Going beyond Probability of Success for Early Development studies

Trevor Smart 10 June 2025

Inspired by **patients.**
Driven by **science.**

Aim

- To get an understanding of why we need more than Probability of Success (PoS) for Early Development

- Basic stats taken a step further to show how
  - operating characteristics,
  - prior elicitation,
  - design priors and
  - pre-posteriors

  can be used to understand how much uncertainty could be removed by a Proof of Concept (PoC) study
- Simple graphics and pictures are shown that could be used to help explain implications of PoC design to the team
  - Lending itself to interactive discussions, homing in on more optimised study designs

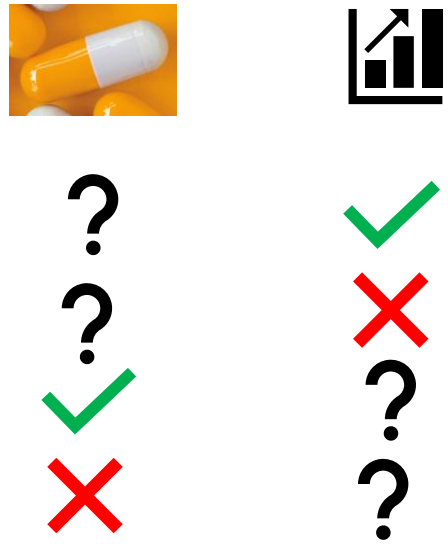
- The numbers from the example presented have been changed and the decision rules do not correspond to decision rules used.

Probability of Success (PoS) of a study

- Early studies such as PoC
  - **Making good decisions is more important**
  - No point in maximising PoS and progressing compounds that are not good enough

Why not appropriate for ED studies

- Other key questions – not just success.
- What is the probability that:
  - the compound is good enough, given the PoC is successful and we progress to the next study;
  - the compound is good enough, given we do not progress;
  - we progress, given the compound is good enough;
  - we progress, given the compound is not good enough.



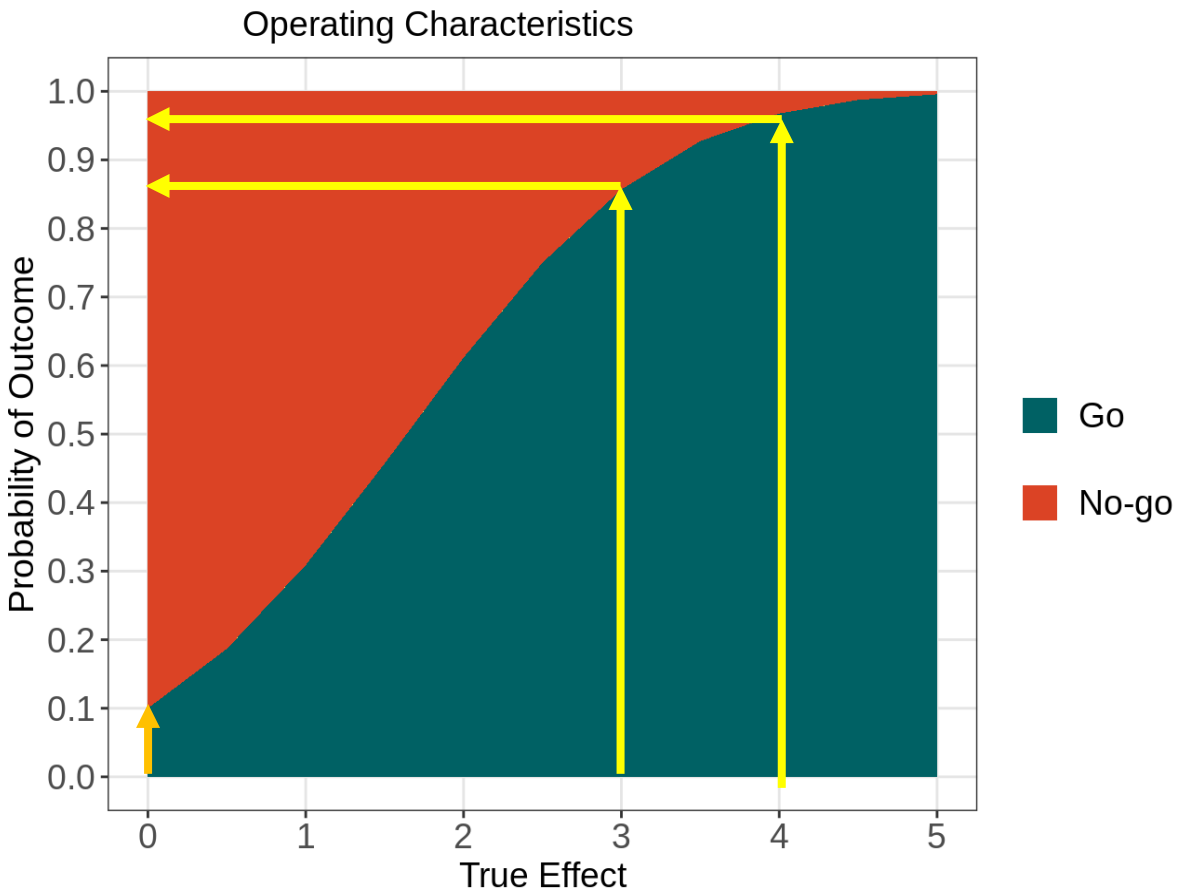
Probability of Success (PoS) of a study

- Terminology
  - Probability of Success – probability that the study meets the decision criteria
  - Operating Characteristic, *p(x)* – Probability that the study meets the decision criteria conditional on the true effect = *x*
  - Design Prior – The prior distribution we assume when simulating or calculating the PoS
  - Pre-posteriors – the expected posterior distribution if we predict what would happen in the study, given the design and our prior belief (design prior).
    - What we expect to get from the study without running it.
    - Pre-posteriors do not require a Bayesian analysis, but a Bayesian analysis could be done

- Simple relationships
  - PoS we integrate over the operating characteristic * design prior
  - Pre-posterior for success = operating characteristic * design prior / PoS
  - Pre-posterior for failure = (1-operating characteristic) * design prior / (1-PoS)

- Good Enough
  - Focus on what is good enough and not good enough (poor), not just the aspirational and placebo like values
  - What is good enough may change over the course of the development of the drug

