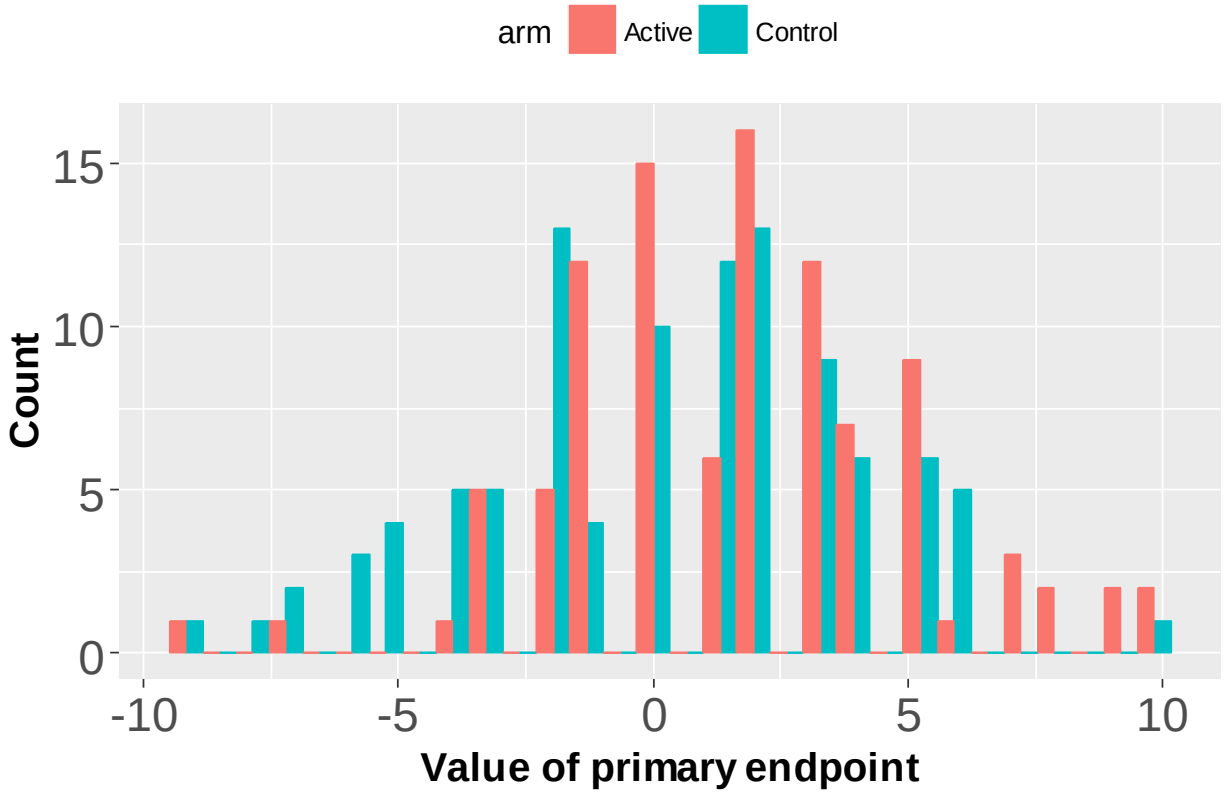
Reducing the CDLC through a Smarter Decision-making Framework Driven by Statistical & Probabilistic Measures

Sakthivel Sivam
GSK Asia Private Ltd.

PhUSE SDE Chennai
12Oct2019

Protocol: "Sample size of n=200 chosen to have 90% power"
That is: a 90% chance to reject the hypothesis of "no difference between treatment groups" at the 5% level of significance, if the true difference is actually 1.6 units (the value specified in the MP).



Observed difference = 1.46, with 95% Confidence Interval of (0.46, 2.43).

P< 0.05 => reject the null hypothesis

Protocol-specified success (a significant p-value) was achieved, but what does it mean?

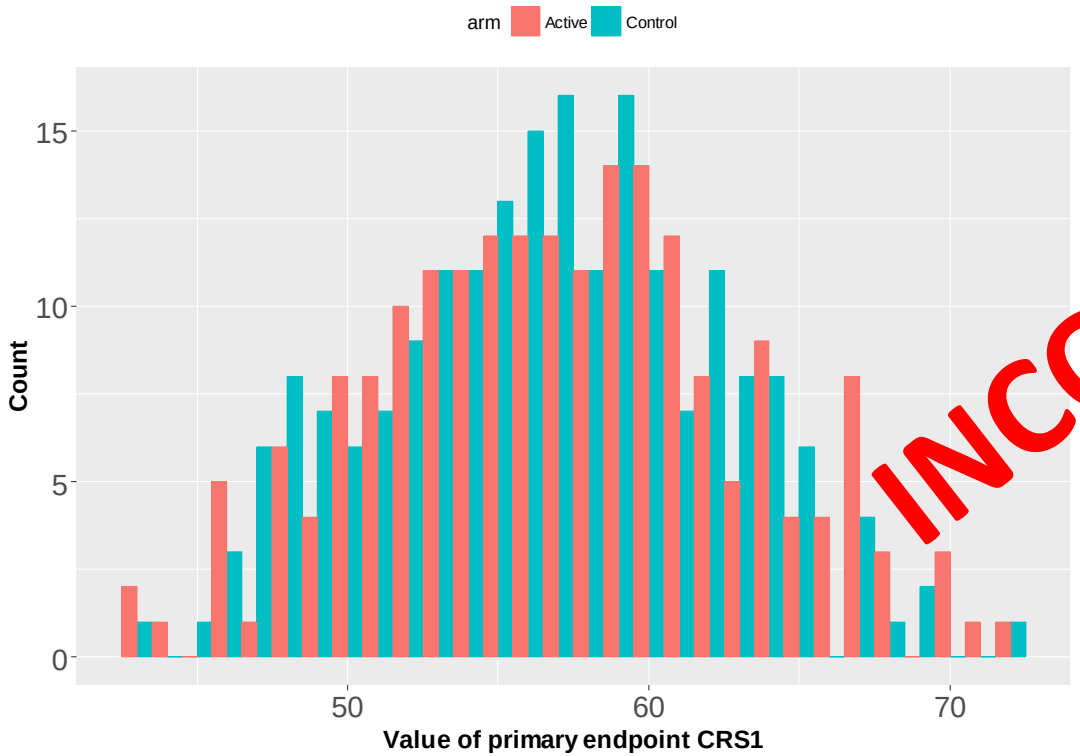
We've rejected a theory of "no difference". But what should we *do*? What should we believe about drug effect?

