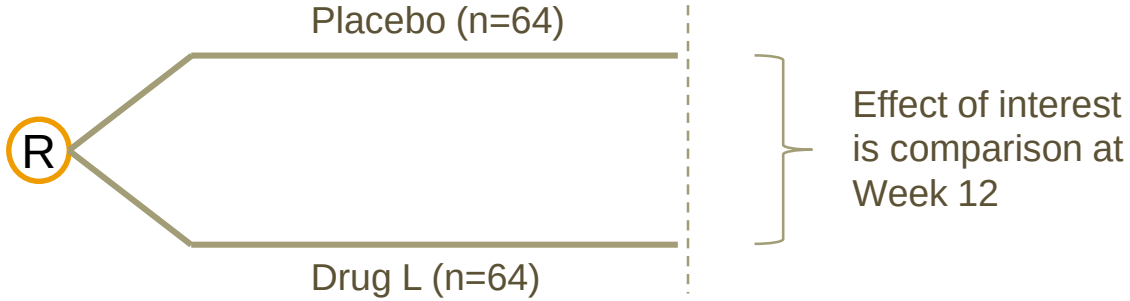


EVALUATING THE POS FOR THE TRIAL



Study Design

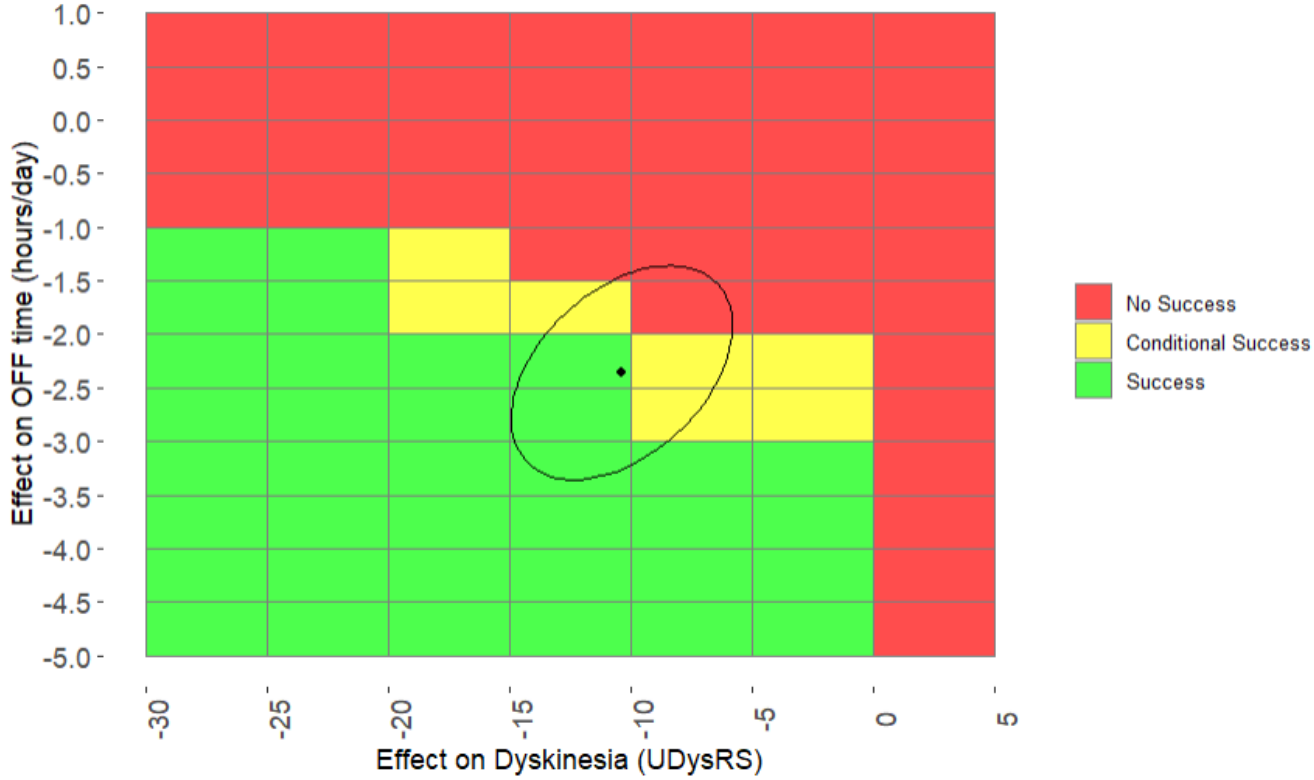
- For this example, a simple T-test is used to compare results on OFF time and UDysRS between arms at week 12.
- The study is powered at 80% for the primary endpoint (OFF time), with a two-sided significance level of 5%, a target effect size of -1.5 and an SD of 3
- UDysRS is a more sensitive endpoint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



