Technical Differences in Implementation: Biomarker Correlation Scaling and Positive-Definite Matrix Handling

N-of-1 Trial Simulation Project

November 18, 2025

Abstract

This document provides detailed technical explanation of two implementation differences between Hendrickson et al. (2020) and our current approach: (1) biomarker correlation scaling with validity clamping, and (2) non-positive definite (non-PD) matrix handling. We explain the mathematical rationale, practical implications, and quality control benefits of these differences.

Contents

1 Biomarker Correlation Scaling and Clamping

1.1 The Problem

Hendrickson's approach:

```
# Simply uses c.bm directly in correlation matrix
R[pb_idx, br_idx] = c.bm  # e.g., 0.4
```

Our approach:

```
# Scales by mean ratio
mean_ratio = mean_br / mean_pb
scaled_correlation = c.bm * mean_ratio
# CLAMP to valid range
scaled_correlation = pmax(-0.99, pmin(0.99, scaled_correlation))
```

1.2 Mathematical Rationale

The scaling exists because **correlation matrices must preserve relationships when means differ**.

1.2.1 Correlation vs. Covariance

Correlation is scale-free, but covariance is not:

$$\text{Correlation:} \quad \rho(X,Y) = \frac{\text{Cov}(X,Y)}{\sigma_X \cdot \sigma_Y} \tag{1}$$

$$\text{Covariance:} \quad \text{Cov}(X,Y) = \rho(X,Y) \cdot \sigma_X \cdot \sigma_Y \tag{2}$$

When X and Y have different means (due to treatment effects and carryover), their covariances need adjustment to maintain the same "strength of relationship."

1.2.2 Concrete Example

Consider a timepoint where BR mean is inflated by treatment and carryover:

$$\mu_{\text{pb}} = 15.0 \quad \text{(baseline)}$$
$$\mu_{\text{br}} = 15.0 + 5.0 \text{ (treatment)} + 2.5 \text{ (carryover)} = 22.5$$

Mean ratio:

$$r = \frac{\mu_{\text{br}}}{\mu_{\text{pb}}} = \frac{22.5}{15.0} = 1.5$$

If $c_{\text{bm}} = 0.4$:

$$\rho_{\text{naive}} = 0.4$$
$$\rho_{\text{scaled}} = 0.4 \times 1.5 = 0.6$$

The scaling reflects that BR has "more room to vary" when its mean is higher, so the correlation strength needs adjustment to maintain consistent relationship across timepoints.

1.3 The Danger Without Clamping

Before we added clamping:
With $c_{\text{bm}} = 0.4$ and mean ratio $r = 2.5$ (high treatment + strong carryover):

$$\rho_{\text{scaled}} = 0.4 \times 2.5 = 1.0 \quad \textbf{INVALID!}$$

Problem: Correlations must satisfy $|\rho| < 1$. A correlation of exactly 1.0 indicates perfect linear dependence, which:

- Produces non-positive definite matrices
- Causes negative eigenvalues
- Creates mathematically invalid covariance structures

After adding clamping:
Same scenario:

$$\rho_{\text{raw}} = 0.4 \times 2.5 = 1.0$$
$$\rho_{\text{clamped}} = \min(0.99, 1.0) = 0.99 \quad \checkmark \text{ VALID}$$

This preserves matrix positive-definiteness. The value 0.99 represents "strong but not perfect correlation," and the slight margin prevents numerical instability.

1.4 Why This Matters

1. **Validity preservation**: Keeps correlation matrices mathematically valid
2. **Numerical stability**: Prevents eigenvalue computation failures
3. **Conservative approach**: When mean ratios are extreme, caps correlation at maximum valid value
4. **Quality control**: Rather than failing silently, ensures all matrices are usable

1.5 Impact After Option 1 Changes

After removing the population mean shift ($c_{\text{bm}} \times \text{tod} \times 2.0$), mean ratios became smaller:

Component	Before Option 1	After Option 1
Baseline	15.0	15.0
Treatment effect	5.0	5.0
Carryover effect	2.5	2.5
Biomarker shift	0.8	**0.0** (removed)
Total BR mean	23.3	22.5
Mean ratio	1.55	1.50

Table 1: Mean values before and after Option 1 implementation

Result: Smaller ratios ⇒ less scaling ⇒ clamping is now **mostly a safety check** rather than frequently invoked.

