Bayesian Changepoint and Predictive Responder Methods for Assessing Drug Efficacy in Multiple N-of-1 Trials of Rare Pediatric Diseases

Gabi Pragier, Nicholas B. Link, Bradley P. Carlin, and Raviv Pryluk

“BAYES2026” Conference / Leiden, NL / October 22, 2026

© Copyright 2026 PhaseV

Background

Rare Disease = Extreme Data Sparsity

- Ultra-rare CNS disorders (NPC1, CLN3, DNM1-related epilepsy) have extremely small patient pools.
- Randomized Placebo-Controlled trials are often **ethically unacceptable** or parentally rejected.
- Traditional statistical paradigms ($p < 0.05$) lack mathematical power in these ultra-sparse regimes.

Shifting Regulatory Landscape

- **FDA 2026 Draft Guidance:** Encourages greater use of Bayesian statistical methods in confirmatory trials.
- **EMA 2026 Reflection Paper:** Recognizes Bayesian hierarchical borrowing to overcome rare pediatric sample bounds.
- **Prasad & Makary (2026):** Shifts preference to 'one pivotal trial' with confirmatory evidence, bypassing the historical double-RCT requirement.

© Copyright 2026 PhaseV

Objective & Context

Produce statistical evidence that the ASO manufacturing process reliably modifies disease trajectories

- Patient-specific antisense oligonucleotides (ASOs) for ultra-rare neurodegenerative diseases
- Each patient is their own control in their own N-of-1 trial, possibly with its own unique endpoint
- Limited number of patients per genetic variant across multiple indications
- Previous work by Schmid and Yang (2022) shows how to use Bayesian meta-analytic techniques to combine across N-of-1 studies having common endpoints; includes implementation via MCMC
- We suggest two complementary statistical approaches to handle the multiple ASO aspect

© Copyright 2026 PhaseV

Statistical Strategy: Two Approaches

- Approach A (Primary): Bayesian Changepoint Model
  - Continuous posteriors on treatment effect magnitude
  - Gradient of evidence: slowing, arrest, improvement
  - Hierarchical pooling across patients and indications
- Approach B (Supporting): Predictive-Responder Analysis
  - Binary responder classification against natural history
  - Simpler to explain: "X of our N patients significantly exceeded expectations"
  - Cross-indication pooling without need for standardization
- We see the two approaches as complementary, not competing

© Copyright 2026 PhaseV

Approach A: Piecewise Linear Changepoint Model

Assume a clinical score of Y_{ij} for patient i at time t_{ij}, bounded in $[Y_{min}, Y_{max}]$ where Y_{max} represents optimal function

