Boehringer Ingelheim

From Hunch to Hedge

Applying Bayesian Thinking in Clinical Development Where It Matters Most

Kevin Kunzmann, 21.10.2026, Bayes Pharma

Life forward

DISCLAIMER

All views expressed in this presentation are my own.

They do not necessarily reflect the views or positions of Boehringer Ingelheim Pharma GmbH & Co. KG

Boehringer Ingelheim

Why does a successful PoC still leave so much risk?

Pr(Program Success | Proof of Concept)

Why it matters?

- Conditional probability determines late phase design decisions and **investment decisions**
- Key factor for **late-stage development efficiency**

What to do about it?

- How does a Bayesian perspective help to identify **actionable drivers** of this key quantity?
- What are the take-aways for program and Go/No Go design?

Key notation

Event	Working definition
PoC	Proof of Concept. The early program crosses a prespecified evidence threshold based on an early-phase endpoint. Early phase efficacy is predictive of late phase efficacy.
Viable (V)	The true late effect meets a clinically relevant threshold. Safety and execution risks not considered in this talk for sake of simplicity.
Success (S)	The later development program achieves its specified objective (e.g., two pivotal trials with significant primary endpoints).

Identifying actionable drivers of post-PoC success probability

Where can one improve Pr(S | PoC) in practice?

Interpret PoC as a diagnostic test [Browner & Newman 1987]

$$\Pr(S \mid PoC) = \Pr(S \mid V, PoC) \times \Pr(V \mid PoC) \quad \textbf{viable path}$$

$$+ \quad \Pr(S \mid \neg V, PoC) \times \Pr(\neg V \mid PoC) \quad \textbf{non-viable path}$$

Under plausible assumptions, the main drivers of Pr(S | PoC) are:

1) Pr(S | V) conditional expected ‘power‘ of phase III program (not assurance, Pr(S))
2) Pr(V | PoC) ‘positive predictive value’ of PoC trial

Pr(S | V) is often less actionable and largely set by regulatory requirements in late phase.

PoC trial design and operating characteristics are a key actionable lever for late-phase success probability.

Boehringer Ingelheim

Approximation 1: Little success without viability

For an efficacy-driven success criterion, the non-viable route is often small.

$$\Pr(S \mid \neg V, PoC) \approx 0$$

$$\Rightarrow \Pr(S \mid PoC) \approx \Pr(S \mid V, PoC) \times \Pr(V \mid PoC)$$

Interpretation

- A non-viable compound rarely meets the intended late-stage efficacy objective.
- Typically, two confirmatory trials needed post-PoC.

Limits

- Chance findings
- Priors with high probability of „almost viable“ effect size (see example later).

Approximation 2: Limited information beyond viability

If we knew the compound were viable, what would the PoC pass still tell us?

$$\Pr(S \mid V, PoC) \approx \Pr(S \mid V)$$

$$\Rightarrow \Pr(S \mid PoC) \approx \Pr(S \mid V) \times \Pr(V \mid PoC)$$

Interpretation

- Due to typically low sample size in PoC, the PoC result carries little information about the true effect size in the class of viable compounds.
- PoC effect size selection is ignorable given ‘viable’

Limits

- Very large PoC trials carry substantial information about the effect magnitude beyond viable/non-viable.
- PoC selection effect largest for high likelihood of ‘almost viable’ effects (see example later)

Two assumptions turn the exact decomposition into B × C

A **Pr(Success | Viable, PoC)**
Success probability given compound is viable and cleared PoC.

B **Pr(Success | Viable)**
Success probability given compound is viable.

C **Pr(Viable | PoC)**
Probability that PoC-positive compound is truly viable.

D **Pr(Success | ¬Viable, PoC)**
Success probability given compound is PoC-positive but NOT viable.

EXACT - LAW OF TOTAL PROBABILITY

$$\Pr(\text{Success} \mid \text{PoC}) = A \times C + D \times (1 - C)$$

viable share not-viable share

NARRATIVE APPROXIMATION

$$\Pr(\text{Success} \mid \text{PoC}) \approx B \times C$$

Replace A with B; omit D(1 – C).