Operating Characteristics



Probability of success given a true effect

Decision rule –

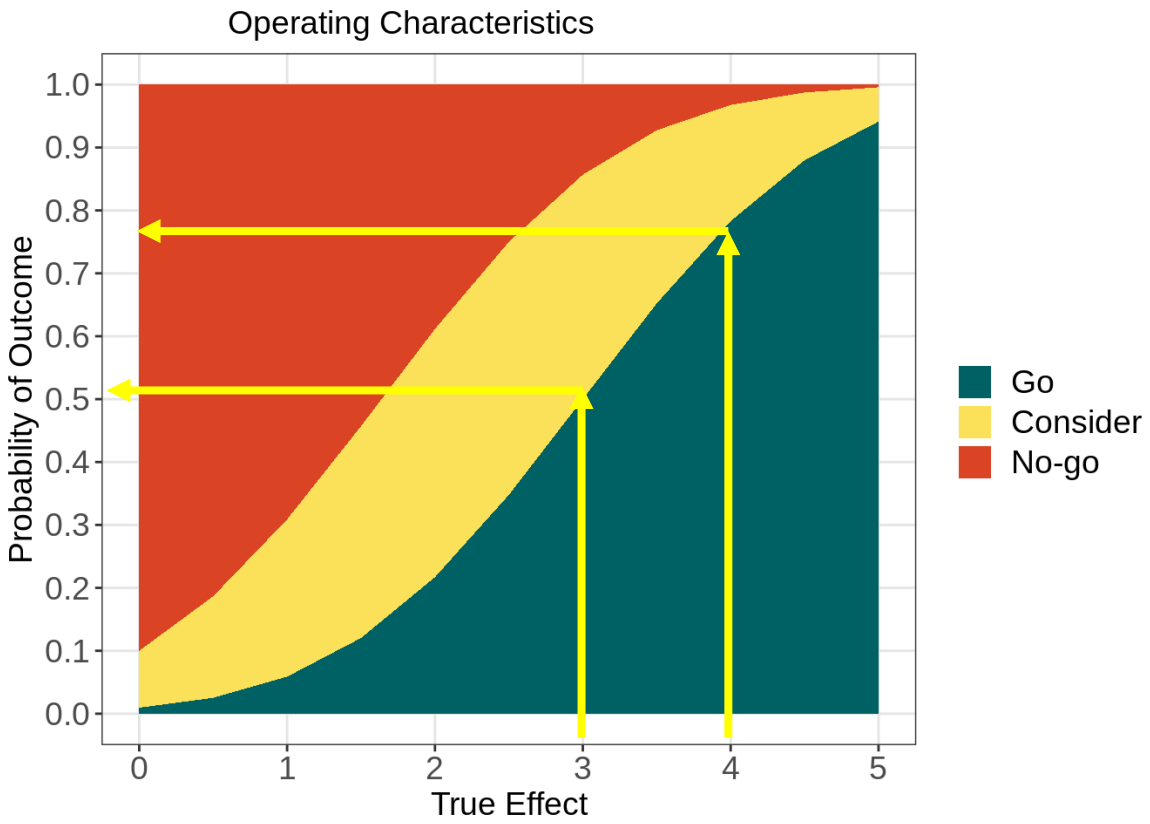
- 90% sure mean score on UCB > mean score on placebo

- If true effect = 0 prob of success = 10%
- If true effect = 3, prob of success = 85%
- If true effect = 4, prob of success = 96%