What is the probability that the true effect lies in the confidence interval?

What is the probability that the true drug effect is larger than 1.6, based on these data?

The problem: what do you tell R&D to do?

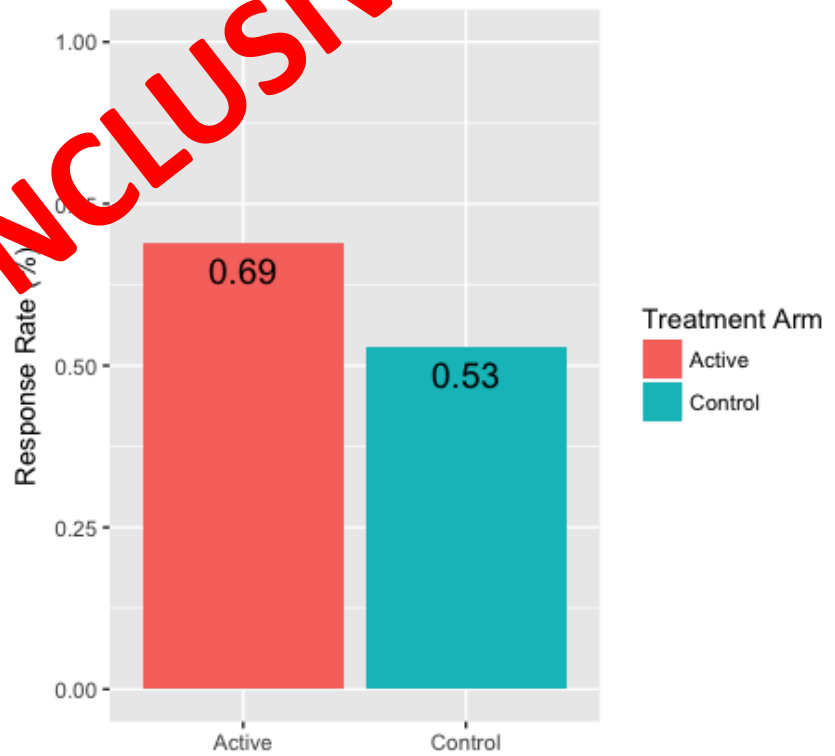
"Non-significant" Primary Endpoint



- Observed mean CRS1 values:
  - Active: 57.26
  - Control: 56.54
- 95% CI for true difference in CRS1 between Active & Control: (-0.41, 1.86)
- No evidence at all for real difference!!!

"Highly significant" secondary endpoint

INCONCLUSIVE



- Observed difference in response rates = 16% points in favour of Active
- Hypothesis test of "no difference" rejected at the 5% level: p<0.05 (Note: there is a drug difference)
- 95% CI for the true difference in response rates is (11.5, 19.5)%

Go? Or Stop? Invest? Or Don't Invest?

What happened to p-values?

- Nothing. But $p<0.05$ isn't a great tool for decision-making
- "Strong evidence to reject $\Delta = 0$ at the 5% level of significance" doesn't tell us how probable it is that Δ is greater than a target value

- Our strong proposition is that (posterior) probabilities are better tools for weighing evidence about endpoints, and to act as inputs to the decision-making process

~50% of Phase III trials fail

Failure Triggers

DRIVERS OF FAILURE	EXAMPLES
Inadequate basic science	• Beneficial effects in animal models not reproduced in humans • Poor understanding of target disease biology
Flawed study design	• Patient population definition changed from Phase II to Phase III • Phase II surrogate endpoint not confirmed by Phase III clinical outcomes • Insufficient sample size
Suboptimal dose selection	• Inadequate dose finding in Phase II • Poor therapeutic indices
Flawed data collection and analysis	• Phase II "false positive" effects were not replicated in Phase III • Overoptimistic assumptions on variability and treatment difference • Missing data; attrition bias; rater bias • Wrong statistical tests; other statistical issues
Problems with study operations	• Data integrity issues; GCP violations • Recruitment, dropouts, noncompliance with protocol • Missing data; unintentional unblinding
Other	• Insufficient landscape assessment of current standard of care and precedents

Source: European Center for Pharmaceutical Medicine; PAREXEL Analysis

Table 2. Specific events in key areas of clinical trial conduct that can result in Phase III failure.

Phase III Trial Failures: Costly, But Preventable, Aug 01, 2016, A Grignolo, S Pretorius, Applied Clinical Trials, Volume 25, Issue 8

Statistical Power Vs Assurance (PoS)

- **Power**: Probability of success assuming that the true treatment difference is exactly θ.
- **Assurance**: Probability of success accounting for uncertainty around the true treatment difference.
- **A protocol might say something like this ...**

Assuming a clinically relevant difference of 2 points on the primary endpoint scale, with a standard deviation of 6.2, 200 subjects per arm are required to provide 90% power at the 5% alpha level (two-sided).

True effect size	Power	Expert belief on the effect size	Power x Belief
0	2.5%	20%	0.5%
1	36%	30%	10.8%
2	90%	40%	36%
3	99.8%	10%	9.9%