EXACT – APPROXIMATION = TWO GAP TERMS

$$D(1 - C) + (A - B)C$$

$D(1 - C)$ “Success requires viability” (Approximation 1)

$(A - B)C$ “PoC adds no information given viability” (Approximation 2)

Boehringer Ingelheim

How close is the approximation in practice? A generative model.

Late effect and viability

$$\theta_L \mid \neg V \sim TN(0, \sigma^2; [0, \delta_V)) \qquad Pr(\neg V) = 1 - p$$

$$\theta_L \mid \quad V \sim TN(\mu, \sigma^2; [\delta_V, 0.6]) \qquad Pr(V) = p$$

TN denotes truncated normal

Early-late effect transport

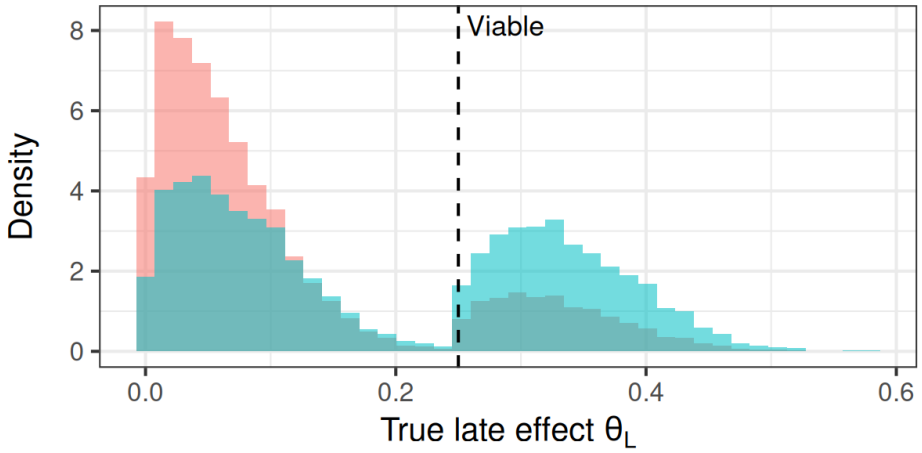
$$\theta_E = \theta_L + \varepsilon, \quad \varepsilon \sim N(0, \tau^2) \quad \text{(no systematic bias assumed)}$$

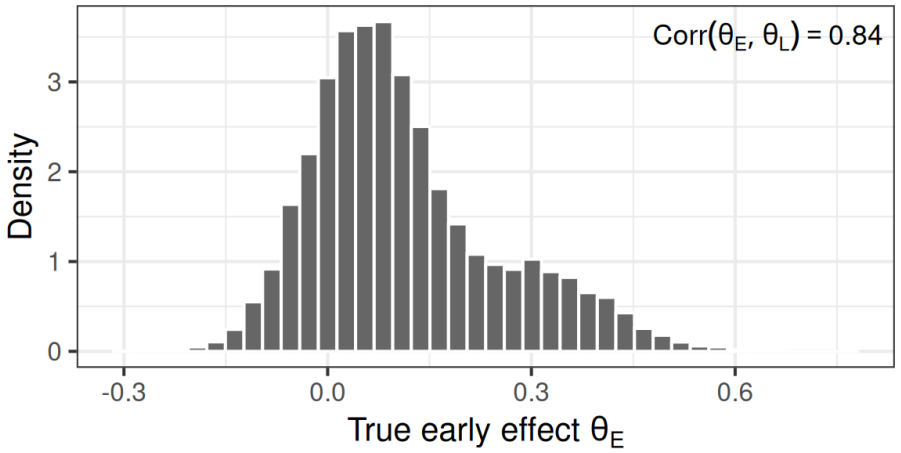
For small σ^2, the prior assumes that the compound either is viable or not, with some uncertainty about the effect size.