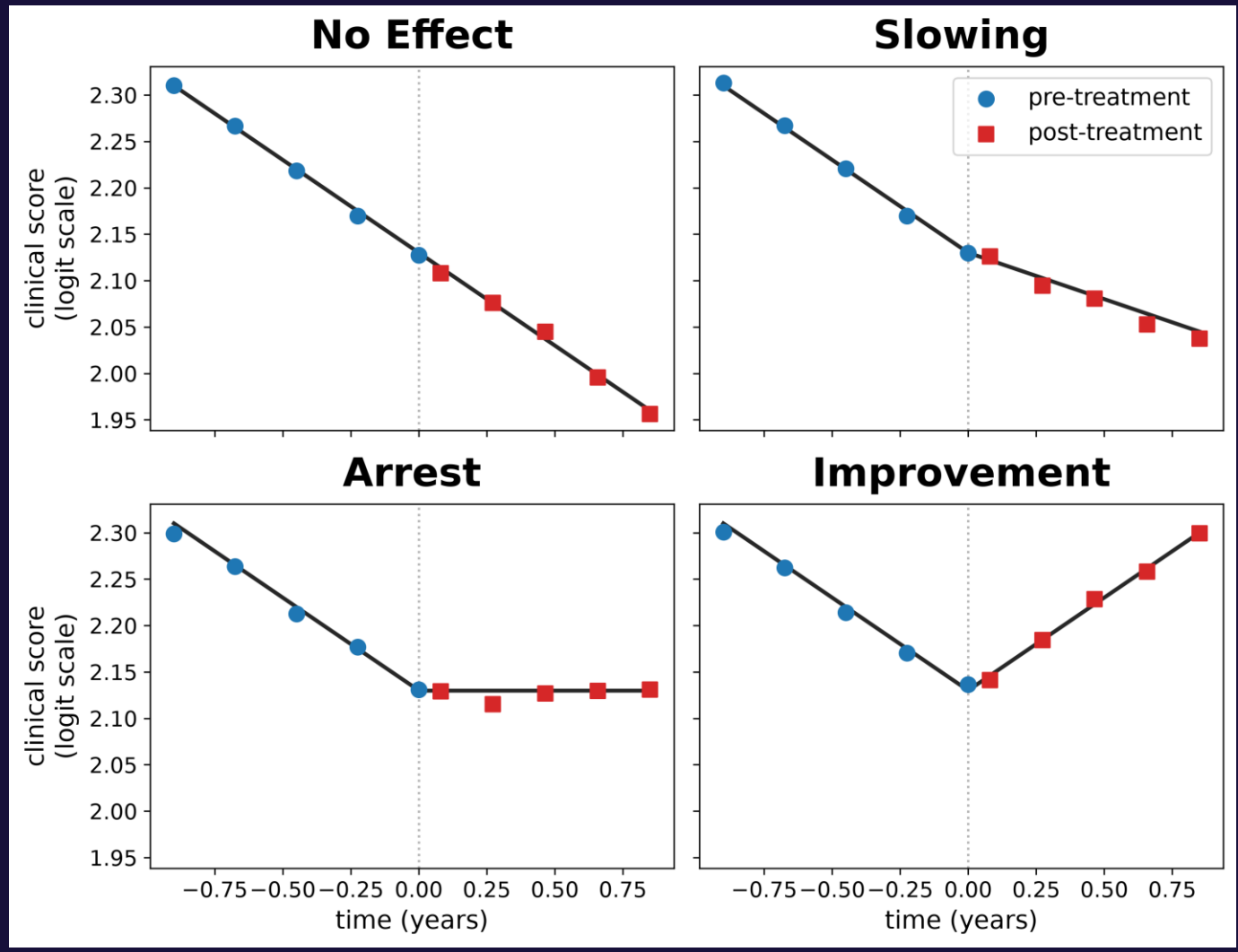
Normalize each response to proportion of optimal function,

$$Z_{ij} = \frac{Y_{ij} - Y_{min}}{Y_{max} - Y_{min}}$$

Reparametrize to $W_{ij} = logit(Z_{ij})$, and model W_{ij} as piecewise linear in time:

$$W_{ij} = \alpha_{1i} + \alpha_{2i}\, t + \alpha_{3i}\, (t - t^*)_+ + \varepsilon_{ij,}$$

where $(x)_+ = \max(0,x)$



© Copyright 2026 PhaseV

Gradient of Evidence

(A) Slowing

$P(\alpha_3 > 0 \mid \text{data})$

Treatment slows decline

▼

(B) Arrest

$P(\alpha_2 + \alpha_3 \geq 0 \mid \text{data})$

Progression is arrested

▼

(C) Improvement

$P(\alpha_2 + \alpha_3 > \delta_k \mid \text{data})$

Clinically meaningful improvement

Nested: $C \subset B \subset A$. Same structure at every level: patient → indication → process

© Copyright 2026 PhaseV

Hierarchical Model Structure (Three Levels)

Level 1: Patients within indications

$\alpha_{pki} \sim N(\mu_{pk}, \tau_{pk}^2)$

τ_{pk}: between-patient variability

▼

Level 2: Indications within process

$\mu_{pk} \sim N(\gamma_p, \omega_p^2)$

ω_p: between-indication variability

▼

Level 3: Process-level effect

$\gamma_p \sim N(\mu_{0p}, \sigma_{0p}^2)$

Prior means μ_{01} *and* μ_{02} (baseline, pre-slope) come from natural history data; treatment effect γ_3 has prior mean 0 (mildly skeptical prior)