54%

Power = 90%

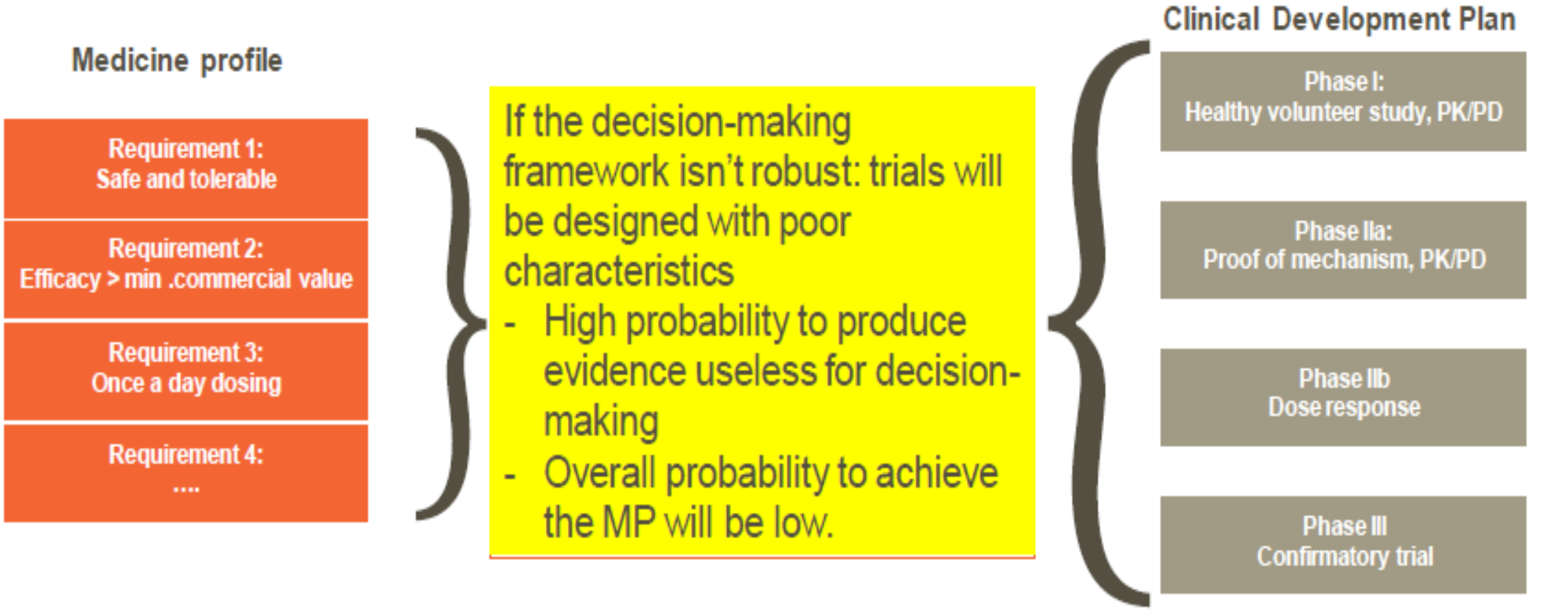
BUT

Assurance = 54%

The probability of success assuming the true (unknown and never known) effect of the drug is 2 points

The average of the power calculations, weighted by the belief about how big the true effect size is

We need a decision-making framework



- What evidence is required from Phase 2b to meet Requirement 2?
- Once phase 2b reads out, what is the probability that Requirement 2 will be met in Phase 3?

Decision Making Framework

- We think a lot about how design choices affect immediate outcomes
  - Will the trial deliver a significant p-value? Etc.

- We should incorporate an investigation of the "decision-making framework" into that design process also.
  - How will your evidence (p-values or whatever it may be) map to decisions about the development plan?
  - How does that *mapping* perform under different conditions?

Why Quantitative Decision Making? – Paradigm Shift !

- "You can't make decisions in the face of uncertainty without obeying the probability calculus..”
- Think about Pension ! How would you like its funds to be invested? Within a decision-making framework that takes account of risk..
- A framework for planning decisions, and measuring evidence, and calibrating risk: a framework that can and should be applied to any study in Development, especially those big Phase 2 "pivot" decision studies.

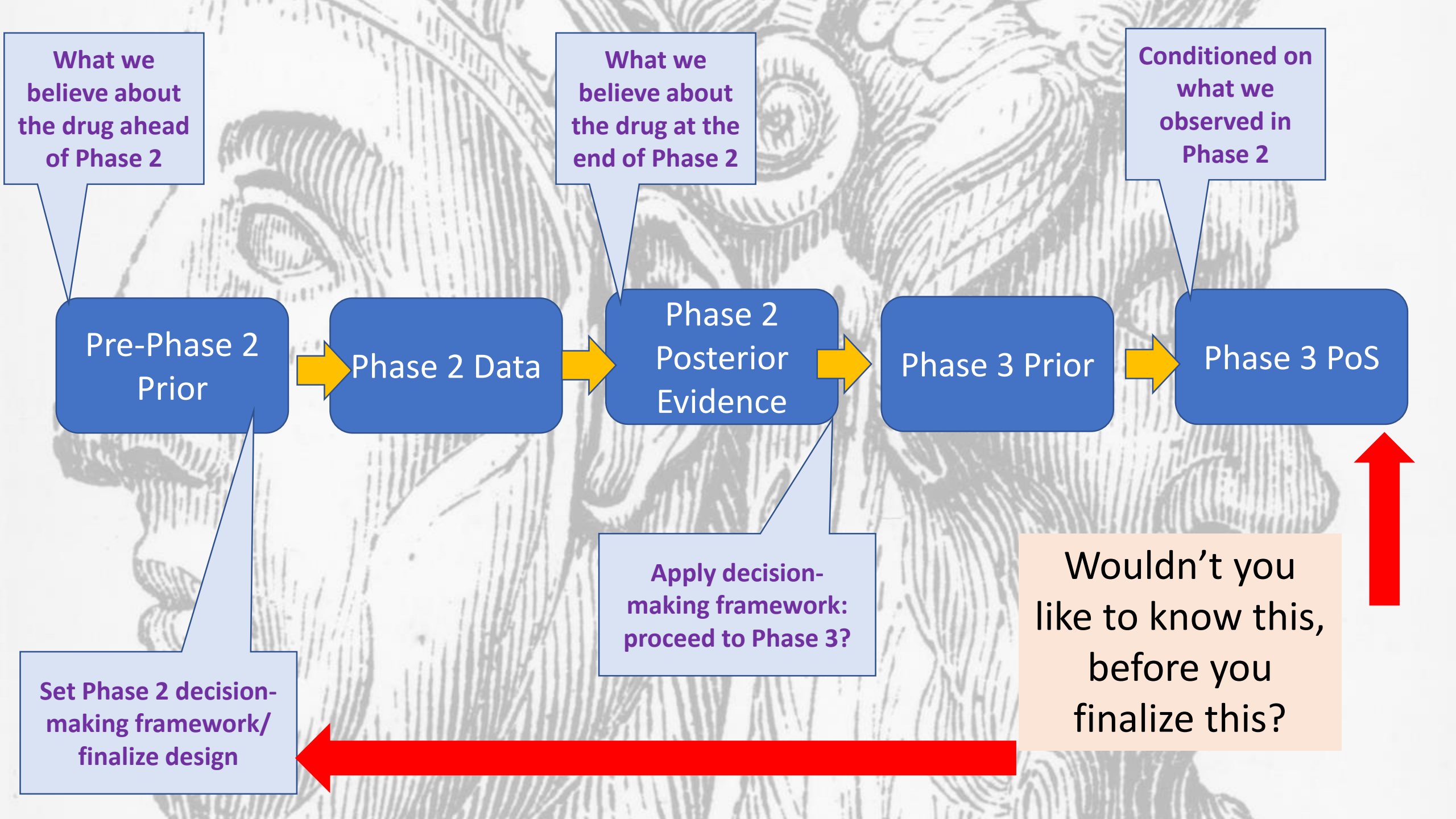
  - does the team understand exactly which numerical outcomes for each endpoint will lead to positive or negative conclusions about those endpoints?
  - If there are multiple endpoints in play, do you know exactly which combinations of readouts from different endpoints map to STOP or GO decisions about future investment?
  - Has the link between conclusions about the endpoints, and the Medicine Profile targets for the medicine, been made, explicitly and quantitatively?
  - Has the proposed design been interrogated with regard to current knowledge/belief about the asset in question?
  - What's the probability of an inconclusive result?

