# Mathematical Derivation of the Two-Step Sigma Matrix Calculation

Clinical Trial Simulation Documentation

December 10, 2025

**Abstract**

This document provides a comprehensive mathematical derivation of the efficient two-step method for computing covariance matrices from correlation matrices in multivariate normal data generation. Specifically, we derive the identity $\Sigma = \mathbf{DRD}$, where $\Sigma$ is the covariance matrix, $\mathbf{R}$ is the correlation matrix, and $\mathbf{D}$ is a diagonal matrix of standard deviations. This decomposition is fundamental to clinical trial simulation, enabling numerical stability and computational efficiency.

# Contents

# 1 Introduction

In multivariate normal data generation for clinical trials, we need to construct covariance matrices that are:

1. **Positive definite** (PD) - ensuring mvrnorm() and mvn sampling work correctly
2. **Numerically stable** - avoiding ill-conditioning and eigenvalue issues
3. **Interpretable** - separating correlation structure from magnitude (standard deviations)

The most efficient approach achieves all three by decomposing the covariance matrix into:

$$\Sigma = \mathbf{DRD}$$

where the correlation matrix $\mathbf{R}$ is validated for PD independently from scaling by standard deviations in $\mathbf{D}$.

This document derives this identity from first principles and explains its computational advantages.

# 2 Foundational Concepts

## 2.1 Covariance and Correlation: Univariate Case

**Definition 1** (Covariance)**.** For two random variables $X$ and $Y$ with means $\mu_X$ and $\mu_Y$:

$$\text{Cov}(X, Y) = E[(X - \mu_X)(Y - \mu_Y)]$$

**Definition 2** (Standard Deviation)**.** The standard deviation of a random variable $X$ is:

$$\sigma_X = \sqrt{\text{Var}(X)} = \sqrt{\text{Cov}(X, X)}$$

**Definition 3** (Correlation Coefficient)**.** The Pearson correlation coefficient between $X$ and $Y$ is:

$$\text{Corr}(X, Y) = \frac{\text{Cov}(X, Y)}{\sigma_X \sigma_Y}$$

**Theorem 1** (Covariance-Correlation Relationship)**.** For any two random variables $X$ and $Y$:

$$\text{Cov}(X, Y) = \text{Corr}(X, Y) \cdot \sigma_X \cdot \sigma_Y$$

*Proof.* By definition of correlation (Definition 3):

$$\text{Corr}(X, Y) = \frac{\text{Cov}(X, Y)}{\sigma_X \sigma_Y}$$

Rearranging:

$$\text{Cov}(X, Y) = \text{Corr}(X, Y) \cdot \sigma_X \cdot \sigma_Y \quad \square$$

$\square$

# 3 Multivariate Extension

## 3.1 Covariance Matrix

**Definition 4** (Covariance Matrix)**.** Let $\mathbf{Z} = [Z_1, Z_2, \ldots, Z_n]^\top$ be a random vector with mean $\boldsymbol{\mu}$. The covariance matrix is:

$$\Sigma = \mathrm{Cov}(\mathbf{Z}) = E[(\mathbf{Z} - \boldsymbol{\mu})(\mathbf{Z} - \boldsymbol{\mu})^\top]$$

with elements:

$$\Sigma_{ij} = \mathrm{Cov}(Z_i, Z_j) = E[(Z_i - \mu_i)(Z_j - \mu_j)]$$

In particular, $\Sigma_{ii} = \mathrm{Var}(Z_i) = \sigma_i^2$.

## 3.2 Correlation Matrix

**Definition 5** (Correlation Matrix)**.** The correlation matrix $\mathbf{R}$ has elements:

$$R_{ij} = \frac{\Sigma_{ij}}{\sigma_i \sigma_j} = \mathrm{Corr}(Z_i, Z_j)$$

By definition, $R_{ii} = 1$ (perfect self-correlation) and $|R_{ij}| \leq 1$ for $i \neq j$.