One fit → posteriors at every level (patient → indication → process)

Indexing in this diagram: i indexes patient; k indexes indication; p indexes model parameter (baseline / pre-slope / treatment effect)

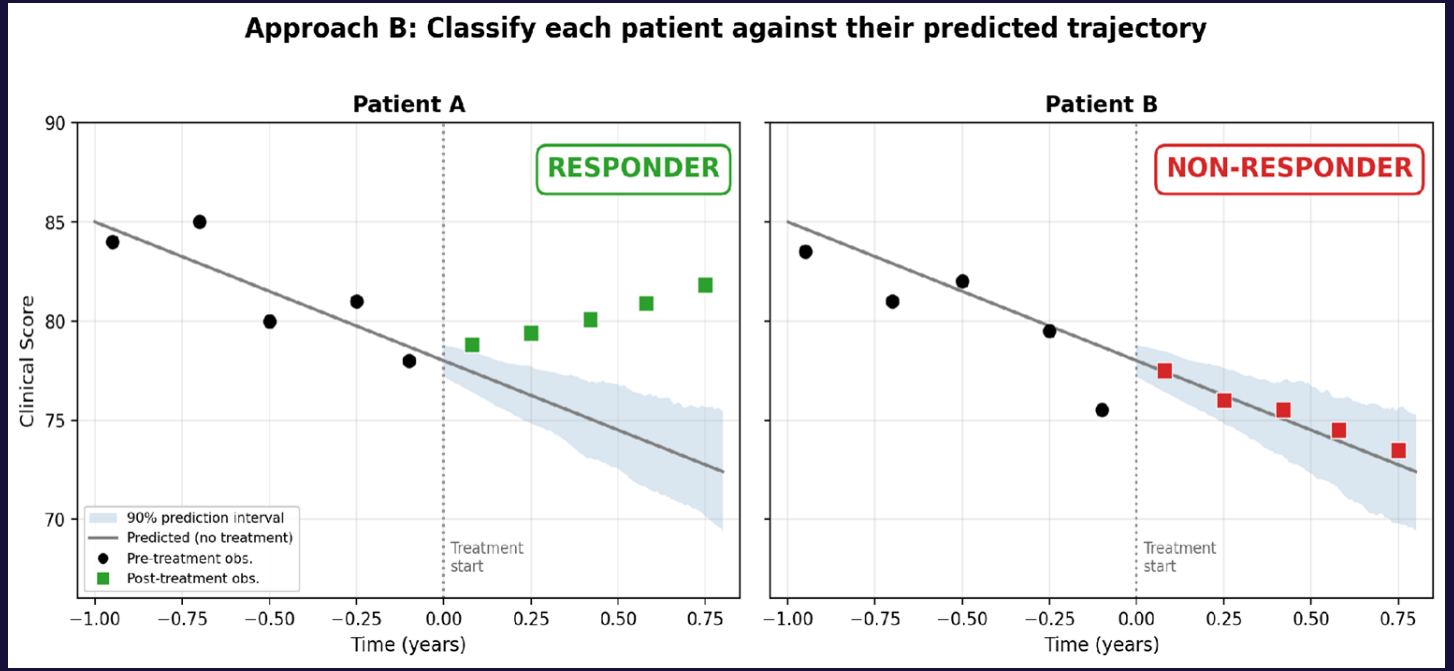
© Copyright 2026 PhaseV

Approach B: Predictive-Responder Analysis

Classify each patient as responder/non-responder, then assess the aggregate response rate

- Aggregate response rate via Beta-Binomial posterior
- Pools across indications (no standardization needed)

Tradeoff: Cannot model effect magnitude



© Copyright 2026 PhaseV

Approach B: Predictive-Responder Details

Model 1: Separate Regression (No Pooling)

- $W_{ij} = b_{0i} + b_{1i}t_{ij} + \epsilon_{ij}$
- Fitted separately per patient via standard OLS on pre-treatment visits.