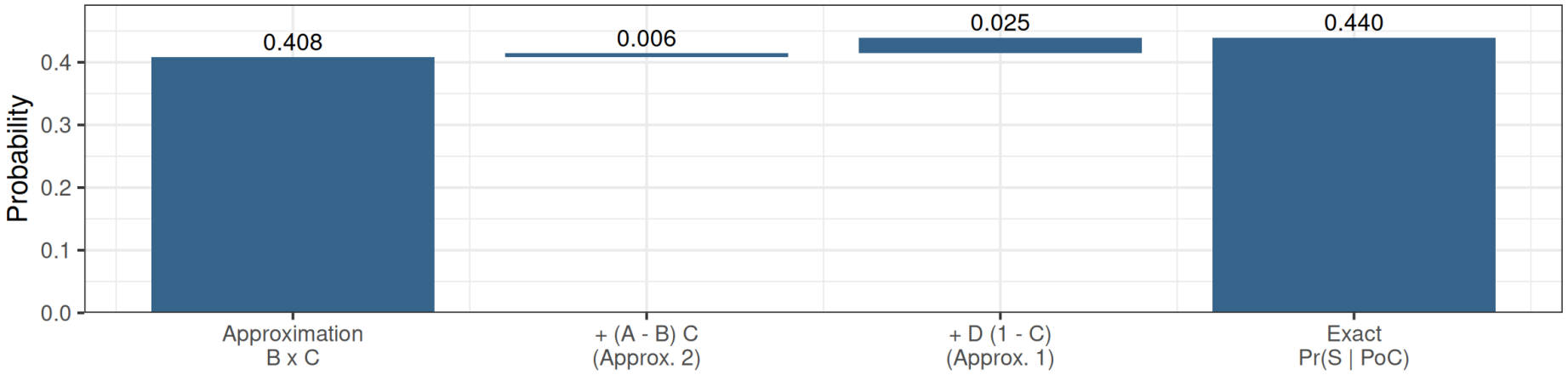
For larger σ^2, effect sizes are similarly likely across [0, 0.6].