- 3 = Standard of Care (SoC)
- 4 = desired effect

Operating Characteristics

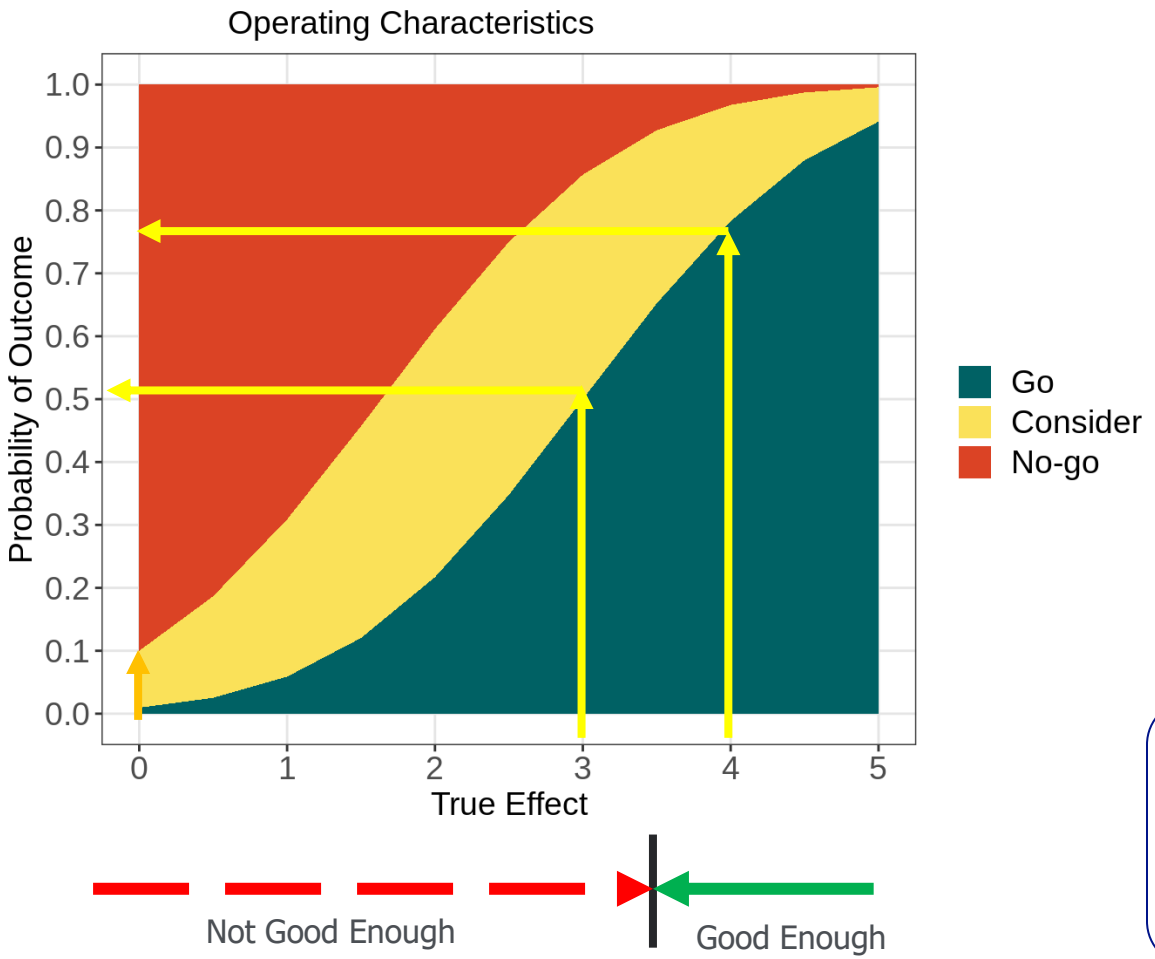
Dual Criteria



- Probability of success given a true effect
- Decision rule –
  - 90% sure mean score on UCB > mean score on placebo
  - 50% sure mean score (UCB – Placebo) > 3
- If comparing against SoC (3pt difference)
  - If true effect = 3, prob of success = 50%
  - If true effect = 4, prob of success = 78%

- Almost 80% power for a difference of 4 and 50% chance of going forward with a compound equivalent to SoC
- Ultimately, we need to be better than SoC

Common team misunderstandings



- Pr(placebo like, given study success) is not the same as Pr(study success, given placebo like) [type 1 error]
- Pr(true effect is at least *d* given study success) is not the same as Pr(study success given true effect = *d*) [power]

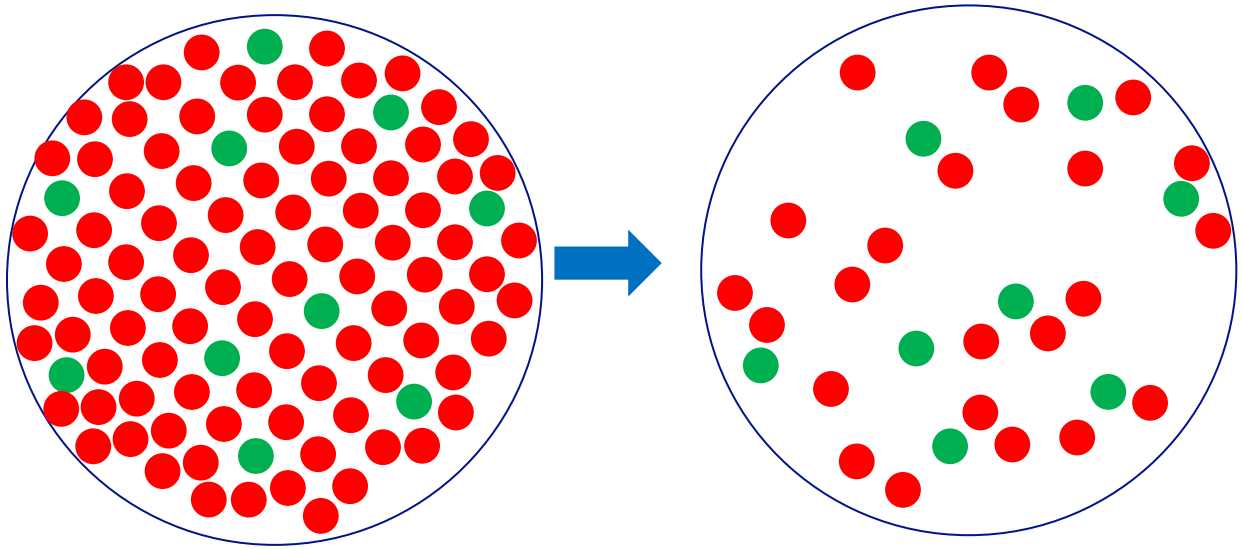
- Pr(Wrongly progressing) is not the same as type 1 error
- Pr(Correctly progressing) is not the same as
  - power (conditional on a single value) or
  - PoS (across all possible values, both good and poor)

To calculate these important probabilities, we need some understanding of our prior belief in the compound

- How do we estimate a prior with no data?
- Prior elicitation

How likely are we to go forward with a good enough compound?

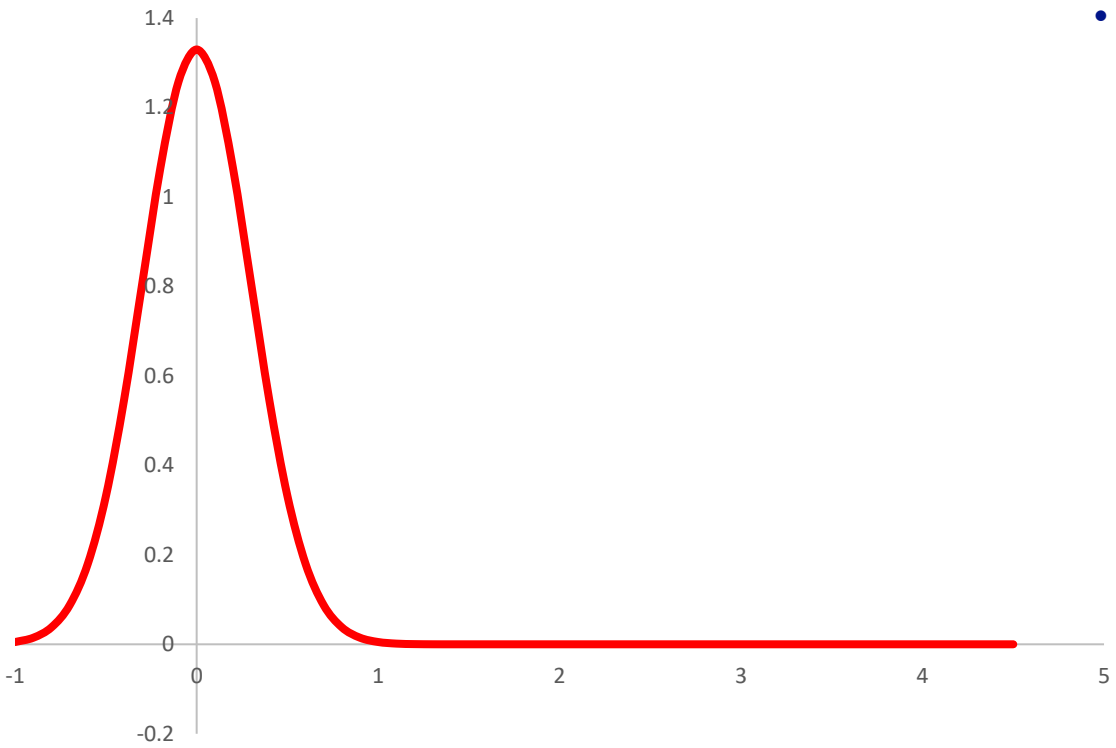
- Depends on Operating Characteristics **AND** how likely are compound is to be good enough
- Eg
  - If before we start, we have a very high chance it is no better than SoC,
  - Then, even if we pass the criteria, it is still likely to be no better than SoC