Quantitative Decision Making (QDM) means

- "to describe in full the quantitative characteristics of the proposed study design, to transparently show the probability of being able to make a decision at the end of the trial and showing how the proposed study de-risks the next stage of development"

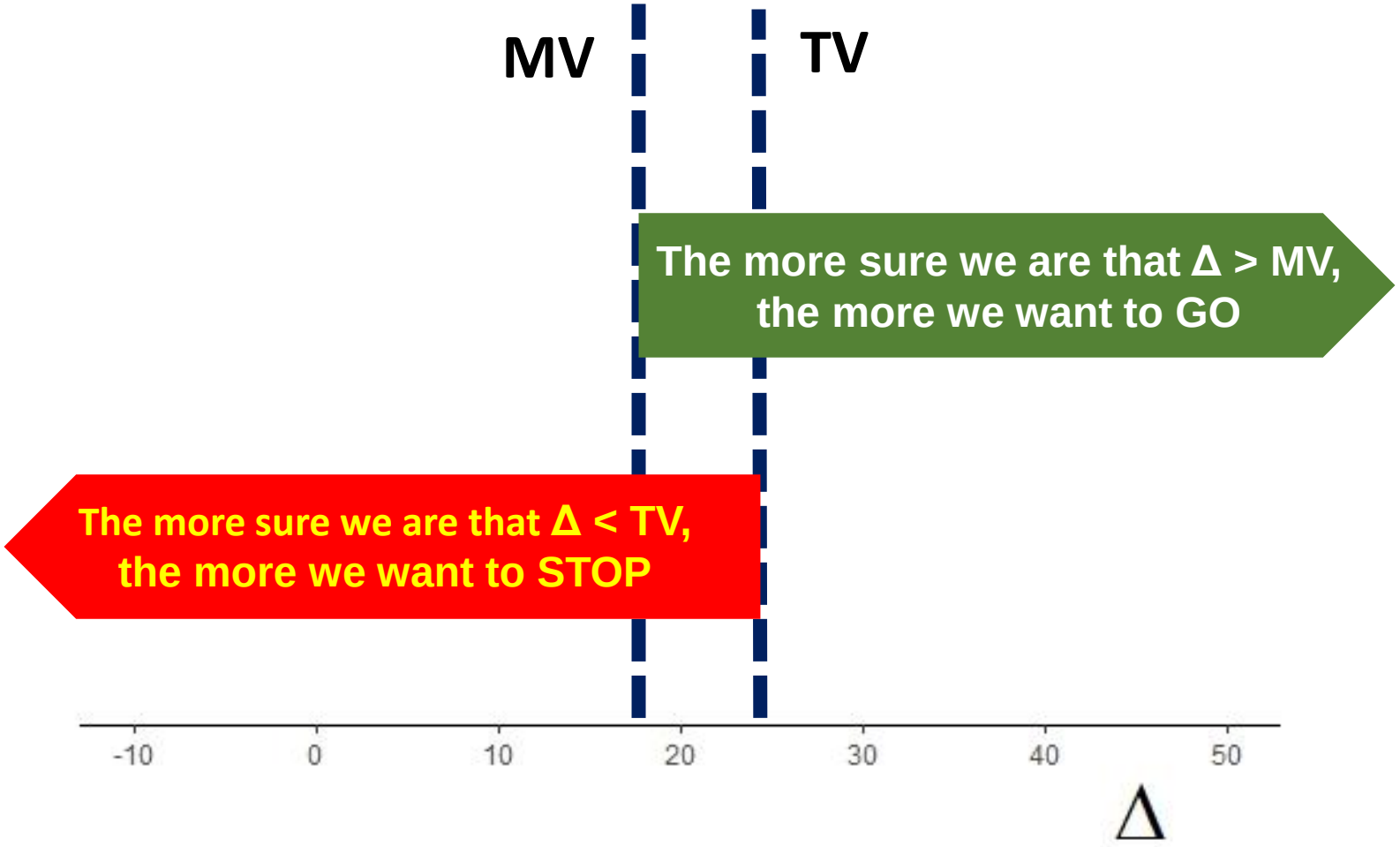
- Two questions that we should be considering at the design stage of all studies we run:
  1. Are we providing all the quantitative assessments that help us to understand the risks for this study and for the future development of the compound?
  2. Can a clear decision be made off of a particular study more often than not?

- The key driver in any study/development plan should be clarity over *which* decisions need to be made, and *how* to generate evidence from studies to trigger, confidently, such go/no-go decisions.
  - **Only studies and plans with a high probability of delivering such decision-making evidence should be endorsed by a data-savvy, patient-focused organisation.**


What we believe about the drug ahead of Phase 2
What we believe about the drug at the end of Phase 2
Conditioned on what we observed in Phase 2
Pre-Phase 2 Prior
Phase 2 Data
Phase 2 Posterior Evidence
Phase 3 Prior
Phase 3 PoS
Apply decision-making framework: proceed to Phase 3?
Wouldn't you like to know this, before you finalize this?
Set Phase 2 decision-making framework/ finalize design

How can these rules be defined?

Based on the Minimum Value (MV) and Target Value (TV)



MV & TV Determined:

- ✓Obtained from the Commercial outlook (i.e. payers perspective)
- ✓Multiple sources of information considered (internal / external)
- ✓Significant input from clinical and others..

