Simplified N-of-1 Trial Simulation: Mathematical Foundations and Design Rationale

pmsimstats2025 Project

2025-11-21

Contents

1 Executive Summary

This white paper documents the systematic simplification of an N-of-1 clinical trial simulation based on Hendrickson et al. (2020). The original implementation used complex Gompertz response curves and a monolithic 26×26 covariance matrix. We present a series of mathematically equivalent but conceptually clearer simplifications:

1. **Rate-based response model** replacing Gompertz curves
2. **Time-based AR(1) correlation** replacing compound symmetry
3. **Two-stage data generation** separating participant and response variables
4. **Guaranteed positive-definiteness** via grid-snapping

Each simplification is justified mathematically and evaluated for conceptual clarity, computational efficiency, and biological plausibility.

---

2 Introduction

2.1 Background

N-of-1 trials are randomized crossover designs where a single participant serves as their own control. The hybrid design combines an open-label run-in with blinded crossover periods to estimate individual treatment effects while accounting for placebo (expectancy) effects.

2.2 Original Complexity

The original Hendrickson-based simulation involved:

- **Three response factors**: Biological Response (BR), Expectancy Response (ER), Time-variant Response (TR)
- **Gompertz trajectories**: Sigmoidal curves with 3 parameters each (max, displacement, rate)
- **26×26 covariance matrix**: 3 factors × 8 timepoints + biomarker + baseline
- **Non-transparent construction**: Correlations filled element-by-element without PD guarantees

2.3 Goals of Simplification

1. **Conceptual clarity**: Each component should be independently understandable
2. **Mathematical transparency**: All assumptions explicit and justified
3. **Robustness**: Guaranteed valid (positive definite) covariance matrices
4. **Flexibility**: Easy to modify individual components

---

3 Simplification 1: Rate-Based Response Model

3.1 Original: Gompertz Curves

The original model used Gompertz functions for each response factor:

$$f(t) = \text{max} \cdot \exp\left(-\text{disp} \cdot \exp(-\text{rate} \cdot t)\right)$$

This S-shaped curve has three parameters:

- **max**: Asymptotic maximum effect
- **disp**: Displacement (horizontal shift)
- **rate**: Growth rate

3.1.1 Problems with Gompertz

1. **Over-parameterized**: 3 parameters per factor × 3 factors = 9 response parameters
2. **Non-intuitive**: Displacement and rate interact in complex ways
3. **Asymptotic behavior**: Effect plateaus, but clinical effects often accumulate linearly

3.2 Simplified: Linear Rate Model

We replace Gompertz with simple linear accumulation:

$$\text{Effect}(t) = \text{rate} \times \text{time}$$

3.2.1 Three-Factor Rate Model

For each factor, we define a single rate parameter (points per week):

$$\text{BR}_{\text{rate}} = 0.5 \text{ points/week on drug} \tag{1}$$
$$\text{ER}_{\text{rate}} = 0.2 \text{ points/week} \times \text{expectancy} \tag{2}$$
$$\text{TR}_{\text{rate}} = 0.1 \text{ points/week} \tag{3}$$

The response at time t is:

$$\text{BR}(t) = \text{BR}_{\text{rate}} \times (\text{cumulative weeks on drug}) \tag{4}$$

$$\text{ER}(t) = \text{ER}_{\text{rate}} \times \sum_{s \le t} \text{expectancy}(s) \tag{5}$$

$$\text{TR}(t) = \text{TR}_{\text{rate}} \times (\text{weeks in trial}) \tag{6}$$

3.2.2 Carryover Model

When drug is discontinued, BR doesn't immediately drop to zero. We model carryover as a partial effect at the first off-drug timepoint:

$$\text{BR}(t) = \begin{cases} \text{BR}_{\text{rate}} \times \text{weeks\_on\_drug} & \text{if on drug} \\ \text{BR}_{\text{accumulated}} \times \text{carryover\_decay\_rate} & \text{if first week off} \\ 0 & \text{if subsequent weeks off} \end{cases}$$

For example, with carryover_decay_rate = 0.5:

- Week 10 (on drug, 4 weeks): BR $= 0.5 \times 4 = 2.0$
- Week 11 (first week off): BR $= 2.0 \times 0.5 = 1.0$
- Week 12 (second week off): BR $= 0$

3.2.3 Why This Is Better

Aspect	Gompertz	Linear Rate
Parameters	9 (3 per factor)	3 (1 per factor)
Interpretation	Complex	Direct (points/week)
Flexibility	Fixed asymptote	Unbounded accumulation
Clinical face validity	Moderate	High

Intuition: Clinicians think in terms of "improvement per week," not asymptotic limits and displacement parameters.