- 80% chance of a great compound passing
- 25% chance of SoC passing
- After study still greater chance of being SoC than a great compound

- We can use our prior belief in the compound to help optimise the sample size and decision rule

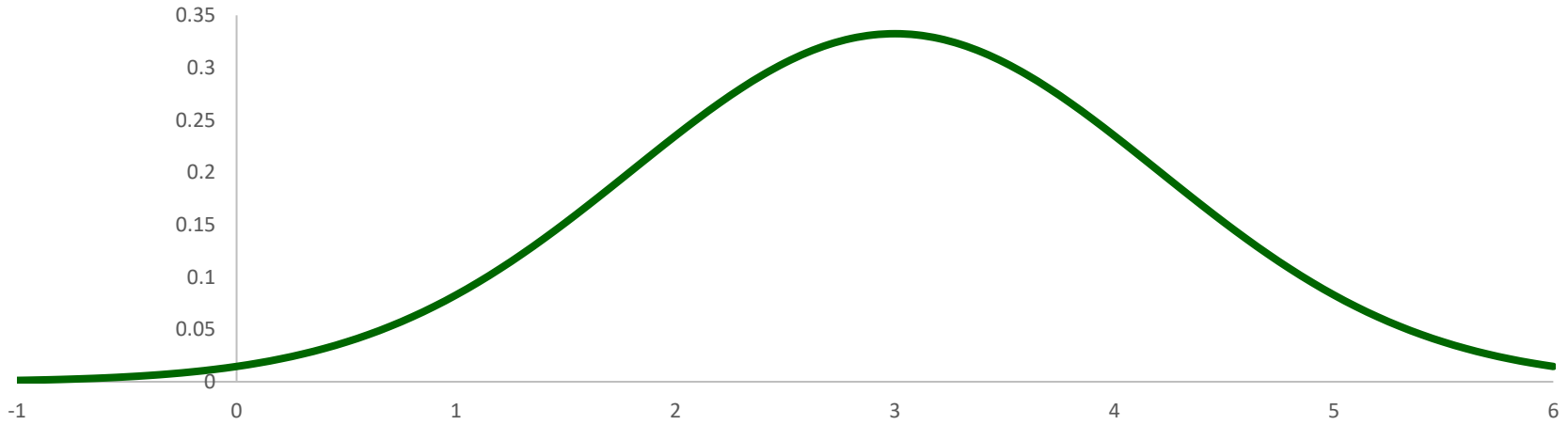
Placebo like



- Placebo Like
- For simplicity take as Normal distribution centered around 0

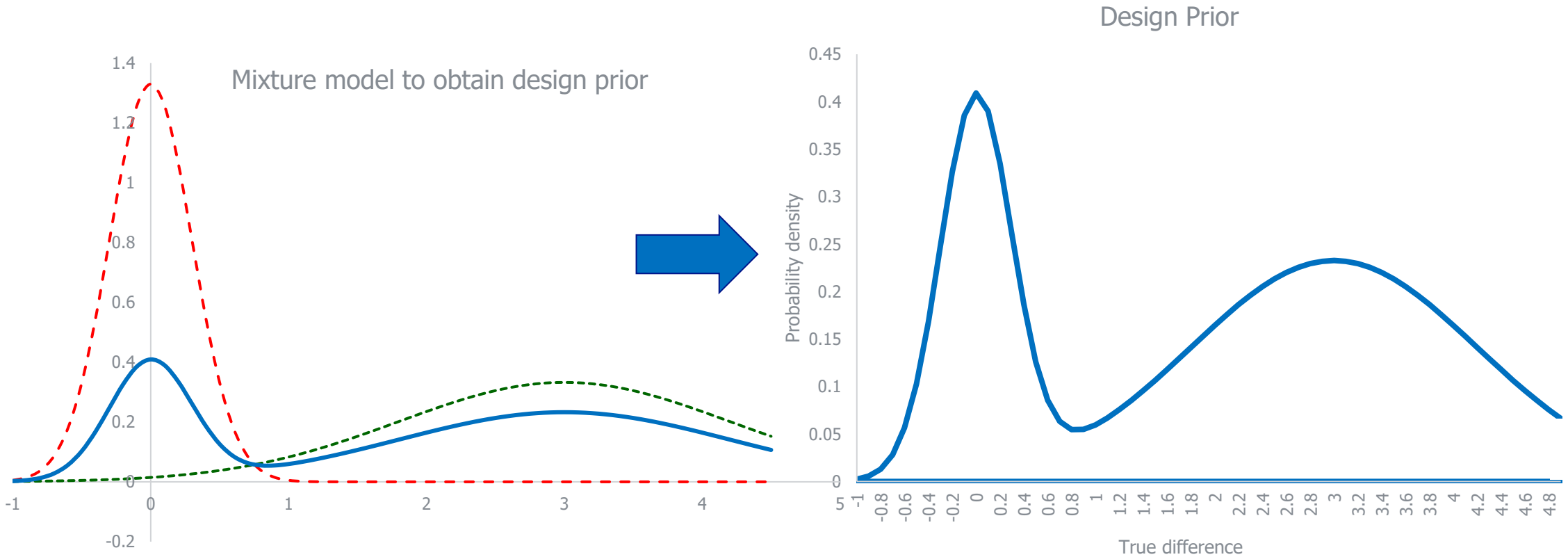
Active, but we don’t know how well

- Works, but we don’t know how well
- Normal distribution centered around 3, but very wide