Boehringer Ingelheim

Reference scenario: Separation viable/non-viable late stage effect



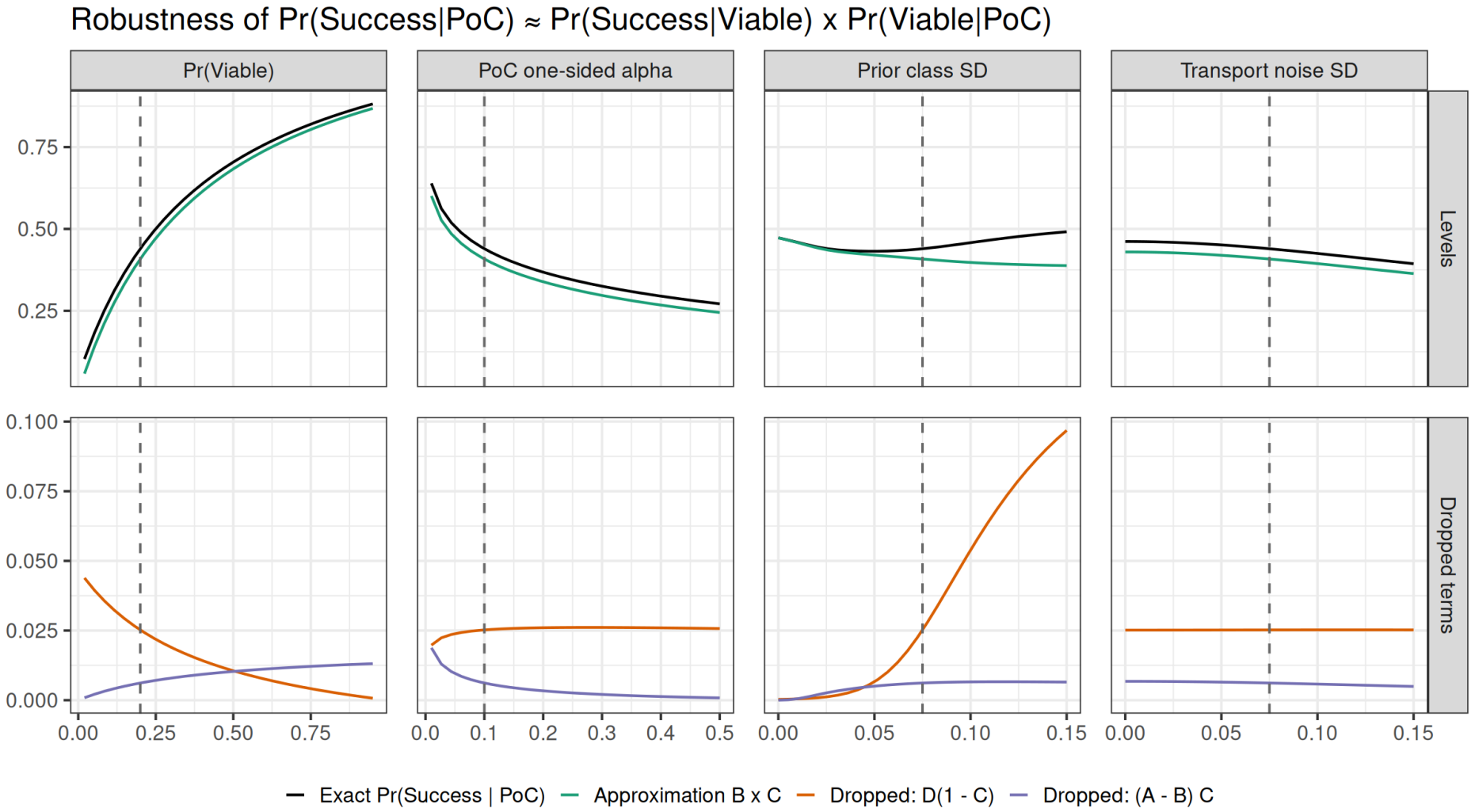




- Compound mostly viable or not ($\sigma^2 = 0.075$)
- PoC sample size of 50 per arm with 10% α_E
- Two phase III trials with 2.5% α_L and 80% power at minimal viable effect $\delta_V = 0.25$
- **Approximation holds well**