---

4 Simplification 2: Time-Based AR(1) Correlation

4.1 Original: Compound Symmetry

The original model used compound symmetry within each response type:

$$\text{Corr}(Y_i, Y_j) = \rho \quad \text{for all } i \neq j$$

This means measurements at week 4 and week 8 (4 weeks apart) have the same correlation as measurements at week 8 and week 9 (1 week apart).

4.1.1 Problems with Compound Symmetry

1. **Biologically implausible**: Nearby measurements should be more correlated
2. **Wastes correlation budget**: High correlation everywhere leaves less room for cross-correlations
3. **More prone to PD failures**: Concentrates eigenvalues

4.2 Simplified: Time-Based AR(1)

We use an autoregressive structure based on actual time lags:

$$\text{Corr}(Y_{t_i}, Y_{t_j}) = \rho^{|t_i - t_j|}$$

where t_i and t_j are the actual week numbers.

4.2.1 Example Correlation Matrix

For measurement weeks $\{4, 8, 9, 10, 11, 12, 16, 20\}$ with $\rho = 0.8$:

Table 2: Time-based AR(1) correlation matrix

	W4	W8	W9	W10	W11	W12	W16	W20
W4	1.00	0.41	0.33	0.26	0.21	0.17	0.07	0.03
W8	0.41	1.00	0.80	0.64	0.51	0.41	0.17	0.07
W9	0.33	0.80	1.00	0.80	0.64	0.51	0.21	0.09
W10	0.26	0.64	0.80	1.00	0.80	0.64	0.26	0.11
W11	0.21	0.51	0.64	0.80	1.00	0.80	0.33	0.13
W12	0.17	0.41	0.51	0.64	0.80	1.00	0.41	0.17
W16	0.07	0.17	0.21	0.26	0.33	0.41	1.00	0.41
W20	0.03	0.07	0.09	0.11	0.13	0.17	0.41	1.00

4.2.2 Key Comparisons

Week Pair	Time Lag	Compound Symmetry	Time-Based AR(1)
W4 - W8	4 weeks	0.80	$0.8^4 = 0.41$
W8 - W9	1 week	0.80	$0.8^1 = 0.80$
W12 - W16	4 weeks	0.80	$0.8^4 = 0.41$
W4 - W20	16 weeks	0.80	$0.8^{16} = 0.03$

4.2.3 Guaranteed Positive Definiteness

The AR(1) correlation function $K(t_1, t_2) = \rho^{|t_1 - t_2|}$ is a valid positive definite kernel for $\rho \in (0, 1)$. This is the exponential covariance function, widely used in spatial statistics and time series.

Proof sketch: The AR(1) process $Y_t = \rho Y_{t-1} + \epsilon_t$ has this covariance structure, and valid stochastic processes always have PD covariance matrices.

4.2.4 Why This Is Better

Aspect	Compound Symmetry	Time-Based AR(1)
Biological realism	Low	High
Eigenvalue spread	Concentrated	Distributed
PD robustness	Lower	Higher
Interpretability	"Same correlation everywhere"	"Correlation decays with time"

Intuition: Your blood pressure yesterday is more predictive of today's than last month's. Correlation should decay with time.

---

5 Simplification 3: Two-Stage Data Generation

5.1 Original: Monolithic 26×26 Matrix

The original approach built a single 26×26 covariance matrix:

$$\Sigma_{26\times 26} = \begin{pmatrix} \Sigma_{\text{BR}} & \Sigma_{\text{BR,ER}} & \Sigma_{\text{BR,TR}} & \Sigma_{\text{BR,BM}} & \Sigma_{\text{BR,BL}} \\ \Sigma_{\text{ER,BR}} & \Sigma_{\text{ER}} & \Sigma_{\text{ER,TR}} & \Sigma_{\text{ER,BM}} & \Sigma_{\text{ER,BL}} \\ \vdots & & \ddots & & \vdots \\ \Sigma_{\text{BL,BR}} & \cdots & & & \Sigma_{\text{BL}} \end{pmatrix}$$