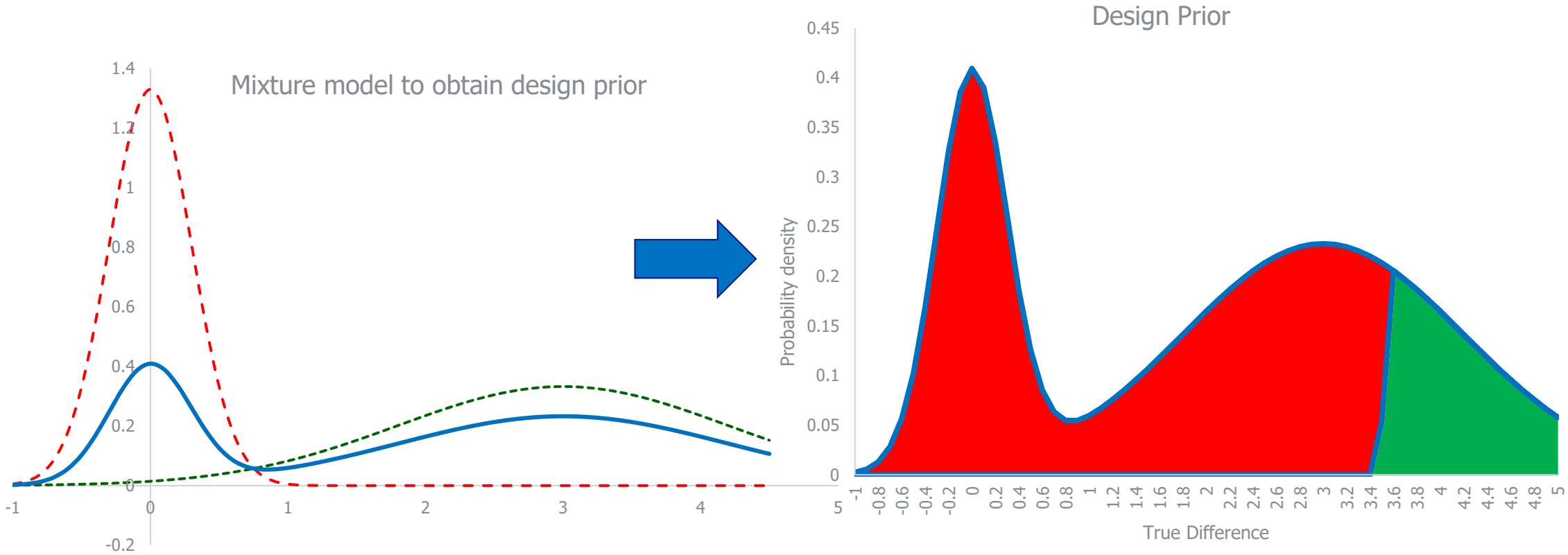
Design Prior

70% chance active + 30% chance placebo like



Design Prior

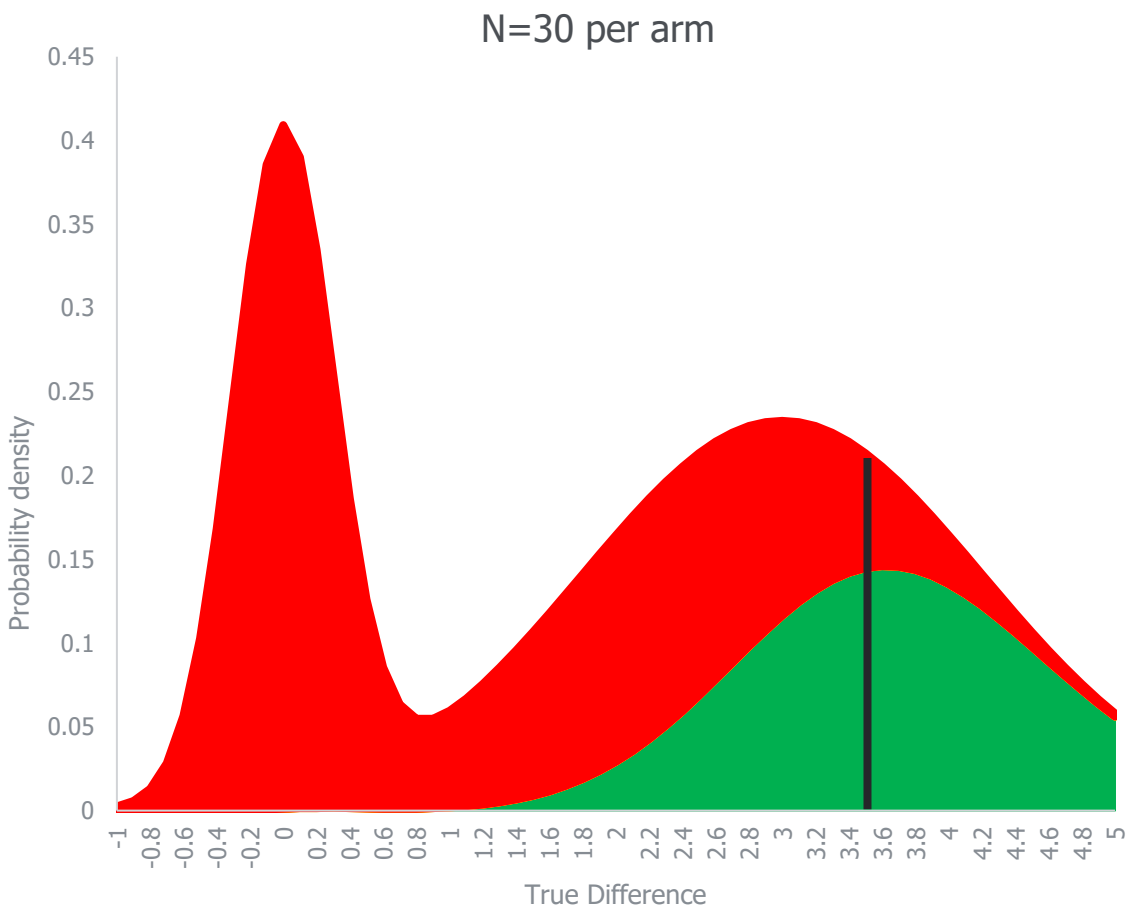
70% chance active + 30% chance placebo like