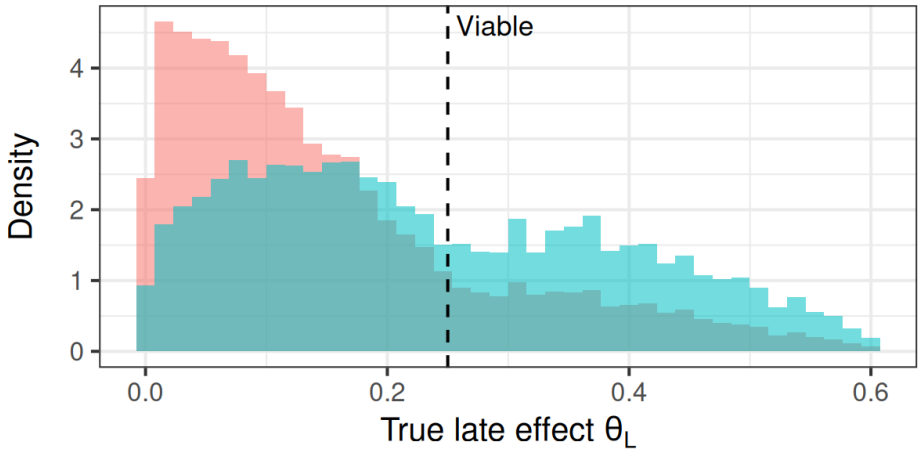
Boehringer Ingelheim

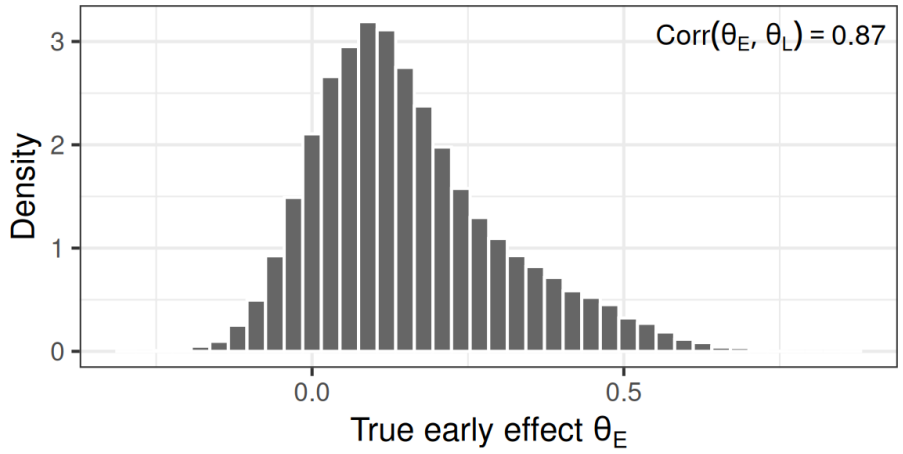
Robust across many situations, exception – weak separation of viability

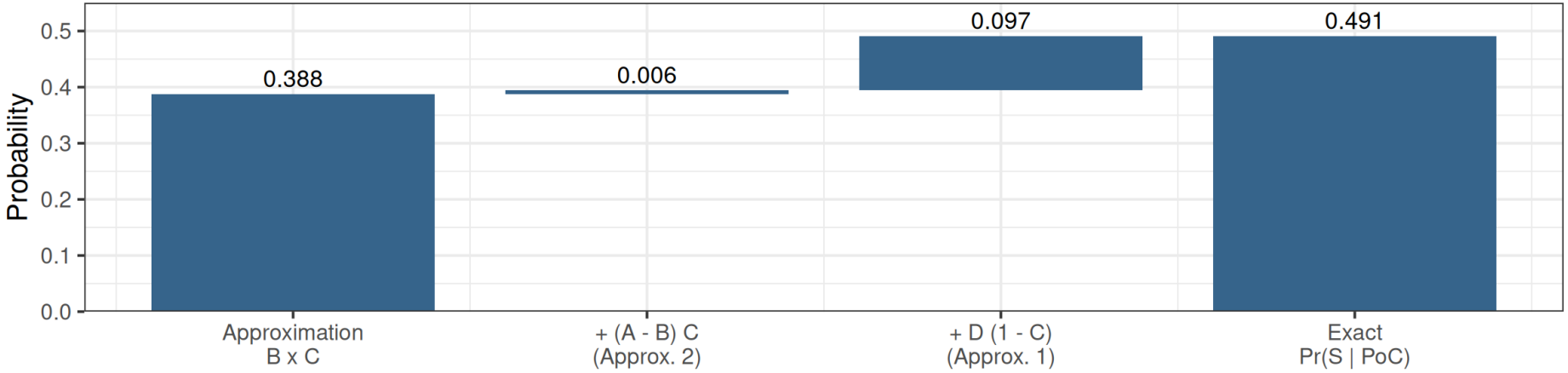


- (A - B) C small
- „PoC adds no information given viability “ is often justified
- D (1- C) increases if prior class SD is large (σ^2)
- “Success requires viability“ less plausible (see next slide)

What happens if “viability” is poorly separated? “False”-positives increase!