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

One possible outcome of a single trial

- The black dot indicates the estimated effect on each outcome
- The black ellipse indicates the 95% confidence area



Confidence in each outcome		
49%	32%	19%

When would we consider a single trial a success?

1. If the point estimate is in the green area and the p-values are significant for both outcomes?

Interpretation of 1: the effect is different from 0 and the best guess for the treatment effect would result in a success

2. If we are > XX% (e.g. 60%) confident that the effect is in the green area?

3. If we are > XX% (e.g. 60%) confident that the effect is in the green or yellow area?

Interpretation of 2 and 3: the confidence that the true effect would result in a success (or success or conditional success for option 3) is at least XX%

We are more used to think in terms of Option 1, but at the end of Phase II, Option 2 or 3 might be what we are more interested in, since it takes the uncertainty around the estimate into account

Idea behind simulating the PoS of the study

- Draw a scenario (effect and SD for OFF time and UDysRS) from the prior distributions – more likely scenarios will be drawn more often
- Simulate the trial data based on the drawn scenario
- Assess whether the simulated trial would be a success
- Do this many times (e.g. 100000) and assess the proportion of successful trials

In the end we will have an evaluation of how often we would see a success given our uncertainty on the effects on OFF time and UDysRS

Note: without priors on the SDs analytical computation of the PoS is quite straightforward