## 3.3 Positive Definiteness

**Definition 6** (Positive Definite Matrix)**.** A symmetric matrix $\mathbf{A} \in \mathbb{R}^{n \times n}$ is positive definite (PD) if:

$$\mathbf{x}^\top \mathbf{A} \mathbf{x} > 0 \quad \forall \mathbf{x} \in \mathbb{R}^n, \mathbf{x} \neq \mathbf{0}$$

Equivalently, all eigenvalues are strictly positive.

**Remark 1** (Valid Correlation Matrices)**.** A valid correlation matrix $\mathbf{R}$ must be:

1. Symmetric: $R_{ij} = R_{ji}$
2. Unit diagonal: $R_{ii} = 1$
3. Bounded: $|R_{ij}| \leq 1$
4. Positive semi-definite (PSD)

Note: Conditions 1-3 are necessary but NOT sufficient for PSD.

# 4 The Sigma Matrix Decomposition: $= \mathbf{D} \cdot \mathbf{R} \cdot \mathbf{D}$

## 4.1 Defining the Diagonal Standard Deviation Matrix

**Definition 7** (Diagonal Standard Deviation Matrix)**.** Define $\mathbf{D}$ as the diagonal matrix of standard deviations:

$$\mathbf{D} = \begin{pmatrix} \sigma_1 & 0 & \cdots & 0 \\ 0 & \sigma_2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \sigma_n \end{pmatrix}$$

where $\sigma_i = \sqrt{\Sigma_{ii}}$ is the standard deviation of $Z_i$.

## 4.2 Main Derivation

**Theorem 2** (Sigma Decomposition)**.** The covariance matrix $\Sigma$ can be decomposed as:

$$\Sigma = \mathbf{DRD}$$

where $\mathbf{D}$ is the diagonal standard deviation matrix and $\mathbf{R}$ is the correlation matrix.

*Proof.* We show this by computing each element of $\mathbf{DRD}$.

**Step 1: Compute DR**

The $(i, j)$-element of $\mathbf{DR}$ is:

$$[\mathbf{DR}]_{ij} = \sum_{k=1}^{n} D_{ik} R_{kj} = D_{ii} R_{ij} = \sigma_i R_{ij}$$

because $\mathbf{D}$ is diagonal (so $D_{ik} = 0$ for $k \neq i$).

**Step 2: Compute (DR)D**

The $(i, j)$-element of $(\mathbf{DR})\mathbf{D}$ is:

$$[(\mathbf{DR})\mathbf{D}]_{ij} = \sum_{k=1}^{n} [\mathbf{DR}]_{ik} D_{kj} = (\sigma_i R_{ij}) D_{jj} = \sigma_i R_{ij} \sigma_j$$

**Step 3: Express in terms of covariance**

From Definition 2 and Theorem 1:

$$R_{ij} = \frac{\Sigma_{ij}}{\sigma_i \sigma_j}$$

Therefore:

$$[(\mathbf{DR})\mathbf{D}]_{ij} = \sigma_i \cdot \frac{\Sigma_{ij}}{\sigma_i \sigma_j} \cdot \sigma_j = \Sigma_{ij}$$

Since this holds for all elements $(i, j)$:

$$\mathbf{DRD} = \Sigma \quad \square$$

□

Key Identity

$$\Sigma = \mathbf{DRD}$$

## 4.3 Element-Wise Illustration

For a $2 \times 2$ case, the decomposition is explicit:

$$\begin{pmatrix} \sigma_1 & 0 \\ 0 & \sigma_2 \end{pmatrix} \begin{pmatrix} 1 & R_{12} \\ R_{12} & 1 \end{pmatrix} \begin{pmatrix} \sigma_1 & 0 \\ 0 & \sigma_2 \end{pmatrix} = \begin{pmatrix} \sigma_1^2 & \sigma_1 R_{12} \sigma_2 \\ \sigma_1 R_{12} \sigma_2 & \sigma_2^2 \end{pmatrix}$$

where the diagonal elements $\sigma_i^2$ are the variances and the off-diagonal elements $\sigma_i R_{ij} \sigma_j$ are the covariances.