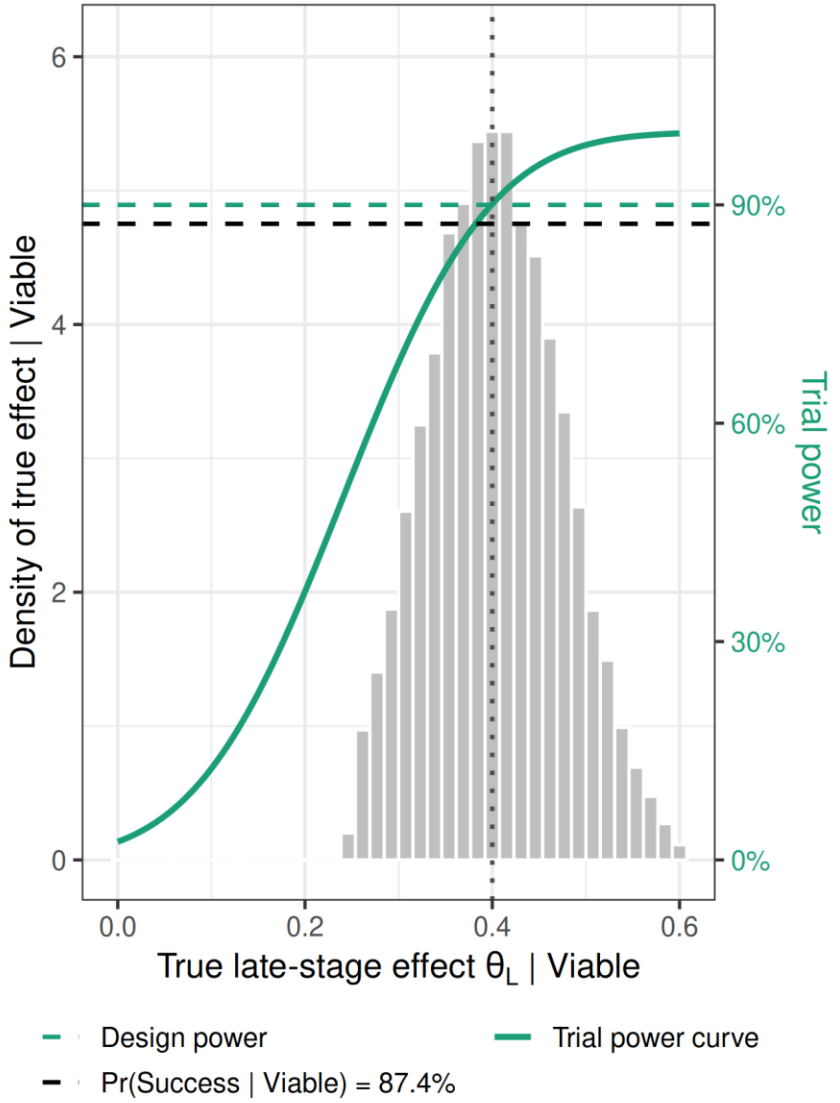


- Larger prior class SD (σ^2) means more „almost viable“ compounds
- More late-stage successes that are technically not viable but still have large effect size
- Pr(S | V) × Pr(V | PoC) still by far biggest term

Boehringer Ingelheim

How actionable is the first factor Pr(S | V) really?

- Late-stage operating characteristics **largely fixed by regulatory**
  - Two pivotal trials with one-sided alpha of 2.5% and around 90% power on target profile alternative
- Averaging power curve with respect to θ | V typically leads to per-trial sensitivity slightly below target power
  - This often leads to Pr(S | V) with two trials between 0.7 and 0.8 (e.g. $0.87^2 = 0.76$)
- **Late phase focus: Find sample size/cost/time effective design for realizing success probability potential given regulatory constraints**
- **Limited potential for increasing directly**



Boehringer Ingelheim

Main *actionable* driver for post-PoC success rates is the positive predictive value of the PoC

$$Pr(S \mid PoC) \approx Pr(S \mid V) \times Pr(V \mid PoC)$$

The positive predictive value of PoC

Among compounds with a positive PoC, what proportion are truly viable?

$$\Pr(V \mid PoC) = \frac{\Pr(PoC \mid V) \times \Pr(V)}{\Pr(PoC \mid V) \times \Pr(V) + \Pr(PoC \mid \neg V) \times \Pr(\neg V)}$$

$$PPV := \frac{Sensitivity \times p}{Sensitivity \times p + (1 - Specificity) \times (1 - p)}$$

Inherently Bayesian since ‘V‘ is not directly observable.