Model 2: Hierarchical Model without Changepoint

- $W_{ij}|\ \alpha_{1ki}, \alpha_{2ki}, \sigma_i \sim N(\alpha_{1ki} + \alpha_{2ki}t_{ij}, \sigma_i^2)$
- Pools across patients & indications. Prediction intervals borrow strength.

Exceedance Success Tests:

Test 1: Last-Observation Exceedance

- Let $R = \sum y_i$, the number of last-observation responders ($y_i$values above the prediction bound)
- Under a $Beta(a, b)$ prior on θ_R, we obtain $\theta_R|R \sim Beta(a + R, b + N - R)$, hence $P(\theta_R > p_0|data)$
- Scored on final post-treatment visit status only. Simple and robust.

Test 2: All-Observation Exceedance

- Let $S_i^+ = \sum_j y_{ij}^+$, the total number of exceedances for patient i
- Standard beta-binomial: $S_i^+ \sim Bin(n_i, \theta_i^+)$ and $\theta_i^+ \sim Beta(\mu^+\kappa, (1-\mu^+)\kappa)$
- Vague priors on $\mu^+$and κ yield inference via the posterior probability $P(\mu^+ > p_0|data)$

© Copyright 2026 PhaseV

Approach Comparison

A: Changepoint Model (Primary)

- Continuous posterior on effect magnitude
- Gradient: slowing / arrest / improvement
- Standardize + hierarchical pooling
- Full magnitude preserved

B: Predictive-Responder (Supporting)

- Binary: responder or not
- Single threshold decision
- Natural pooling (count responders)
- Magnitude collapsed to binary

Again, approaches are complementary; A may be better for per-patient & per-indication evidence, but B better for process-level support