Then generated all 26 variables jointly:

$$\mathbf{X} \sim \mathcal{N}(\mathbf{0}, \Sigma_{26\times 26})$$

5.1.1 Problems with Monolithic Approach

1. **Opaque structure**: Hard to see how biomarker affects responses
2. **All-or-nothing PD**: If not PD, entire matrix rejected
3. **No clear causal interpretation**: Everything generated simultaneously

5.2 Simplified: Two-Stage Conditional Generation

We partition variables into:

- $\mathbf{X}_2$: Participant variables (biomarker, baseline) - 2 dimensions
- $\mathbf{X}_1$: Response variables (BR, ER, TR at 8 timepoints) - 24 dimensions

5.2.1 Partitioned Covariance

$$\Sigma = \begin{pmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{21} & \Sigma_{22} \end{pmatrix}$$

where:

- Σ_{22}: 2×2 covariance of (biomarker, baseline)
- Σ_{11}: 24×24 covariance of responses
- Σ_{12}: 24×2 cross-covariance (how biomarker/baseline relate to responses)

5.2.2 Conditional Distribution Theorem

For jointly normal variables:

$$\mathbf{X}_1 | \mathbf{X}_2 \sim \mathcal{N}(\mu_{1|2}, \Sigma_{1|2})$$

where:

$$\mu_{1|2} = \Sigma_{12}\Sigma_{22}^{-1}(\mathbf{X}_2 - \mu_2) \tag{7}$$

$$\Sigma_{1|2} = \Sigma_{11} - \Sigma_{12}\Sigma_{22}^{-1}\Sigma_{21} \tag{8}$$

5.2.3 Two-Stage Algorithm

Stage 1: Generate participant characteristics

$$\begin{pmatrix}\text{biomarker}\\ \text{baseline}\end{pmatrix} \sim \mathcal{N}\left(\begin{pmatrix}\mu_{\text{BM}}\\ \mu_{\text{BL}}\end{pmatrix}, \Sigma_{22}\right)$$

Stage 2: Generate responses conditional on participant characteristics

$$\begin{pmatrix}\text{BR}_1\\ \vdots\\ \text{TR}_8\end{pmatrix} \sim \mathcal{N}\left(\Sigma_{12}\Sigma_{22}^{-1}\begin{pmatrix}\text{biomarker} - \mu_{\text{BM}}\\ \text{baseline} - \mu_{\text{BL}}\end{pmatrix}, \Sigma_{1|2}\right)$$

5.2.4 Mathematical Equivalence

This two-stage procedure is **exactly equivalent** to generating from the joint 26×26 distribution. The conditional distribution formula preserves all correlations.

Proof: By construction, the joint density factors as $p(\mathbf{X}_1, \mathbf{X}_2) = p(\mathbf{X}_2) \cdot p(\mathbf{X}_1|\mathbf{X}_2)$.

5.2.5 Why This Is Better

Aspect	Monolithic	Two-Stage
Causal interpretation	Unclear	Clear (BM → responses)
Matrix inversion	None	2×2 only (trivial)
Debugging	Hard	Test each stage
Conceptual model	Simultaneous	Sequential

Intuition: A participant's biomarker level is determined before the trial starts. Then their responses depend on this biomarker. The two-stage approach matches this causal structure.

---

6 Simplification 4: Guaranteed Positive Definiteness

6.1 The Problem

For a covariance matrix to be valid, it must be positive definite (PD): all eigenvalues must be positive. When constructing correlation matrices element-by-element, PD is not guaranteed.

6.1.1 Common Failure Mode

High correlations "use up" the positive definiteness budget:

$$\lambda_{\min}(\Sigma) = \sigma^2(1 - \rho_{\max})$$

When cross-correlations are added, $\lambda_{\min}$ can become negative.

6.2 Solution: Independent Construction with Schur Complement

6.2.1 The Schur Complement Condition

For the partitioned matrix to be PD:

$$\Sigma_{1|2} = \Sigma_{11} - \Sigma_{12}\Sigma_{22}^{-1}\Sigma_{21} > 0$$

This is the **Schur complement** condition.