2 Non-Positive Definite Matrix Handling

2.1 What is Positive-Definiteness?

A covariance matrix $\mathbf{\Sigma}$ is **positive-definite** (PD) if:

$$\mathbf{v}^T\mathbf{\Sigma}\mathbf{v} > 0 \quad \forall \mathbf{v} \neq \mathbf{0}$$

Equivalently: All eigenvalues $\lambda_i > 0$.

2.1.1 Why It Matters

- Only PD matrices can represent valid covariance structures
- Multivariate normal (MVN) distribution requires PD matrix for valid probability density:

$$f(\mathbf{x}) = \frac{1}{(2\pi)^{k/2}|\mathbf{\Sigma}|^{1/2}} \exp\left(-\frac{1}{2}(\mathbf{x}-\boldsymbol{\mu})^T\mathbf{\Sigma}^{-1}(\mathbf{x}-\boldsymbol{\mu})\right)$$

- Non-PD $\Rightarrow$ some linear combinations of variables have negative variance (impossible!)

2.2 Hendrickson's Approach: Auto-Fix

```r
# Check if matrix is PD
eigenvalues = eigen(Sigma)$values
is_pd = all(eigenvalues > 0)

if (!is_pd) {
  # Automatically "fix" by adjusting eigenvalues
  Sigma = corpcor::make.positive.definite(Sigma, tol = 1e-3)
  # This finds the "nearest" PD matrix
}

return(Sigma)  # Always returns something
```

2.2.1 What `make.positive.definite()` Does

1. Compute eigendecomposition: $\mathbf{\Sigma} = \mathbf{Q}\mathbf{\Lambda}\mathbf{Q}^T$
2. Find negative eigenvalues: $\lambda_i < 0$
3. Replace them: $\lambda_i \to \max(\lambda_i, \tau)$ where $\tau = 10^{-3}$
4. Reconstruct: $\mathbf{\Sigma}_{\text{fixed}} = \mathbf{Q}\mathbf{\Lambda}_{\text{fixed}}\mathbf{Q}^T$

Effect: You always get a "valid" matrix, but it's **not the matrix you asked for**—it's been perturbed.

2.3 Our Approach: Reject Invalid

```
# Check if matrix is PD
eigenvalues = eigen(Sigma)$values
is_pd = all(eigenvalues > 0)

if (!is_pd) {
  cat("REJECTED: Non-positive definite matrix detected.\n")
  cat("Eigenvalue range:", min(eigenvalues), "to",
      max(eigenvalues), "\n")
  cat("Parameter combination:", params, "\n")
  return(NULL)  # Signal: skip this combination
}

return(Sigma)  # Only return if truly valid
```

Effect: Simulation skips invalid parameter combinations entirely.

2.4 Why This Difference Matters

Philosophical question: What should you do when parameters produce invalid matrices?

2.4.1 Option A: Fix and Continue (Hendrickson)

Pros:

- Never lose data, complete parameter grid
- Computational efficiency (no failed runs)

Cons:

- Using a *different* matrix than intended
- Masks potential parameter issues
- "Fixing" might introduce bias

2.4.2 Option B: Reject and Skip (Our Approach)

Pros:

- Only use mathematically correct matrices
- Reveals problematic parameter combinations
- No silent perturbations

Cons:

- Lose some parameter combinations
- Need to ensure grid has valid combinations

2.5 When Does Non-PD Occur?

2.5.1 1. High Correlations

When multiple correlations near 1.0 create impossible relationships:

```
# Example: Can't have all three simultaneously
Cor(X,Y) = 0.95
Cor(Y,Z) = 0.95
Cor(X,Z) = 0.10  # Inconsistent!
# X and Z should be correlated through Y
```

If X and Y are highly correlated, and Y and Z are highly correlated, then X and Z *must* also be correlated. A low correlation between X and Z creates an inconsistent constraint system.

2.5.2 2. Scaling Issues

When mean-ratio scaling pushes correlations too high:

```
# Before clamping fix
c.bm = 0.4
mean_ratio = 2.6
# Result:
correlation = 0.4 * 2.6 = 1.04  # Invalid!
# Produces non-PD matrix
```

2.5.3 3. Autocorrelation + Cross-Correlation Conflicts

If:

- Autocorrelations are very high (e.g., $c_{\text{tv}} = c_{\text{pb}} = c_{\text{br}} = 0.8$)
- AND cross-correlations are substantial ($c_{\text{cf1t}} = 0.2$, $c_{\text{cfct}} = 0.1$)
- AND biomarker correlation varies by timepoint (on/off treatment)

This can create impossible constraint systems, leading to non-PD matrices.