© Copyright 2026 PhaseV

Simulation Findings: The Evidence Gradient A ≥ B ≥ C

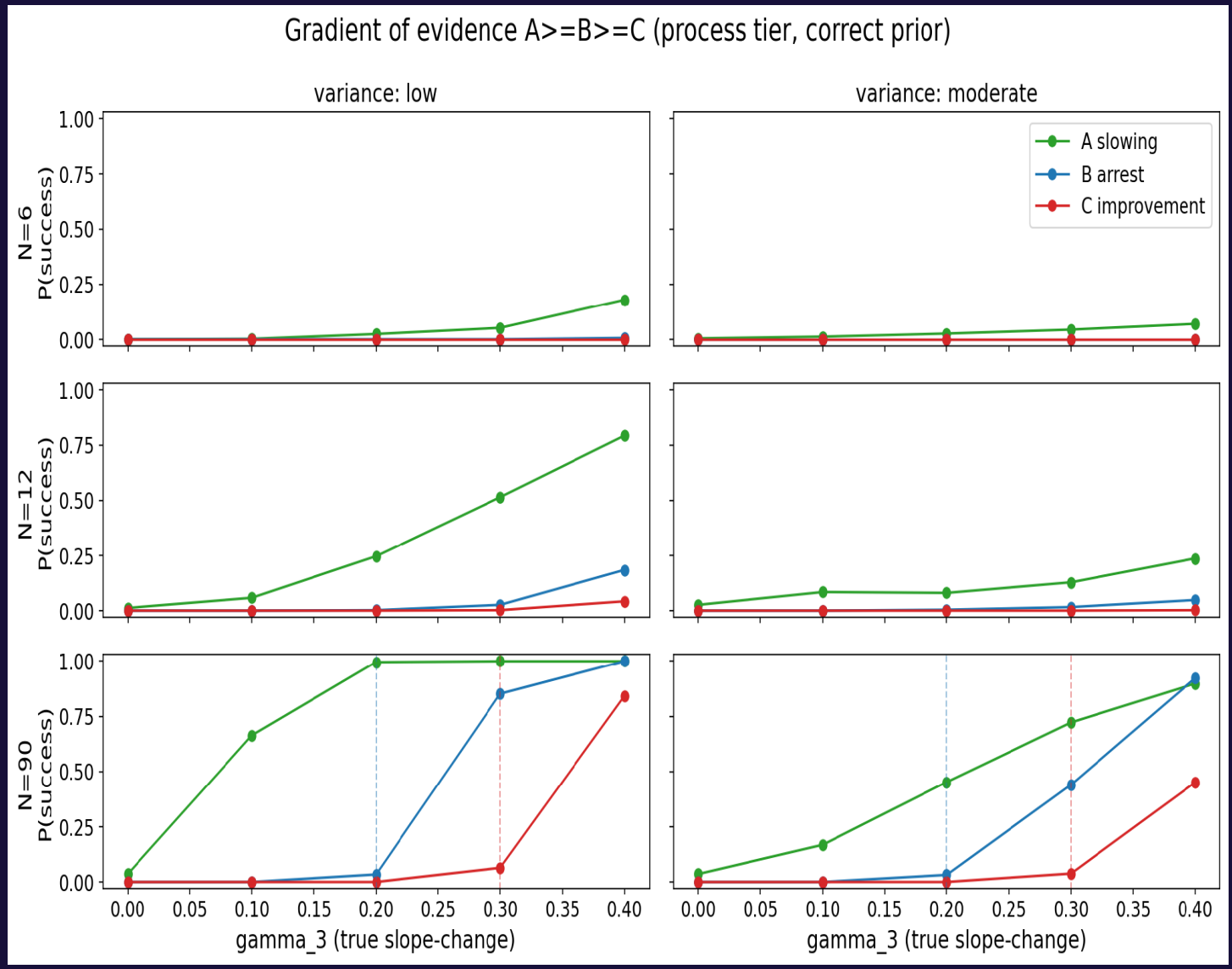
Ordering holds empirically
Power for A (slowing) ≥ B (arrest) ≥ C (improvement) at every sample size and variance regime

Improvement requires large γ_3
Near-zero power until $\gamma_3 > 0.3$, since the truth must offset the pre-slope $\gamma_2 = -0.2$.

N=6 underpowered
With only K=2 indications, even strong effects (γ_3=0.4) yield ~20% power at low variance

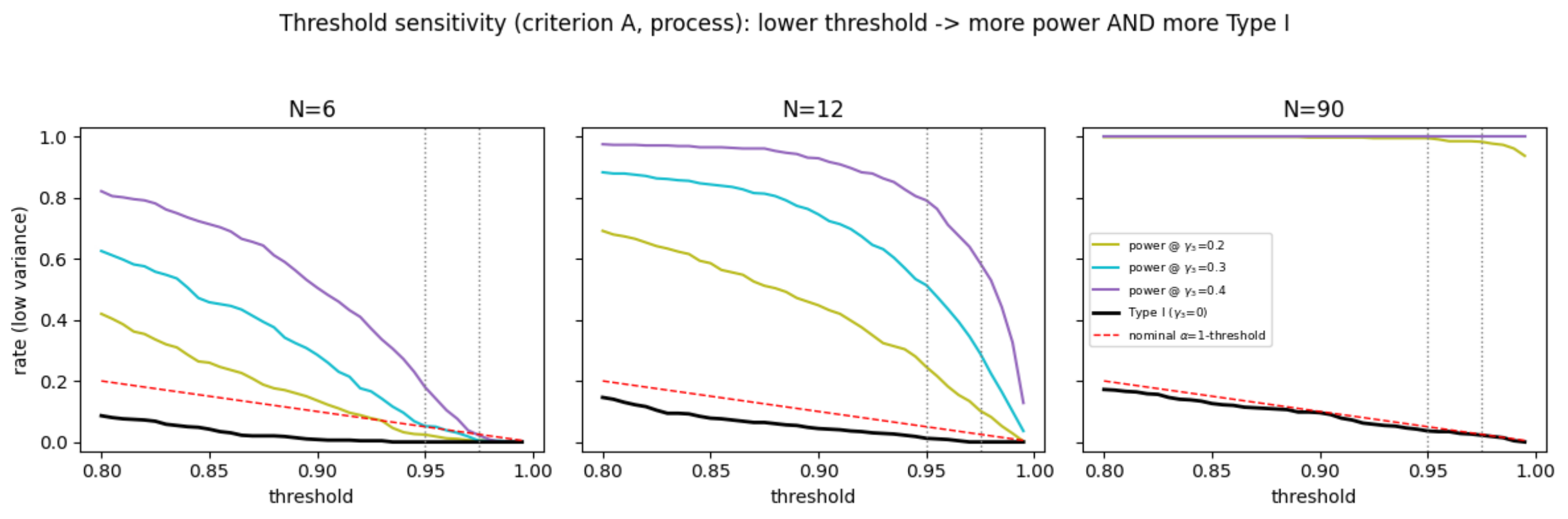
N=12 shows meaningful power
Under low variance and γ_3=0.4, power for slowing reaches ~80%.