# 5 Key Properties

## 5.1 Preservation of Positive Definiteness

**Lemma 1** (PD Preservation)**.** If $\mathbf{R}$ is positive definite and $\mathbf{D}$ is diagonal with all positive diagonal entries, then $\Sigma = \mathbf{DRD}$ is positive definite.

*Proof.* For any non-zero vector $\mathbf{x}$:

$$\mathbf{x}^\top \Sigma \mathbf{x} = \mathbf{x}^\top (\mathbf{DRD})\mathbf{x} = (\mathbf{Dx})^\top \mathbf{R}(\mathbf{Dx})$$

Let $\mathbf{y} = \mathbf{Dx}$. Since $\mathbf{D}$ has positive diagonal entries and $\mathbf{x} \neq \mathbf{0}$, we have $\mathbf{y} \neq \mathbf{0}$. Since $\mathbf{R}$ is positive definite:

$$\mathbf{y}^\top \mathbf{R}\mathbf{y} > 0$$

Therefore:

$$\mathbf{x}^\top \Sigma \mathbf{x} > 0 \quad \square$$

$\square$

**Corollary 1** (Validity of Derived Covariance)**.** If the correlation matrix $\mathbf{R}$ is positive definite, then the covariance matrix $\Sigma = \mathbf{DRD}$ is automatically positive definite, regardless of the values in $\mathbf{D}$ (provided all $\sigma_i > 0$).

## 5.2 Symmetry Preservation

**Lemma 2** (Symmetry Preservation)**.** If $\mathbf{R}$ and $\mathbf{D}$ are symmetric, then $\Sigma = \mathbf{DRD}$ is symmetric.

*Proof.*

$$\Sigma^\top = (\mathbf{DRD})^\top = \mathbf{D}^\top \mathbf{R}^\top \mathbf{D}^\top = \mathbf{DRD} = \Sigma \quad \square$$

$\square$

# 6 Computational Implementation

## 6.1 R Implementation

In R, the efficient computation is:

```
sigma <- outer(standard_deviations, standard_deviations) * correlations
```

where:

- `standard_deviations` is a vector of $\sigma_1, \ldots, \sigma_n$
- `correlations` is the matrix $\mathbf{R}$
- `outer(x, y)` computes the outer product: $\mathbf{x}\mathbf{y}^\top$
- `*` is element-wise multiplication

## 6.2 Explicit Computation

The `outer()` function computes:

$$\text{outer}(\sigma, \sigma) = \begin{pmatrix} \sigma_1\sigma_1 & \sigma_1\sigma_2 & \cdots & \sigma_1\sigma_n \\ \sigma_2\sigma_1 & \sigma_2\sigma_2 & \cdots & \sigma_2\sigma_n \\ \vdots & \vdots & \ddots & \vdots \\ \sigma_n\sigma_1 & \sigma_n\sigma_2 & \cdots & \sigma_n\sigma_n \end{pmatrix} = \mathbf{D}^2$$

where $\mathbf{D}^2$ denotes the matrix of all pairwise products.
Then, element-wise multiplication by $\mathbf{R}$:

$$\Sigma_{ij} = (\sigma_i\sigma_j) \times R_{ij} = \sigma_i R_{ij} \sigma_j$$

which is exactly the $(i, j)$-element of $\mathbf{DRD}$.

## 6.3 Computational Advantages

This two-step approach provides several benefits:

1. **Numerical Stability**: Correlation matrices have entries bounded in $[-1, 1]$, avoiding overflow/underflow issues when combined with scaling.
2. **PD Validation Efficiency**: Only $\mathbf{R}$ needs to be checked for positive definiteness. Once $\mathbf{R}$ is validated (Lemma 1), $\Sigma$ is automatically PD regardless of $\sigma_i$ values.
3. **Parameter Separation**: Correlation structure (affecting relationships) and magnitude (affecting variance) are computed independently, simplifying interpretation and parameter selection.
4. **Memory Efficiency**: The computation $\texttt{outer}(\sigma, \sigma) \times \mathbf{R}$ avoids explicit construction of the $n \times n$ matrix $\mathbf{D}$.
5. **Reusability**: A single correlation matrix $\mathbf{R}$ can be scaled with different $\mathbf{D}$ matrices without recomputation.
6. **Dimension Independence**: For fixed $\mathbf{R}$, increasing $n$ (number of variables) only requires updating the outer product computation.