2.6 Practical Impact in Our Simulations

With fixed Hendrickson correlations + clamping:

✓ All 12 parameter combinations produce valid (PD) matrices

✓ No rejections in current parameter grid

✓ Eigenvalue ranges all positive

✓ Condition numbers acceptable (< 100)

So our rejection mechanism is currently a **quality assurance check** rather than frequently invoked.

2.7 Quality Control Benefits

Our approach provides **diagnostic information**:

```
# When rejection occurs, we print:
REJECTED: Non-positive definite matrix detected.
Eigenvalue range: -0.0234 to 2.156
Condition number: 847.3 (ill-conditioned!)

Parameter combination:
  c.bm = 0.8
  carryover_t1half = 2.0
  biomarker_correlation = 0.8

# This tells us: "These parameters create
#   inconsistent correlation structure"
# Action: Revise parameters or investigate why
```

Hendrickson's auto-fix would silently continue with a perturbed matrix—we'd never know there was a problem.

3 Comparison Summary

Aspect	Hendrickson	Our Approach	Trade-off
Scaling	No scaling, use c_{bm} directly	Scale by mean ratio, clamp to $[-0.99, 0.99]$	Complexity vs. validity
Philosophy	Trust parameters	Validate parameters	Permissive vs. strict
Non-PD handling	Auto-fix and continue	Reject and skip	Completeness vs. correctness
Diagnostics	Silent	Explicit warnings	Convenience vs. transparency
Result	Always complete grid	Only valid combinations	Quantity vs. quality

Table 2: Comparison of approaches

3.1 Which Approach is "Better"?

It depends on the goal:

3.1.1 For Exploratory Simulation (Hendrickson's Goal)

- Wide parameter sweep
- Want complete results
- Auto-fix acceptable
- Published methodology

3.1.2 For Methodological Validation (Our Goal)

- Strict adherence to assumptions
- Only mathematically valid matrices
- Diagnostic transparency
- Quality over quantity

Both approaches are defensible. We chose the stricter approach because:

1. We're extending published methodology (need rigor)
2. Current parameter grid produces all valid matrices anyway
3. Provides quality assurance for future parameters
4. Diagnostic output helps identify issues

4 Practical Recommendations

4.1 When to Use Scaling/Clamping

Always use it when:

- Means vary across timepoints (treatment effects, carryover)
- You want correlation "strength" to remain consistent
- You need to prevent invalid correlations

Monitor for:

- Frequent clamping at 0.99 $\Rightarrow$ parameters may be too extreme
- Mean ratios $> 2.0 \Rightarrow$ consider parameter adjustment

4.2 When to Reject Non-PD Matrices

Reject when:

- Methodological rigor required
- Parameter validation needed
- Publishing results (need defensibility)

Auto-fix when:

- Exploratory analysis
- Computational constraints (can't afford rejections)
- Parameters known to be near boundary

4.3 Our Current Status

After Option 1 alignment:

- ✓ Mean ratios are modest (< 1.5 typically)
- ✓ Clamping rarely invoked
- ✓ All matrices PD without rejection
- ✓ Both mechanisms serving as quality checks

5 Conclusion

These differences reflect **stricter quality control** in our implementation, ensuring we only use mathematically valid, correctly specified covariance structures. While Hendrickson's approach prioritizes computational completeness (always obtaining results), our approach prioritizes mathematical correctness (only using valid matrices).

Both philosophies are valid for their respective goals. For our purpose of extending and validating Hendrickson's methodology with novel carryover modeling, the stricter approach provides:

1. Confidence that all results are mathematically sound
2. Transparency about parameter validity
3. Early detection of problematic parameter combinations
4. Documentation trail for methodological rigor

The practical impact is minimal in our current simulations (all parameters produce valid matrices), but the infrastructure ensures quality control as we extend to new parameter ranges or designs.

Documentation created: November 18, 2025
Repository: `/Users/zenn/Dropbox/prj/d08/pmsimstats2025`