N=90 saturates at low variance
Power hits 1.0 at γ_3=0.2 — confirming correct model implementation.



© Copyright 2026 PhaseV

Simulation Findings: Threshold Sensitivity



N=6 emerges as feasible!

- At γ_3=0.4 (optimistic), setting the threshold=0.80 delivers ~82% power with ~10% Type I error -- Remarkable performance for only 6 patients!

N=12 delivers good power and low Type I error with more traditional (higher) thresholds

N=90 offers a "sanity check"

- Power is essentially 1 and Type I error close to nominal across all decision thresholds

© Copyright 2026 PhaseV

Simulation Findings: K and N Both Drive Process Power

Process Tier Hardest to Estimate

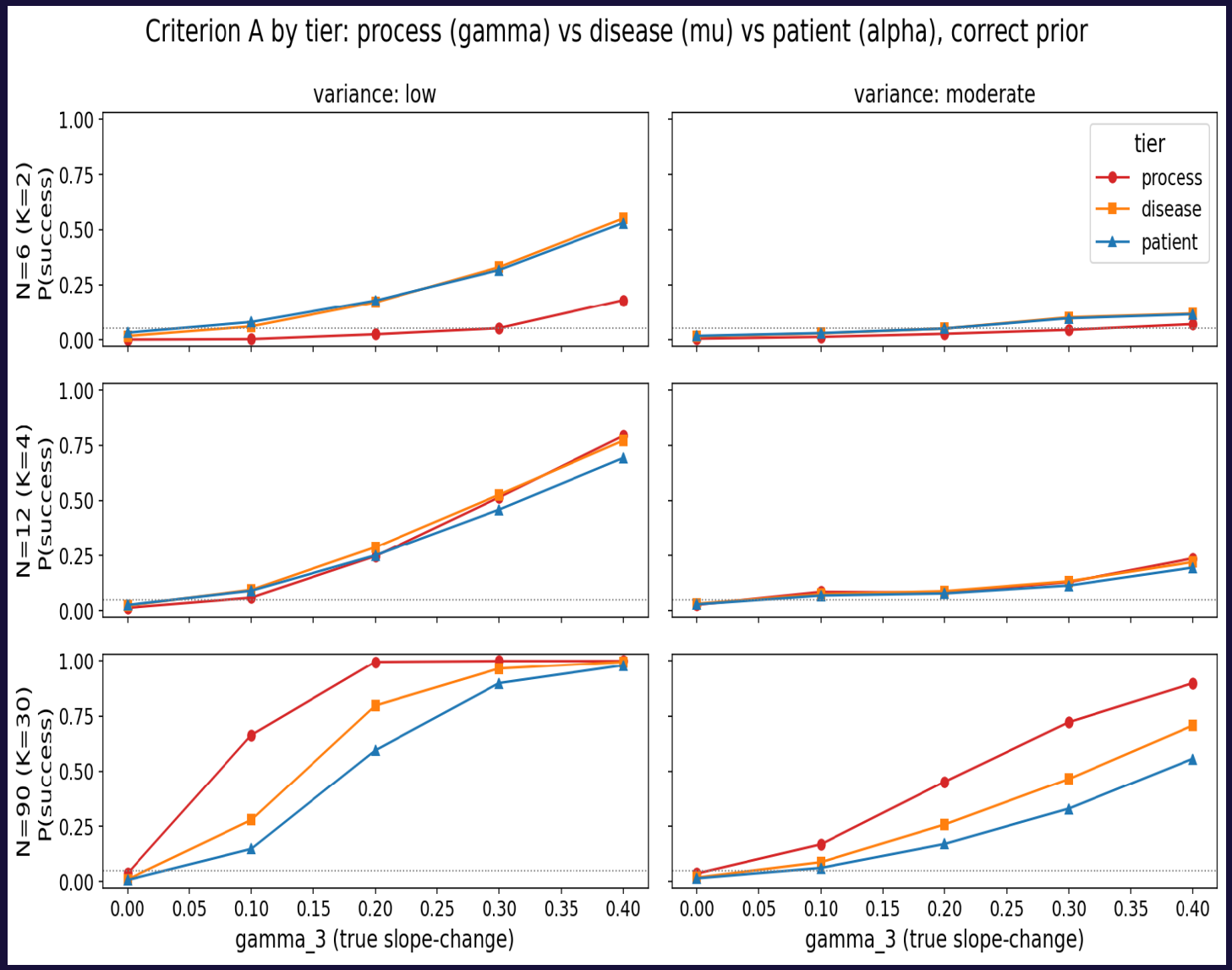
- At small N and K, process-level power is very low; patient- and disease-level power is moderate