6.2.2 Construction Strategy

1. **Build Σ_{22} (2×2)**: Always PD for $|\rho| < 1$
2. **Build Σ_{11} (24×24)**: Use time-based AR(1), guaranteed PD
3. **Build Σ_{12} (24×2)**: Regression coefficients
4. **Check Schur complement**: If not PD, scale down Σ_{12}

6.2.3 Grid-Snapping Algorithm

Rather than continuous scaling, we snap to a predefined grid of correlation values:

```
allowed_correlations <- c(0, 0.1, 0.2, 0.3, 0.4, 0.5, 0.6)
```

Algorithm:

1. User requests biomarker correlation $\rho_{\text{requested}}$
2. Try $\rho \in \{\rho_{\text{requested}}, \rho_{\text{requested}} - 0.1, \ldots, 0\}$
3. For each ρ, check if $\Sigma_{1|2}$ is PD
4. Use largest valid ρ from grid

6.2.4 Example

- Requested: $\rho = 0.5$
- System checks: 0.5 (not PD), 0.4 (PD!)
- Uses: $\rho_{\text{effective}} = 0.4$
- Reports: "Snapped biomarker correlation: 0.50 → 0.40"

6.2.5 Why Grid-Snapping Is Better Than Continuous Scaling

Aspect	Continuous	Grid-Snapping
Result values	Arbitrary (0.423…)	Clean (0.4)
Visualization	Hard to bin	Natural grid
Reproducibility	Exact but odd values	Clean categories
Interpretation	"Scaled by 0.847"	"Using 0.4"

Intuition: For reporting and visualization, we want results on a regular grid. Grid-snapping ensures this automatically.

---

7 Complete Response Model

7.1 Total Response Formula

Combining all simplifications, the response at timepoint t for participant i is:

$$Y_{it} = \underbrace{\text{baseline}_i}_{\text{from } \Sigma_{22}} \tag{9}$$

$$+ \underbrace{\text{BR}_{\text{rate}} \times \text{weeks\_on\_drug}_{it} + \text{br\_random}_{it}}_{\text{Biological Response}} \tag{10}$$

$$+ \underbrace{\text{ER}_{\text{rate}} \times \sum_{s \le t} \text{expectancy}_{is} + \text{er\_random}_{it}}_{\text{Expectancy Response}} \tag{11}$$

$$+ \underbrace{\text{TR}_{\text{rate}} \times (t - t_0) + \text{tr\_random}_{it}}_{\text{Time-variant Response}} \tag{12}$$

where:

- baseline_i: Participant's baseline (from Σ_{22}, correlated with biomarker)
- br_random_{it}, er_random_{it}, tr_random_{it}: Correlated random effects (from $\Sigma_{1|2}$)
- Carryover modifies BR when off drug

7.2 Correlation Structure

The random components ($\text{br\_random}, \text{er\_random}, \text{tr\_random}$) are correlated:

1. **Within factor, across time**: AR(1) with time-based decay
2. **Across factors, same time**: Cross-correlation $c_{\text{cf1t}} = 0.2$
3. **Across factors, different time**: Cross-correlation $c_{\text{cfct}} = 0.1$ with decay
4. **With biomarker**: BR correlated at c_{bm}, ER/TR at $0.5 \times c_{\text{bm}}$

---

8 Hybrid Trial Design

8.1 Design Structure

Week	Phase	Treatment	Expectancy	Description
4, 8	Open-label	All active	1.0	Run-in period
9	Blinded	All active	0.5	Transition
10	Blinded	Randomized	0.5	Paths 1,2 active; 3,4 placebo
11, 12	Blinded	All placebo	0.5	Washout
16	Blinded	Crossover	0.5	Paths 1,3 active; 2,4 placebo
20	Blinded	Crossover	0.5	Paths 1,3 placebo; 2,4 active

8.2 Four-Path Randomization

Participants are randomized to one of four paths, ensuring balanced treatment sequences:

Path	W10	W11-12	W16	W20
1	Active	Placebo	Active	Placebo
2	Active	Placebo	Placebo	Active
3	Placebo	Placebo	Active	Placebo
4	Placebo	Placebo	Placebo	Active