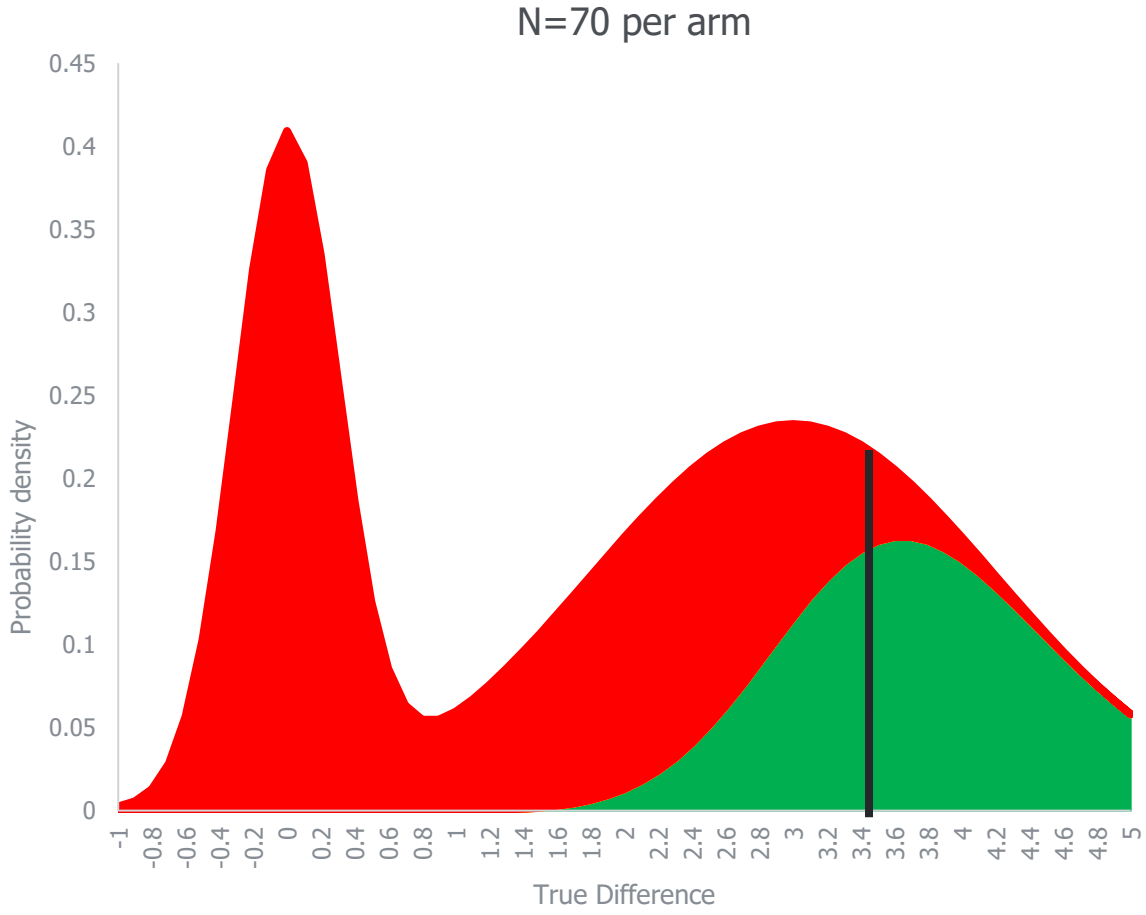


Good Enough > 3.5

Decision Criteria: 50% sure mean score (UCB – Placebo) > 3

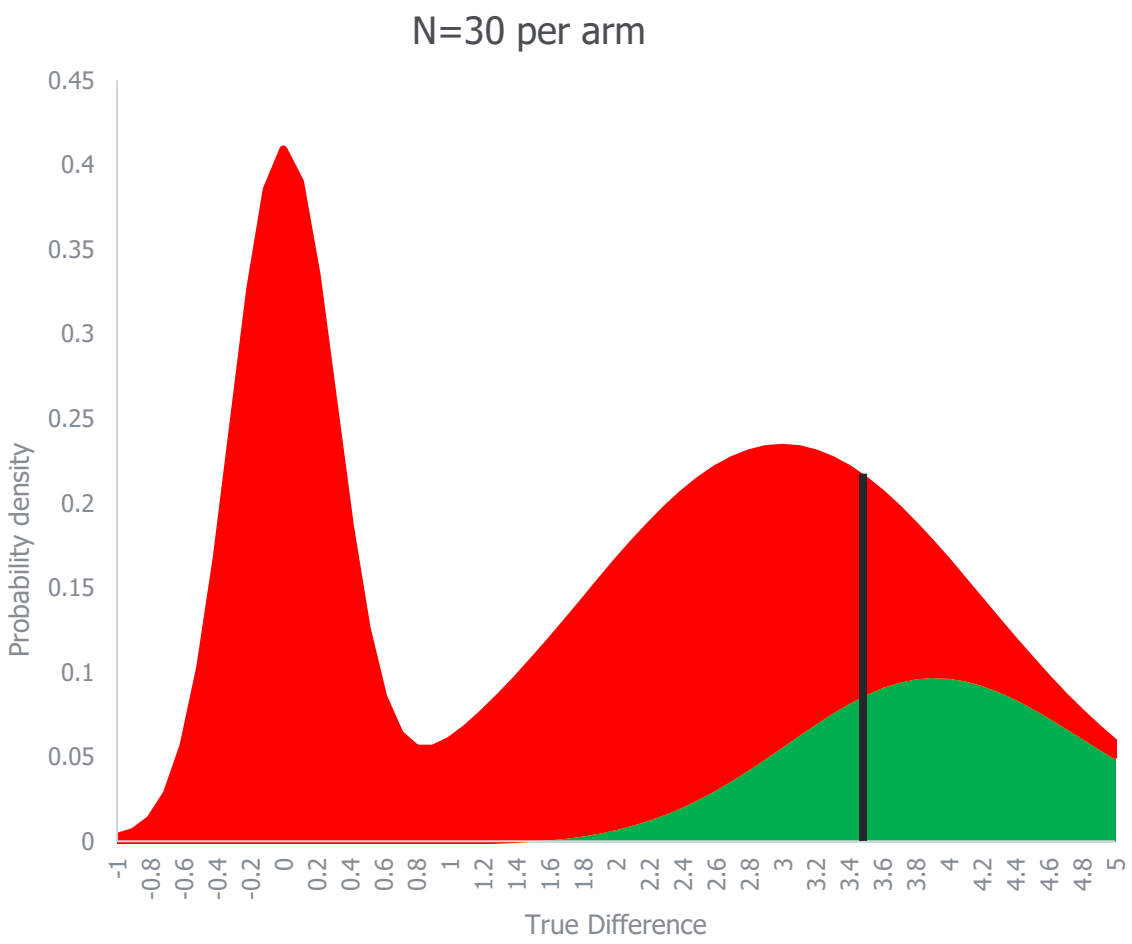
- High probability of success if the compound is good enough
- But also high probability if the study is successful, the compound is not good enough
- Some but limited improvement as sample size increases