Process Dominates as K Grows

- As K grows from 2 to 30, process-level power overtakes and saturates, dominating individual and disease-specific power.

Crucial Design Insight:

- Process-level inference benefits from multiple indications (↑K)
- Adding more patients per indication (↑N/K) will of course increase indication-level power
- We are currently investigating this issue with both this approach and Approach B



© Copyright 2026 PhaseV

Simulation Findings: Comparing the Two Approaches

We calibrate each approach to a common realized Type I error before comparing power

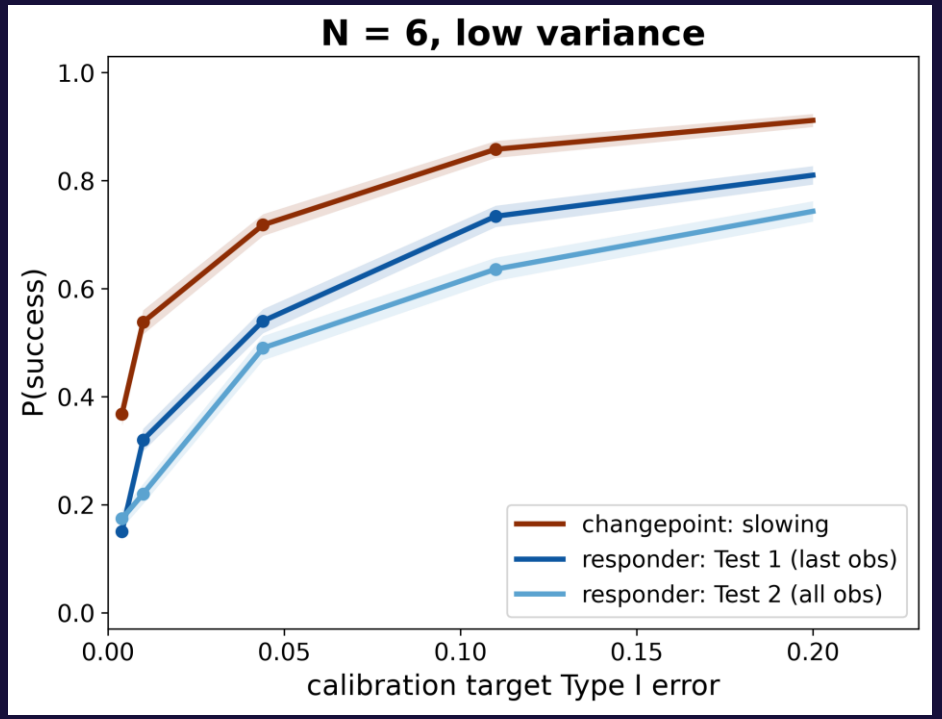
- Compare the “slowing” hypothesis, since both approaches share a null of no change between the pre- and post-intervention trajectory