---

9 Statistical Analysis

9.1 Mixed Effects Model

The analysis model is:

$$Y_{it} = \beta_0 + \beta_1 \text{treatment}_{it} + \beta_2 \text{biomarker}_i + \beta_3 (\text{treatment} \times \text{biomarker})_{it} + \beta_4 \text{week}_t + \gamma_i + \epsilon_{it}$$

where:

- β_3: Treatment × biomarker interaction (primary outcome)
- $\gamma_i \sim \mathcal{N}(0, \sigma^2_{\text{between}})$: Random intercept
- $\epsilon_{it} \sim \mathcal{N}(0, \sigma^2_{\text{within}})$: Residual

9.2 Power Calculation

Statistical power is the probability of detecting a significant treatment × biomarker interaction:

$$\text{Power} = P(p < 0.05 | H_1 \text{ true})$$

Estimated via Monte Carlo simulation over multiple iterations.

---

10 Computational Considerations

10.1 Efficiency Comparison

Operation	Old Method	New Method
Build covariance	26×26 at once	2×2 + 24×24
PD check	Full eigendecomposition	Schur complement
Failure mode	Reject	Snap to grid
Cholesky	26×26	2×2 + 24×24
Matrix inverse	None	2×2 (trivial)

10.2 Cholesky Decomposition Cost

$$\text{Cost} \propto n^3$$

- Old: $26^3 = 17,576$
- New: $2^3 + 24^3 = 8 + 13,824 = 13,832$

Approximately 20% reduction, plus avoided failures.

---

11 Conclusions

11.1 Summary of Simplifications

1. **Rate-based response**: 3 parameters vs 9, clinically interpretable
2. **Time-based AR(1)**: Biologically plausible decay, more robust
3. **Two-stage generation**: Clear causal structure, easier debugging
4. **Grid-snapping**: Guaranteed PD, clean visualization

11.2 Trade-offs

Simplification	Gained	Lost
Rate model	Interpretability	Asymptotic saturation
AR(1)	Realism, robustness	Compound symmetry option
Two-stage	Clarity	Nothing (mathematically equivalent)
Grid-snapping	Guaranteed PD	Exact requested correlation

11.3 Recommendations

These simplifications are recommended for:

- **Teaching**: Much easier to understand
- **Development**: Easier to debug and modify
- **Production**: More robust, fewer failures

The original Gompertz/monolithic approach may still be preferred for:

- **Publication**: Exact replication of Hendrickson
- **Asymptotic effects**: When saturation is clinically meaningful

---

12 References

Hendrickson, E., et al. (2020). N-of-1 trials with multiple randomization structures for individualized treatment. *Statistics in Medicine*.

Raskind, M., et al. (2013). Pilot RCT data used for biomarker and baseline estimates. [Source for parameter values]

---

13 Appendix: R Implementation

13.1 Key Functions

```r
# Build guaranteed-PD sigma with time-based AR(1)
build_sigma_guaranteed_pd <- function(weeks, c.bm, params) {
  # Stage 1: Sigma_22 (2×2)
  # Stage 2: Sigma_11 (24×24) with AR(1)
  # Stage 3: Sigma_12 (24×2)
  # Stage 4: Grid-snap to ensure PD
}

# Two-stage data generation
generate_participant_twostage <- function(sigma_parts, idx) {
  # Stage 1: Generate (biomarker, baseline)
  x2 <- mvrnorm(1, mu = c(0, 0), Sigma = sigma_parts$Sigma_22)

  # Stage 2: Generate responses | participant vars
  mu_cond <- Sigma_12 %*% Sigma_22_inv %*% x2
  x1 <- mvrnorm(1, mu = mu_cond, Sigma = sigma_parts$Sigma_cond)
}
```

13.2 Grid-Snapping Algorithm

```r
allowed_correlations <- c(0, 0.1, 0.2, 0.3, 0.4, 0.5, 0.6)

find_valid_correlation <- function(Sigma_11, Sigma_22_inv, c.bm_requested, ...) {
  for (c.bm_try in sort(allowed[allowed <= c.bm_requested], decreasing = TRUE)) {
    # Build Sigma_12 with c.bm_try
    # Check if Schur complement is PD
    if (min_eigenvalue > 0) return(c.bm_try)
  }
  return(0)  # Fallback
}
```