The Decision-Making Framework Applied

You define the evidence thresholds (80%, 90%) & they don't need to be the same for each threshold, and nor do they need to be 80 or 90%

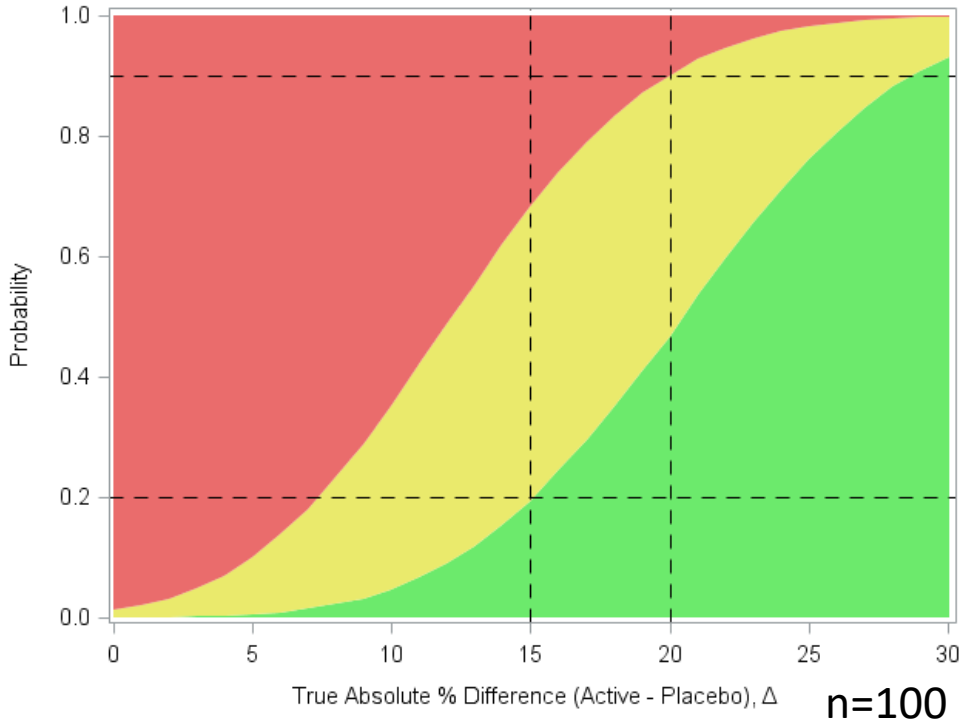
$\Pr(\Delta > MV \mid data) \geq 80\% \rightarrow$ STRONG **POSITIVE** EVIDENCE
$\Pr(\Delta < TV \mid data) \geq 90\% \rightarrow$ STRONG **NEGATIVE** EVIDENCE

		Positive Evidence	
Negative Evidence		$\Pr(\Delta > MV \mid data) < \varepsilon$	$\Pr(\Delta > MV \mid data) \geq \varepsilon$
	$\Pr(\Delta < TV \mid data) < \eta$	CONSIDER	GO
	$\Pr(\Delta < TV \mid data) \geq \eta$	STOP	CONSIDER

Evidence Thresholds

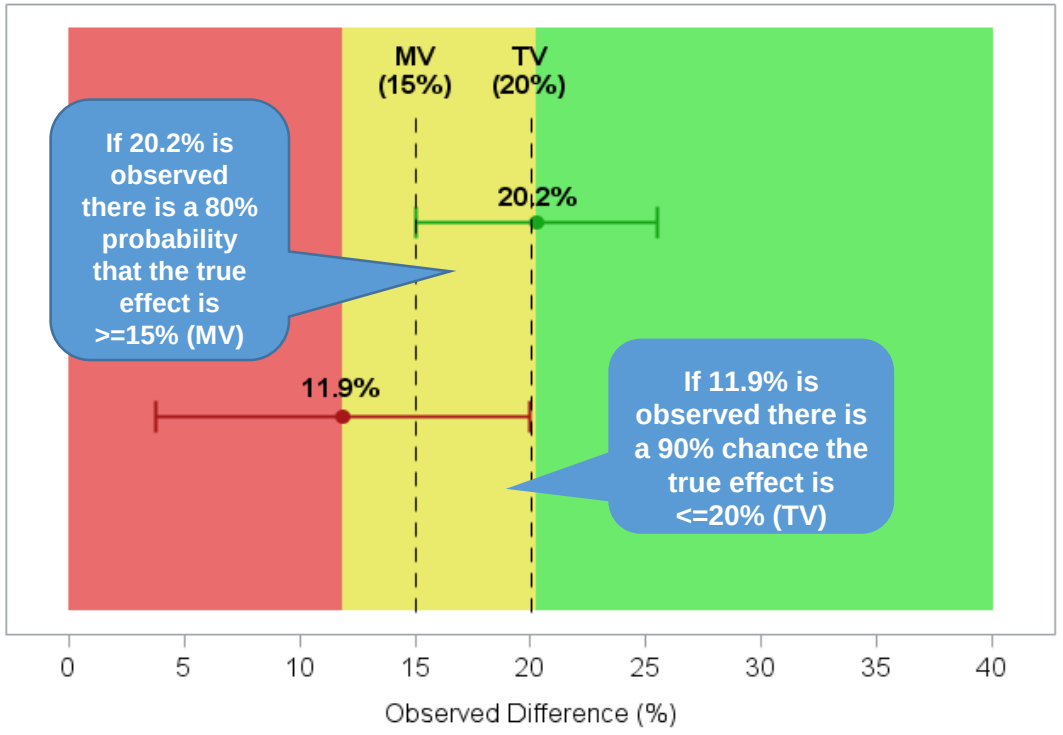
- In addition to the MV & TV, the desired level of evidence for each rule is required
  - **POSITIVE Evidence**: High probability that the **true drug effect is > MV**
  - **NEGATIVE Evidence**: High probability that the **true drug effect < TV**

Pre-study: OC of the rule (risks)



n=100

Minimum critical values and decisions



Instead of a single power curve, there are two, which together make a decision-making framework to guide whether the drug should:

Continue in Development (green area); STOP Development (red area); OR Consider (yellow area)

Multiple endpoints

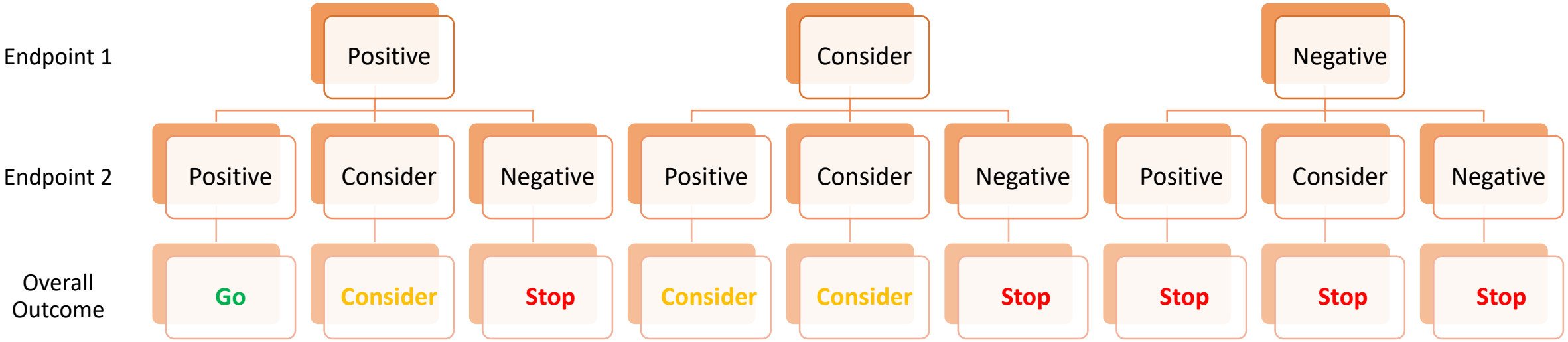
- The number of decision-making endpoints is nearly always >1