At matched Type I error, the changepoint approach outperforms the responder approach

- At N=6 under low variance, calibrating every test to have 11% Type I error gives the changepoint 86% power for slowing, whereas the responder delivers just 73% and 64% under Test 1 and Test 2, respectively
- Changepoint approach is similarly dominant for every sample size and variance regime we investigated

Responder approach might be more palatable to medical investigators or regulators

- Easier to interpret clinically; provides interpretable statistics on proportions of people responding to treatment



© Copyright 2026 PhaseV

Additional Issues for Investigation

- *Treatment delay:* The effect of the treatment is presumably not immediate; is this lag time estimable, and how shall we model it (see Appendix 1)
- *Placebo effects:* Not currently estimable due to N-of-1 design, but tipping point analyses could help
- *Standardization sensitivity:* logit transform approximation, comparability across endpoints; candidate for regulatory input?
- Nonlinearity near scale bounds: Linearity on the logit scale typically fails near boundaries; might handle via beta regression (so we can avoid the logit transformation entirely; see Appendix 2)
- δ_k *thresholds:* require clinical input per indication
- *Robustness*: how sensitive are our conclusions to prior choice? Missing data? Heavy-tailed errors? Can we hope for a sample size that will help control this sensitivity?
  - (A referee has criticized our simulation studies as "too favorable to the proposed method"; we agree!)

© Copyright 2026 PhaseV

PHASE V

Adaptive. Causal. Intuitive.

Thank you for your attention!



Brad Carlin, PhD
brad@phasevtrials.com

© Copyright 2026 PhaseV

Appendix 1: Treatment Delay Modeling Options

Option A: Sharp Changepoint + Delay

- $W_{ij} = \alpha_1 + \alpha_2 \cdot t + \alpha_3 \cdot (t - t^* - \lambda)_+$
- $\lambda \geq 0$: estimated delay before slope change
- Pre-treatment trajectory continues for λ
- Abrupt onset once the delay elapses

Option B: Smooth Sigmoid Onset

- $\alpha_3 \cdot S(t-t^*; \lambda, \kappa) \cdot (t-t^*)$
- $S(x) = 1 / (1 + \exp(-\kappa(x-\lambda)))$
- κ sets the steepness of the ramp
- Gradual onset of treatment effect

Base model assumes no delay. With only 6 post-treatment points, λ is poorly identified –
Could be explored as a sensitivity analysis, or benefit from a clinically-informed prior distribution

© Copyright 2026 PhaseV

Appendix 2: Nonlinearity Near Scale Bounds

Option A: Logistic / Sigmoid Curves

- Replace linear segments with logistic functions
- Naturally respect scale bounds [Y_{min}, Y_{max}]
- Changepoint structure can be retained
- Cost: added model complexity, reduced model interpretability

Option B: Beta Regression

- Model standardized Z_{ij} on (0,1) directly
- Beta distribution lives naturally on (0,1)
- Mean trajectory keeps changepoint structure
- Bound-respecting without the logit approximation, but less standard and interpretation still tricky

Linear decline must eventually fail near a scale floor/ceiling. Pragmatic view: for short windows linearity holds locally, and the Level 3 logit transform already compresses near the bounds. Decision needed: how long is follow-up vs total disease course?

© Copyright 2026 PhaseV