A back of the envelope estimate for post-PoC success probability

Illustrative PoC setting: low prior viability and limited precision (n < 50 per arm) constrain PoC PPV.

$Pr(V) \approx 0.15$ **$Pr(PoC \mid V) \approx 0.85$** **$Pr(PoC \mid \neg V) \approx 0.10$**

10% false-positive rate and 85% sensitivity correspond to a Cohen's d of around 0.54 with n = 50 per arm.

$Pr(S \mid V) \approx 0.80$

Two phase III trials at 90% power

$$PoC\ PPV \approx (0.85 \times 0.15) \div (0.85 \times 0.15 + 0.10 \times 0.85) = 60\%$$

$$Pr(S \mid PoC) \approx Pr(S \mid V) \times Pr(V \mid PoC) = 0.8 \times 0.6 = 48\%$$

In practice, PoC PPV often lower due to transportability issues (population/endpoint) or sample size constraint.

Boehringer Ingelheim

Improving PoC PPV

- **Maximize information about late-stage effect in population / endpoint**
  - Magnitude of effect is less important than transportability to late-stage setting,
  - The less predictive the PoC endpoint / population is, the worse the PoC PPV for late-stage success
- PPV is highly sensitive to the PoC specificity (lower false go), **keep PoC alpha tight**
  - Minimizing PoC sample size by accepting higher PoC alpha reduces late-stage efficiency
- **Increase PoC sample size**
  - often more cost-effective than failure in late phase [Kirchner et al. 2016]
- Low **PoC PPV can be compensated for by more aggressive futility stopping** in late phase

Two caveats define where the simple framework stops

Use the framework to build intuition - but match the model to the decision and the underlying biology.

01 Viability may be continuous

A binary viable / non-viable split can be artificial when late-stage success changes gradually with the true effect size.
[Pater & Willan 1984] [Sailer & Kunzmann 2026]

WHAT REMAINS TRUE

PoC design remains a key driver of late-stage success.

BETTER LENS

Ask how PoC filters or enriches the effect-size distribution.

02 The approximation is not an optimizer

The product approximation is useful for back-of-the-envelope checks and for explaining the key drivers of late-stage success.
[Kirchner et al. 2016]

USE IT FOR

Intuition, communication, and rapid plausibility checks.

FOR DECISIONS

Use exact calculation or simulation for program and portfolio optimization.

Bottom line: use the approximation to explain; use the full model to decide.

Where Bayesian thinking matters most

Two factors largely determine late-stage success given PoC. Often, only one is truly actionable.

$$\Pr(\text{Success} \mid \text{PoC}) \approx \Pr(\text{Success} \mid \text{Viable}) \times \mathbf{\Pr(\text{Viable} \mid \text{PoC})}$$

The first factor - Late phase program sensitivity

Often difficult to increase meaningfully beyond regulatory minimum requirements.

The second factor – PoC PPV

Can be meaningfully affected by investing in PoC operating characteristics,
often cost-effective due to small sample sizes.

The probability for viability given PoC makes Bayesian thinking essential to right-sized PoC design.

The biggest opportunity to increase your late-phase success rate after PoC …
is the quality of your PoC decision framework.

Bayesian thinking helps to make this connection clearer.

References

- Browner WS, Newman TB. Are all significant p values created equal?: The analogy between diagnostic tests and clinical research. JAMA. 1987;257(18):2459–2463. doi:10.1001/jama.1987.03390180077027.
- Pater JL, Willan AR. Clinical trials as diagnostic tests. Controlled Clinical Trials. 1984;5(2):107–113. doi:10.1016/0197-2456(84)90117-X.
- Sailer O, Kunzmann K. What drug do we want to develop? Interpretation of results from proof-of-concept trials and de-risking of phase 3. Poster P064, PSI 2026.
- Kirchner M, Kieser M, Götte H, Schüler A. Utility-based optimization of phase II/III programs. Statistics in Medicine. 2016;35(2):305–316. doi:10.1002/sim.6624

Thank you!