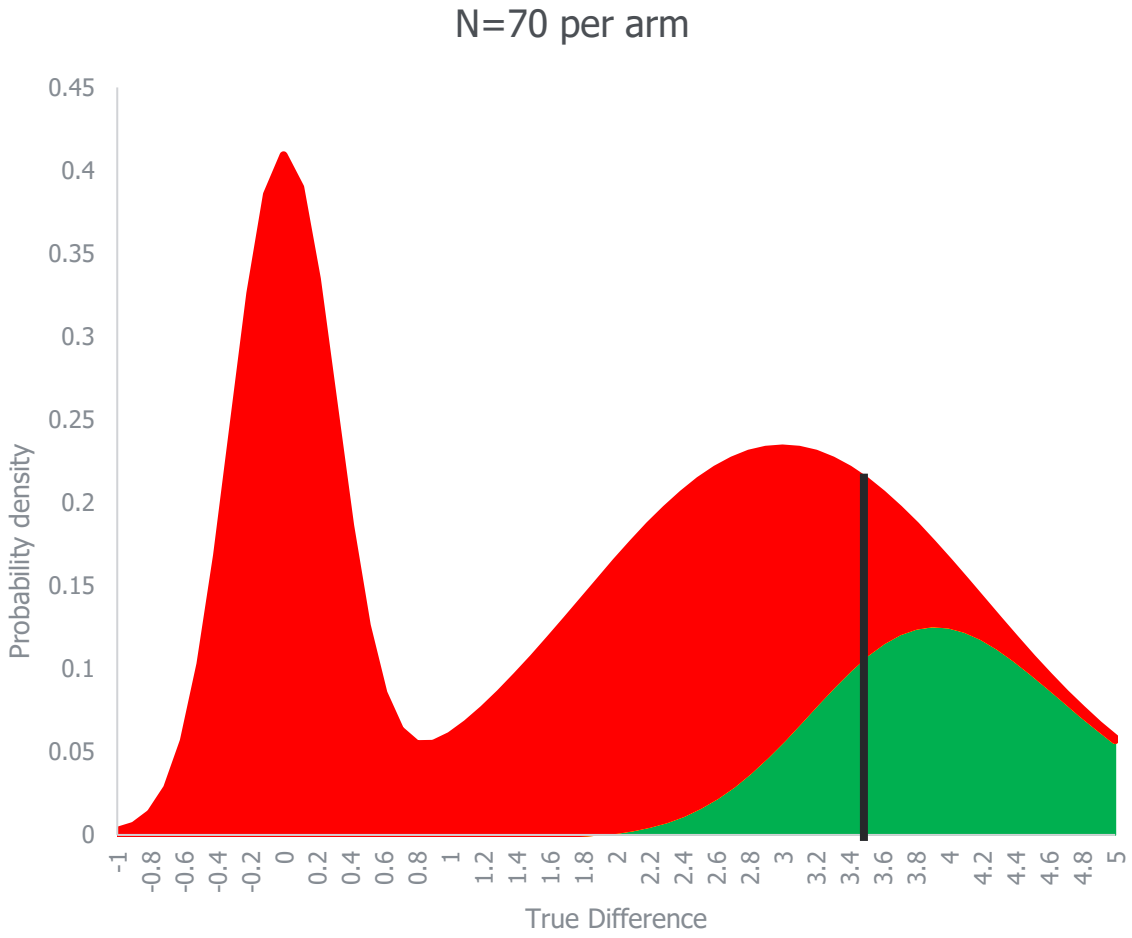




Decision Criteria: 75% sure mean score (UCB – Placebo) > 3

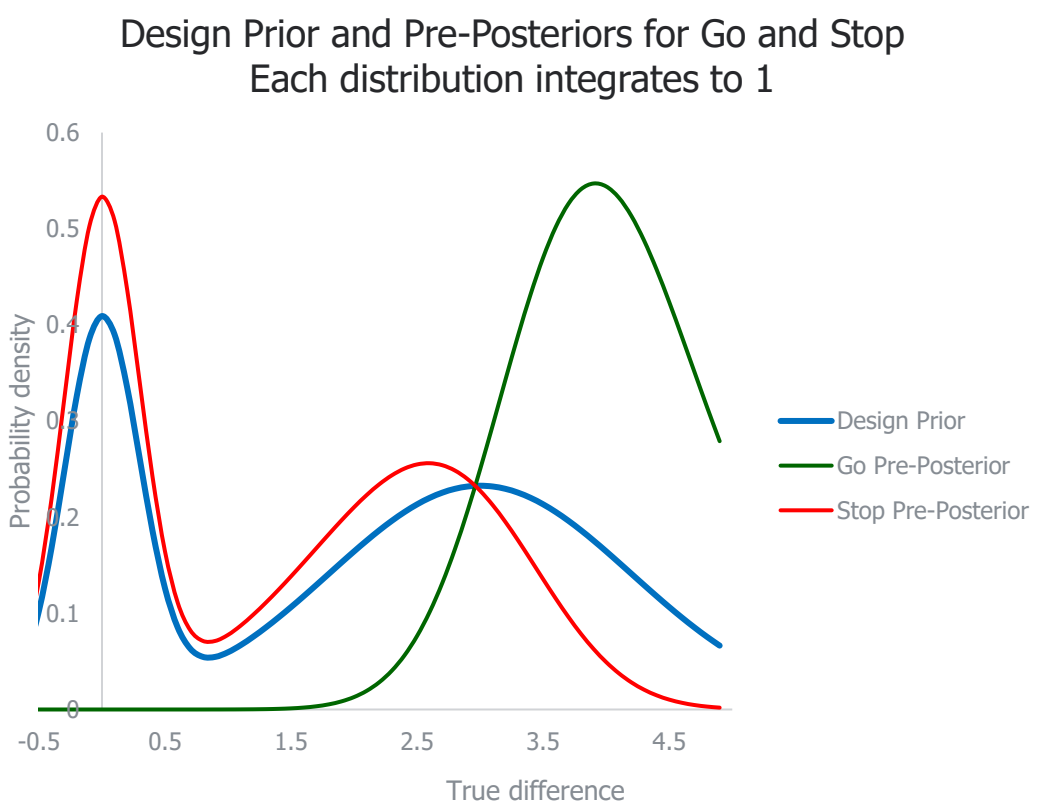
- Reduced probability if the study is successful, the compound is not good enough
- But also reduced probability of success if the compound is good enough