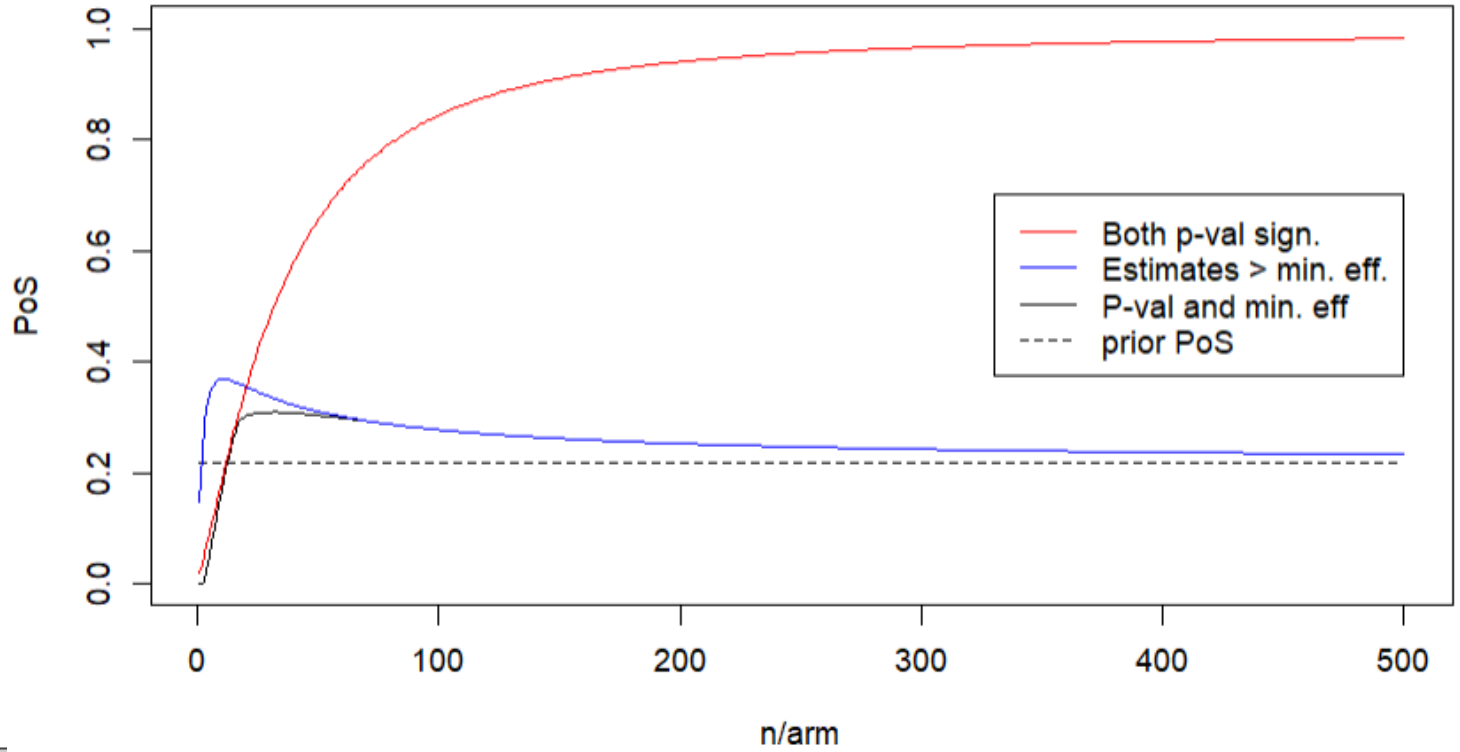
Results on PoS

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

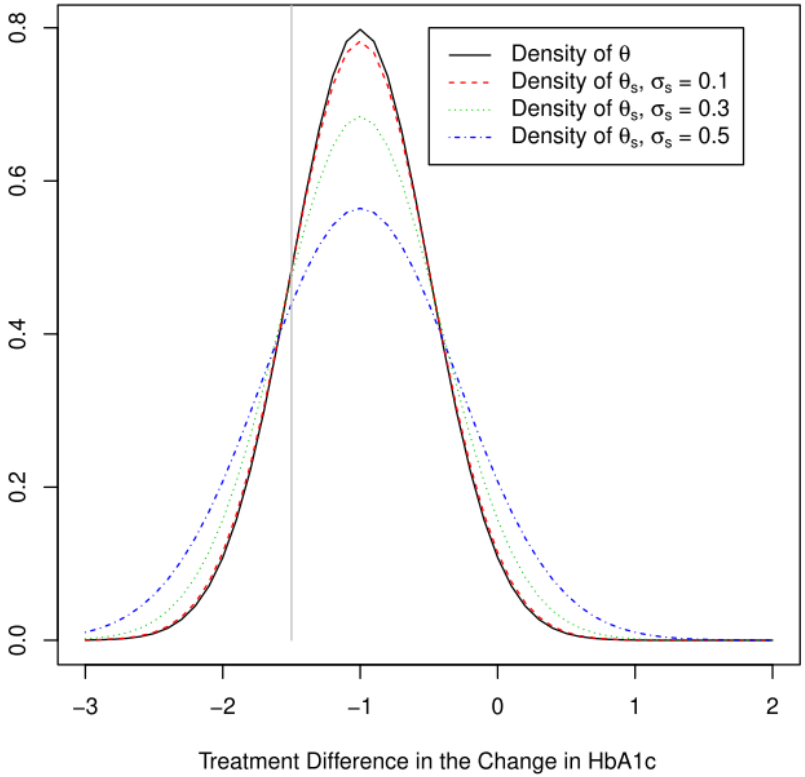
Success criteria for the trial	PoS (10000 sims)	PoS (analytical, fixed σ)	Proportion of true success
Significant results on both endpoints	72%	73%	20%
Significant results on both endpoints Estimates in Success scenario	29%	30%	15%
Significant results on both endpoints Estimates in Success or Conditional Success scenario	57%	58%	53%
Confidence that effect is in Success scenario is 60%	23%		13%
Confidence that effect is in Success or Conditional Success scenario is 60%	50%		44%