- QDM framework extends easily to this situation
  - Define MV and TV for each endpoint
  - Define level of evidence required to declare positive/negative/consider for each endpoint
  - Map every possible combination of evidence outcomes to an overall STOP, GO, CONSIDER decision

- The OC of such a mapping can be investigated as before.
- Failure to map outcomes to decisions leads to an explosion in Type I error as the no. of endpoints increases.

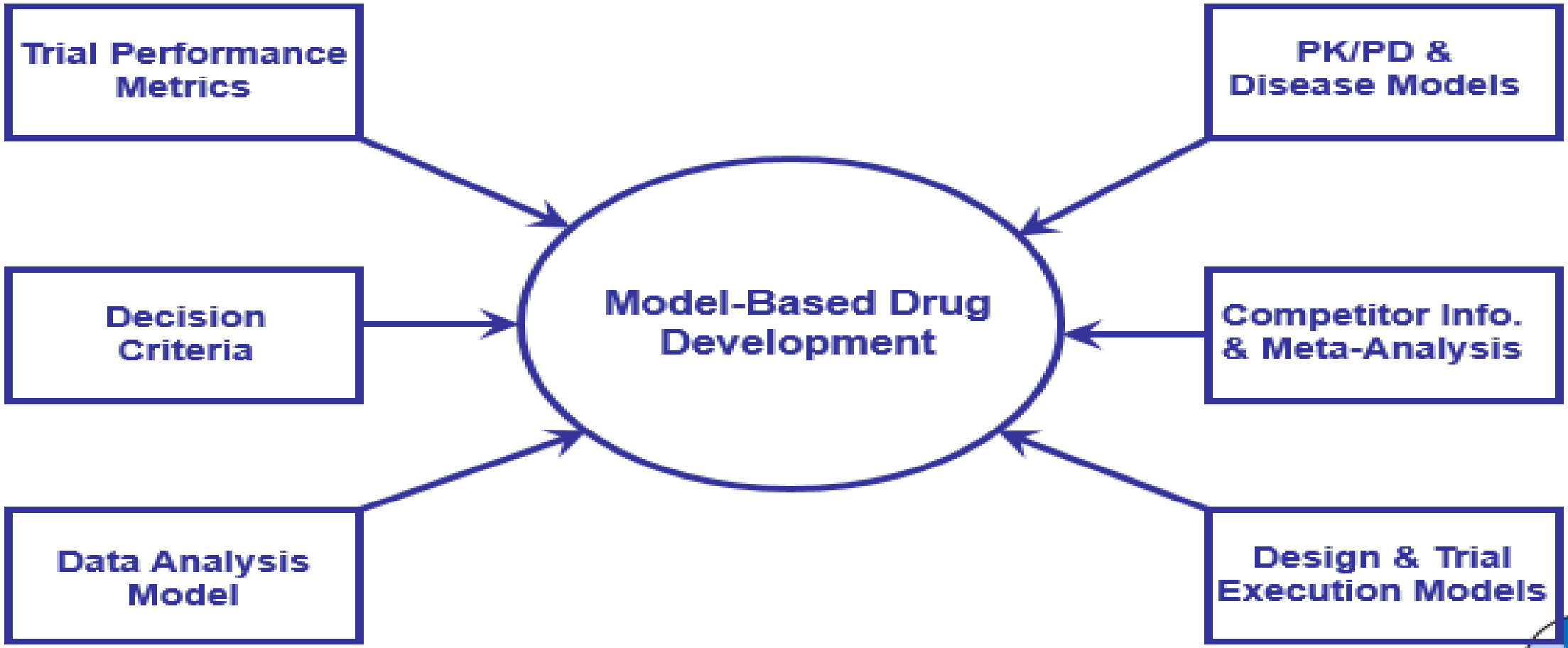
Decision Flow

Designing Framework Becomes More Complex…



- Once in mapping in place, its OC must be determined
- What is the overall probability of achieving, GO, STOP, Consider
- Exactly the same process with single endpoint …. But requires more thought!