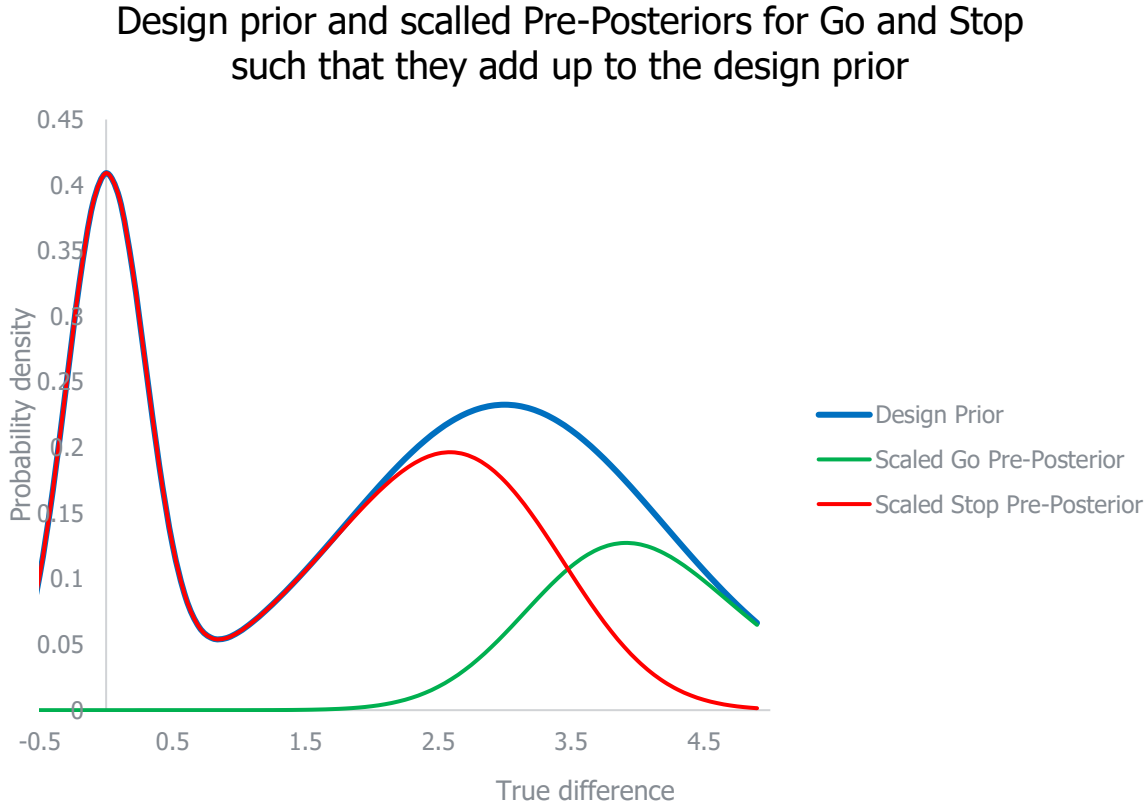


Other displays could be used

Decision Criteria: 75% sure mean score (UCB – Placebo) > 3, N=70 per arm



If we are successful in the study, is this increase in confidence what is expected?



Is the split between stopping and going acceptable across the range of possible true differences?

Probabilities can be calculated

- The diagrams give a good picture to interactively compare different designs
  - My preference
- The key values can still be easily calculated.
  - E.g.

Decision Criteria	50% sure better than 3			75% sure better than 3		
Number per arm	30	70	100	30	70	100
Pr(Success)	33%	33%	33%	21%	23%	24%
Pr(Success and Good)	18%	20%	21%	14%	17%	18%
Pr(Good given Success)	56%	62%	63%	67%	73%	75%
Pr(Success given Good)	82%	91%	93%	62%	76%	82%
Pr(Success and Poor*)	15%	13%	12%	7%	6%	6%
Pr(Poor given Success)	44%	38%	37%	33%	27%	25%
Pr(Success given Poor)	19%	16%	16%	9%	8%	8%

- Despite sample size changes, PoS remains the same, but other key probabilities change
- For 50% decision rule Pr(Good|Success) is low [and Pr(Poor|Success) is high], but Pr(Success|Good) is high
- Properties of 75% decision rule look better, but potentially larger sample size

* Poor = not good enough

POCs are always a compromise
No one answer fits all situations

How does it fit into the clinical plan and the overall portfolio?

- Is there another study prior to any pivotal study?
  - More important to keep good compounds, than stop poor compounds
- Is this the last study before pivotal studies
  - Need to increase the probability that if we go forward the compound is good enough
- If you have a large portfolio
  - More important that if you go forward the compound is good
- If you have a small portfolio
  - More important that if you have a good compound, you don't kill it
- If huge unmet need or potential blockbuster indication
  - More important that if you have a good compound, you don't kill it

- Etc

Summary

We need more than Probability of Success to optimise our Early Development studies

The use of operating characteristics, design priors and pre-posteriors are key

Prior to Proof of Concept, we do know some things that will help create informative design priors