Prior probabilities for each outcome		
22%	40%	38%

PoS for case study with increasing sample size

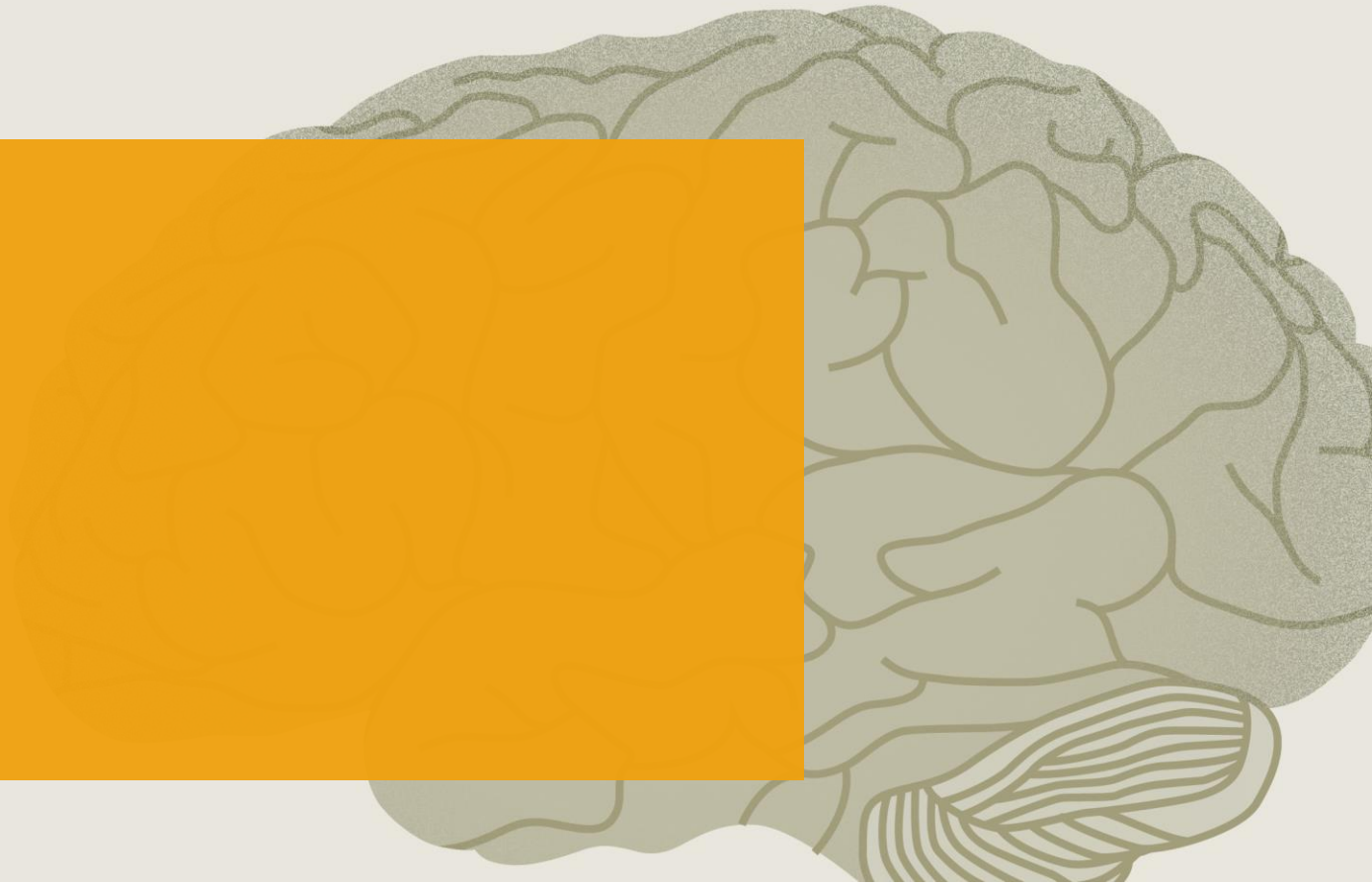


The probability of obtaining an effect estimate ≥ X is larger for smaller sample sizes



- 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.

DISCUSSION



Discussion on use of assurance

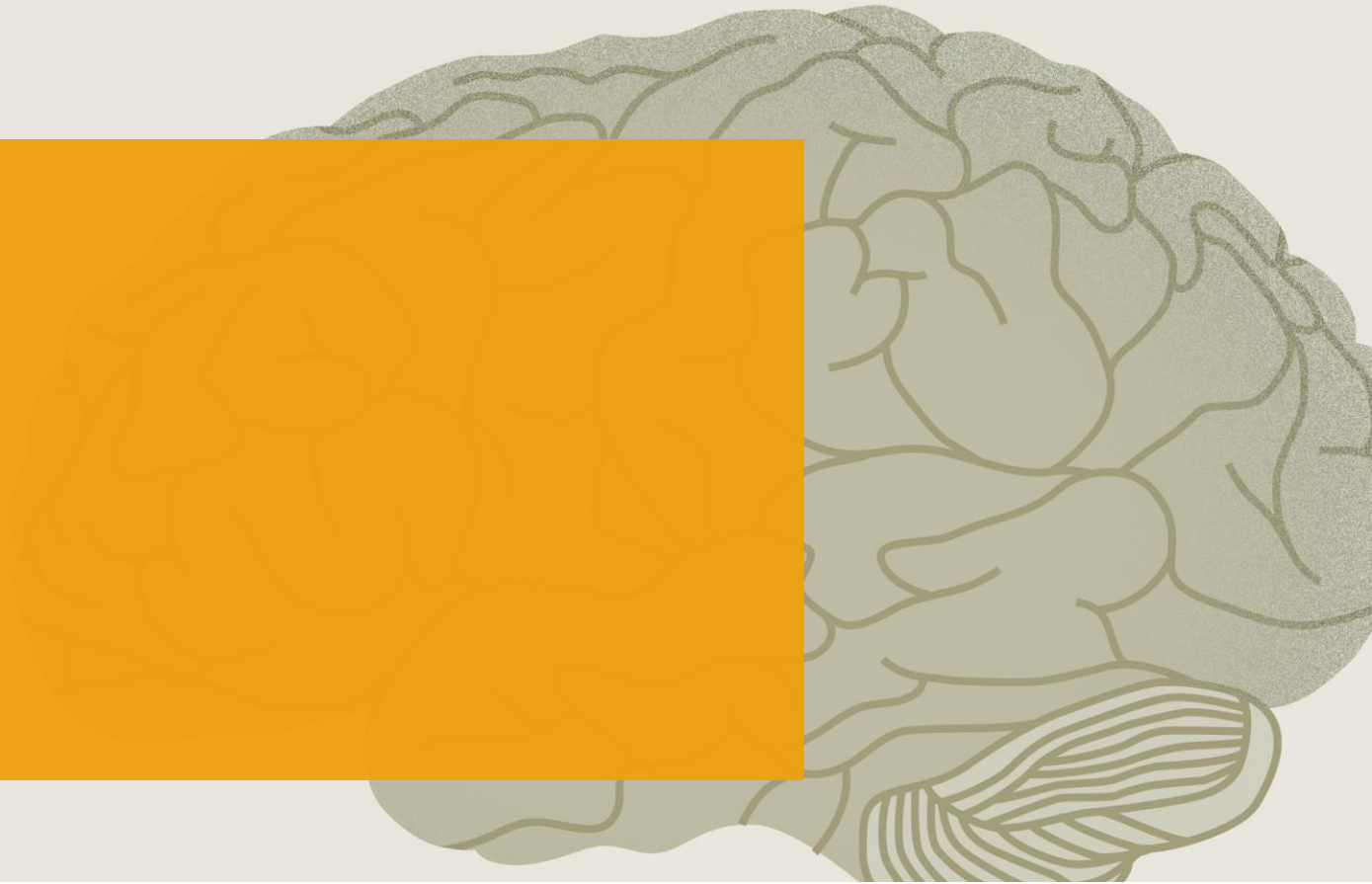
- One can easily extend the approach to assess the PoS of an entire program (e.g. the PoC study and 2 Phase III studies)
  - Can also calculate the PoS of the full development program, conditional on a successful PoC study (how much are we de-risking the program with this PoC study?)
  - Will need to down weight the results from small early phase studies*
- Priors on the SD don't seem to impact the results much**
- Discrepancy between p-value based on null hypothesis of no effect and success criteria based on minimum clinically relevant effect