Key elements of Model-Based drug development

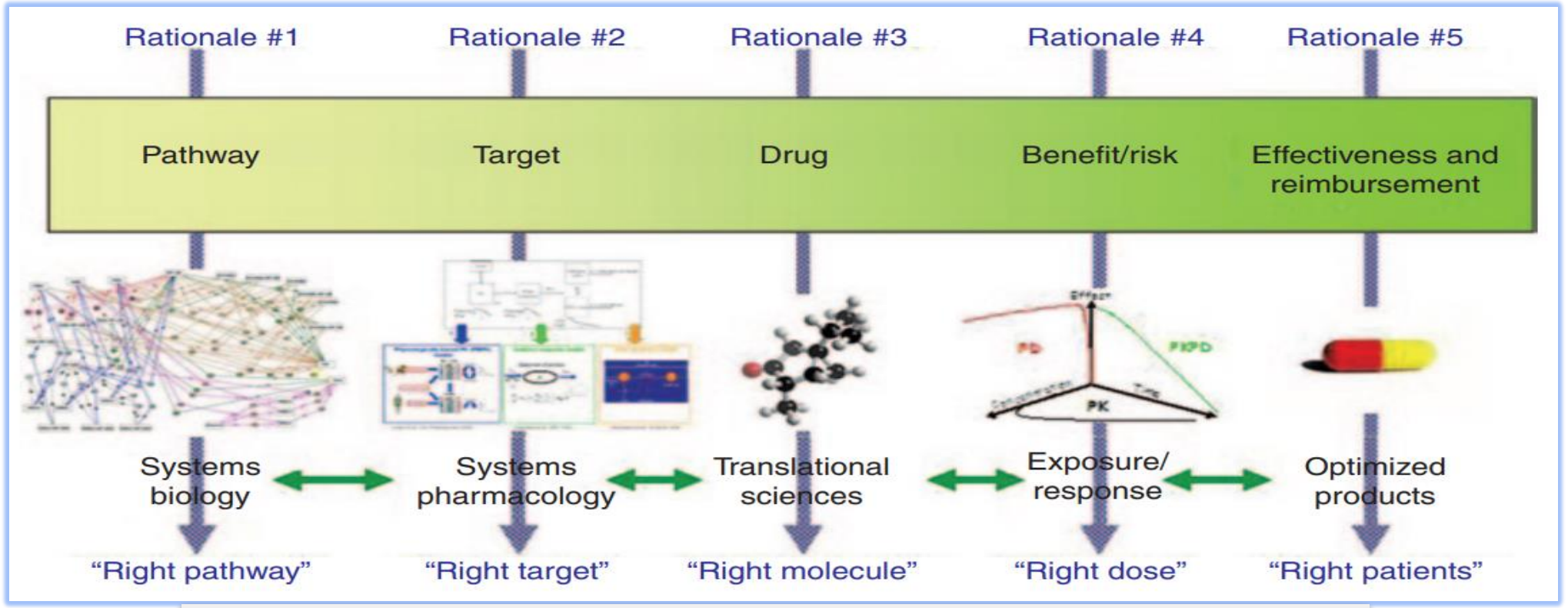


2

* Lalonde et al, Clin Pharm & Ther, 2007; **82**: pp21-32

Model-Based Drug Discovery & Development

MBD3: constructing the quantitative framework



"A quantitative framework for **prediction** and **extrapolation** centered on **knowledge** and **inference** generated from integrated models of **compound**, **mechanism** and **disease level data** aimed at improving the quality, efficiency and cost effectiveness of decision making"

Why is this framework important for decision makers?

MBD3 components required for ROI

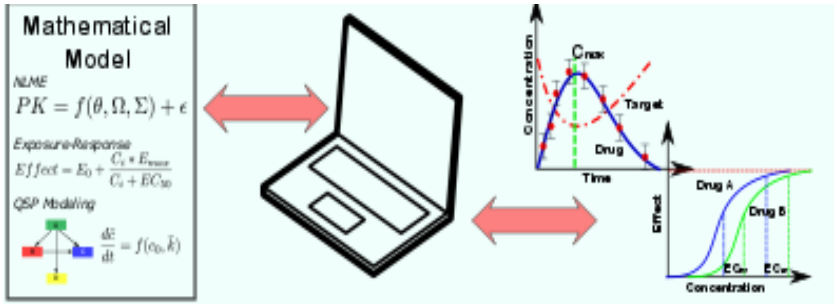
Impact and Return of Investment

- Increased confidence (PoS)
- Increased efficiency
- Cost savings
- Cost avoidance

Foundational Clinical Pharmacology Understanding of Asset

Integrated Team Effort in Quantitative Development Strategy

Pull From Decision Makers (Internal and Regulatory)

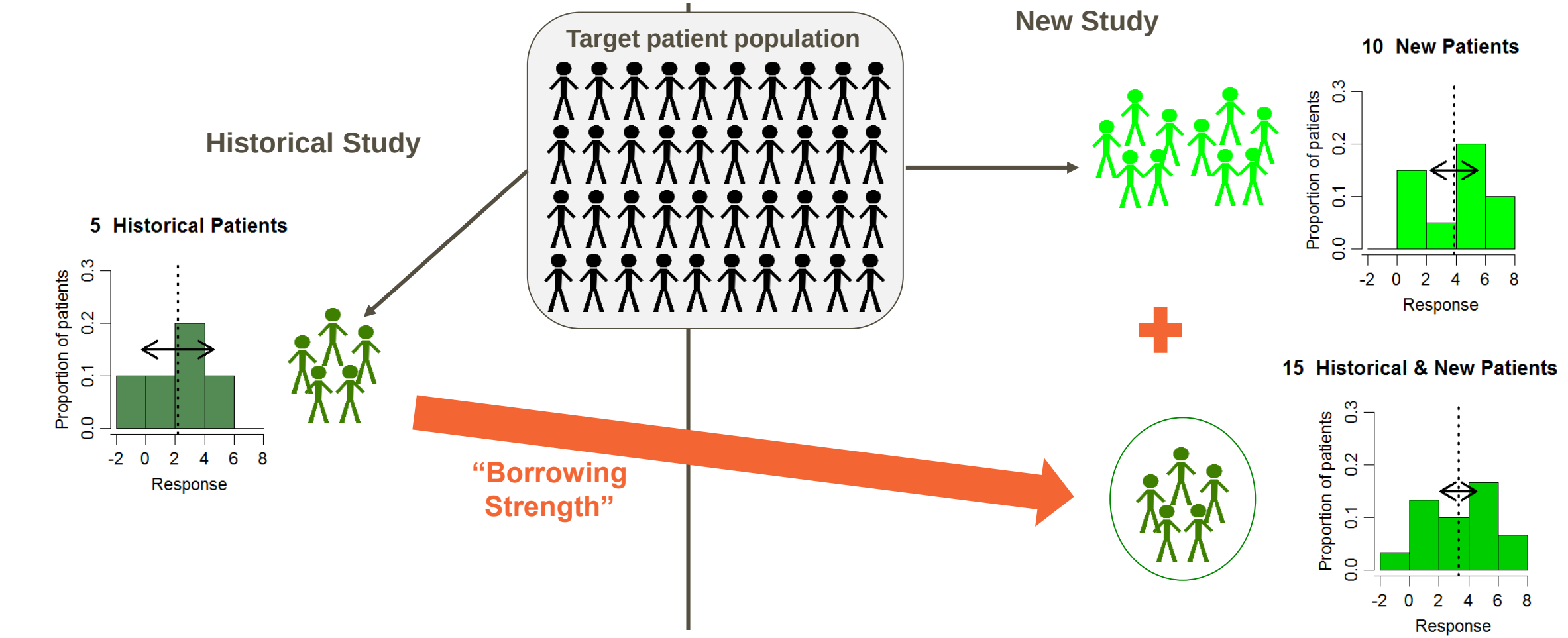


Model-Informed Knowledge Generation

Nayak CPT: 2018; https://www.ncbi.nlm.nih.gov/pubmed/29330855

- **Faster access for Patients (~ 0.5 - 4 y reduction) for first registration and expansion of therapeutic use**
  - Nayak S et al., Getting Innovative Therapies Faster to Patients at the Right Dose: Impact of Quantitative Pharmacology towards First Registration and Expanding Therapeutic Use. Clin. Pharmacol. Ther. 2018. http://onlinelibrary.wiley.com/doi/10.1002/cpt.978/epdf
- **$100M annual reduction in clinical study costs (Pfizer)**
  - Milligan, P.A., Brown, M.J., Marchant, B., Martin, S.W., van der Graaf, P.H. & Benson, N. Model-based drug development: a rational approach to efficiently accelerate drug development. Clin. Pharmacol. Ther. 93, 502–514 (2013).
- **$500M cost reduction in 3 year R&D spend (Merck & Co)**
  - Allerheiligen, S.R. Impact of modeling and simulation: myth or fact? Clin. Pharmacol. Ther. 96, 413–415 (2014).
  - Visser, S.A. et al., Implementation of quantitative and systems pharmacology in large pharma. CPT Pharmacometrics Syst. Pharmacol. 22, e142 (2014).