*Qu, Y. et al. (2020). Understanding and adjusting for the selection bias from a proof-of-concept study to a more confirmatory study. Statistics in Medicine, 1–12.
**Walley et al. (2015). Advantages of a wholly Bayesian approach to assessing efficacy in early drug development: A case study. *Pharmaceutical Statistics*, *14*(3), 205–215.

Benefits of using this approach

- 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

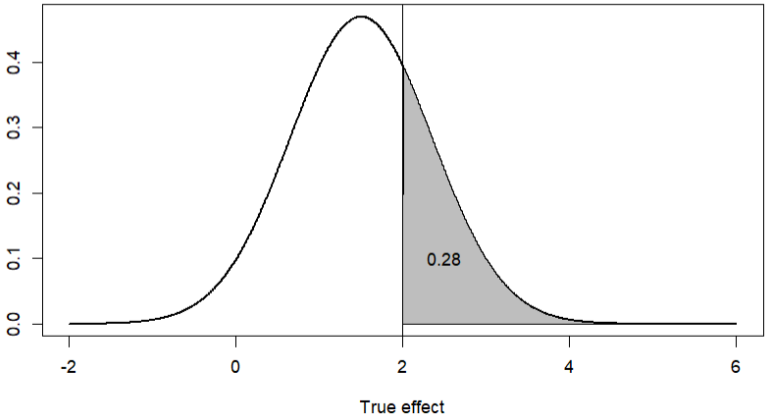
BACKUP

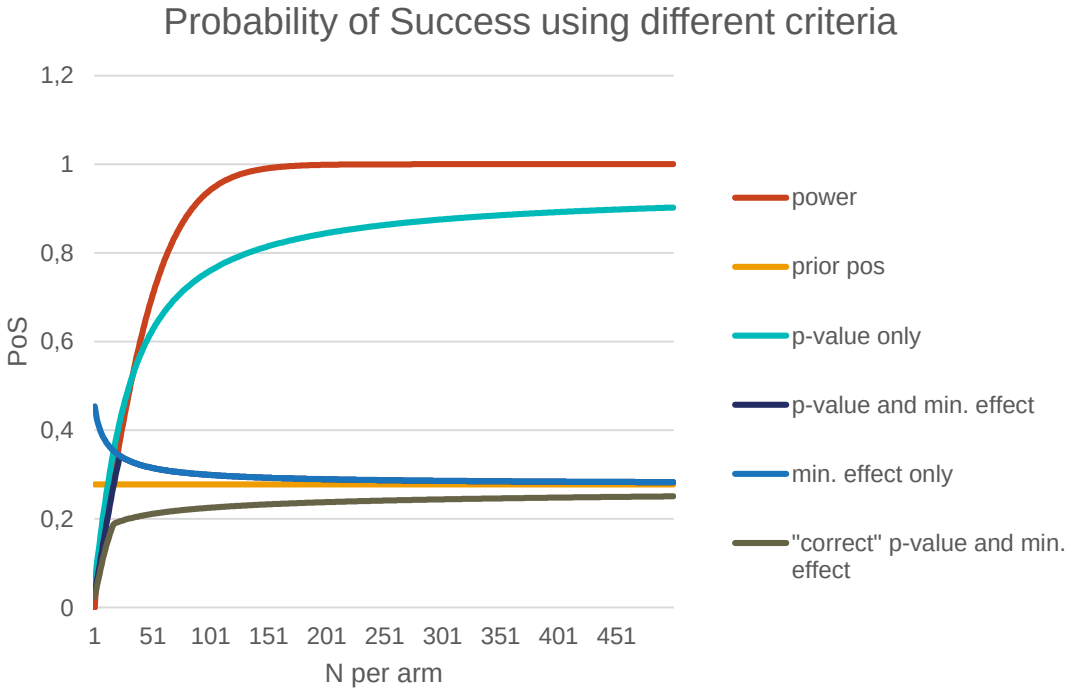


Assurance adding requirement that the observed effect is ≥ a minimum relevant effect

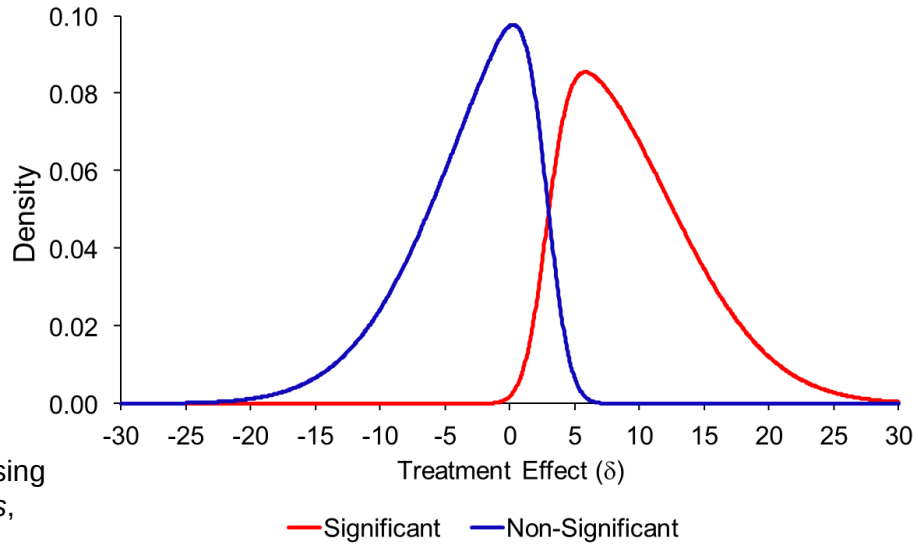
- Standard deviation of outcome σ=3
- Mean of prior for effect δ=1.5
- Standard deviation for prior v=0.85
- Minimum relevant effect = 2

Prior:





- The PoS seems to have limited use as a tool to select between study designs*
- "We expect most PoC studies to be negative in early development. Consequentially, as the study design improves, the PoS typically will tend towards a low value and not simply increase" – use Posterior Conditional Success and Failure Distributions?



*Walley et al. (2015). Advantages of a wholly Bayesian approach to assessing efficacy in early drug development: A case study. *Pharmaceutical Statistics*, *14*(3), 205–215.

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%
Significant results on both endpoints Estimates in Success or Conditional Success scenario	57%	58%	53%

Prior PoS

Prior probabilities for each outcome		
22%	40%	38%