Borrowing strength from historical data

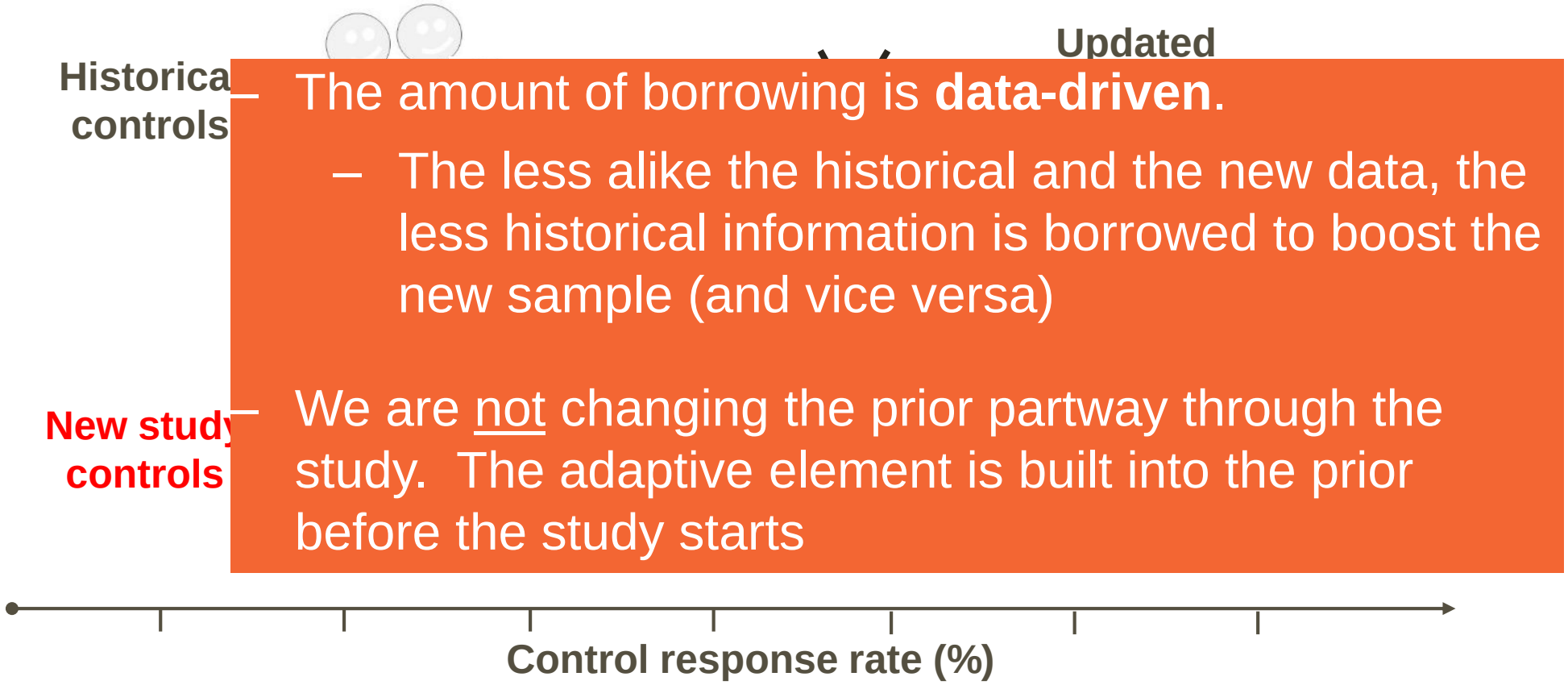


Choosing the right historical data to re-use

- If the historical data are representative of the target population for the new study, incorporating historical data into the analysis of the new study will always improve the operating characteristics of our decision rule, regardless of the rule itself or the true drug effect.
- Randomizing more subjects into the active arm and supplementing the control arm with historical subjects (effectively increasing the overall size of the study) can improve these operating characteristics even further.

But how can we be sure that the historical data are representative of the target population for the new study, and what happens when they are not?

Dynamic borrowing of historical data



- The amount of borrowing is **data-driven**.
  - The less alike the historical and the new data, the less historical information is borrowed to boost the new sample (and vice versa)
- We are not changing the prior partway through the study. The adaptive element is built into the prior before the study starts

Benefits of data re-use

- Fundamentally we are always trying to get the right balance between the risks of making incorrect GO decisions and the risks of making incorrect STOP decisions

- As with “traditional” study designs, we expect to be wrong some of the time, but can data re-use reduce how often we are wrong?

- If we routinely consider **“smart data re-use”** across the portfolio, then over the long run, we should win more often than we lose, and improvements in decision-making power will accrue (see next slide).

Benefits of data re-use

Illustrative savings from routine application of smart data re-use across the portfolio

Option 1:

Supplement planned sample size with historical data to increase precision

Traditional design

Data re-use design

Progress ~7% more “good” drugs and ~33% fewer “bad” drugs across the portfolio for ~5% less total spend*

Option 2:

Replace some planned subjects by historical data to achieve required sample size

Traditional design

Data re-use design

~$15M/year savings and ~12-20 weeks faster enrolment across development lifecycle‡

*Illustrative scenario – see backups for details

‡ Calculations from TransCelerate – see backups for details

Key messages

- Greater collaboration required among kineticists/modelers, statisticians and clinicians
- Quantitatively define clinically relevant effects and commercial targets
- Explicitly and quantitatively defined decision rules will help to reduce cycle times, robustify investment decisions & improve Probability of Success (PoS)
- Mining historical data (data re-use) to add to our cumulative knowledge is a low-cost and high-yield Activity.
- Bayesian methods and innovative designs improve decision making throughout drug development
- These methods enable more efficient deliver medicines to patients
- Regulators are open to discuss these options throughout the drug development process