# 7 Application to Clinical Trial Simulation

## 7.1 General Framework

In the N-of-1 trial simulation, the complete process is:

1. Define correlation parameters following Hendrickson et al.: $c_{tv}, c_{pb}, c_{br}, c_{cf1t}, c_{cfct}, c_{bm}$
2. Build the correlation matrix $\mathbf{R}$ from these parameters (with dimension $\approx 62$ for 20 timepoints $\times$ 3 factors + 2 baseline variables)

3. Validate $\mathbf{R}$ is positive definite (eigenvalue check)
4. Define standard deviations $\sigma_1, \ldots, \sigma_n$ from response and baseline parameters
5. Compute $\Sigma = \mathbf{DRD}$ using the efficient two-step method
6. Sample from $\mathbf{Z} \sim \mathcal{N}(\boldsymbol{\mu}, \Sigma)$ using mvrnorm()

### 7.2 Why This Matters

The decomposition $\Sigma = \mathbf{DRD}$ is critical because:

- High-dimensional correlation matrices (62×62) are at risk of numerical failure
- Separating $\mathbf{R}$ and $\mathbf{D}$ isolates the PD problem to $\mathbf{R}$ only
- This allows systematic validation of correlation structures independently of scales
- The final $\Sigma$ inherits PD guarantee from $\mathbf{R}$ via Lemma 1

## 8 Example: 3-Variable Case

For illustration, consider three biomarker components at one timepoint:

$$\mathbf{R} = \begin{pmatrix} 1.00 & 0.20 & 0.15 \\ 0.20 & 1.00 & 0.10 \\ 0.15 & 0.10 & 1.00 \end{pmatrix}$$

with standard deviations $\sigma_1 = 2.0, \sigma_2 = 1.5, \sigma_3 = 2.5$.
The covariance matrix is:

$$\Sigma = \begin{pmatrix} 2.0 & 0 & 0 \\ 0 & 1.5 & 0 \\ 0 & 0 & 2.5 \end{pmatrix} \begin{pmatrix} 1.00 & 0.20 & 0.15 \\ 0.20 & 1.00 & 0.10 \\ 0.15 & 0.10 & 1.00 \end{pmatrix} \begin{pmatrix} 2.0 & 0 & 0 \\ 0 & 1.5 & 0 \\ 0 & 0 & 2.5 \end{pmatrix}$$

Computing element (1,2):

$$\Sigma_{12} = 2.0 \times 0.20 \times 1.5 = 0.60$$

and the full covariance matrix:

$$\Sigma = \begin{pmatrix} 4.00 & 0.60 & 0.75 \\ 0.60 & 2.25 & 0.375 \\ 0.75 & 0.375 & 6.25 \end{pmatrix}$$

Note that the diagonal elements are $\sigma_i^2$:

$$\Sigma_{11} = 2.0^2 = 4.00$$
$$\Sigma_{22} = 1.5^2 = 2.25$$
$$\Sigma_{33} = 2.5^2 = 6.25$$

# 9 Block Partitioning: From 26×26 to 2×2 Inversion

## 9.1 Motivation: Curse of Dimensionality

In clinical trial simulation with 8 measurement timepoints and 3 response factors plus 2 baseline variables, the full covariance matrix is:

$$\text{Dimension} = 2 \text{ (baseline)} + 3 \times 8 \text{ (responses)} = 26 \times 26$$

A naive approach would:

1. Construct the full 26×26 covariance matrix $\Sigma$
2. Compute its Cholesky decomposition $\Sigma = \mathbf{L}\mathbf{L}^\top$
3. Sample via $\mathbf{x} = \boldsymbol{\mu} + \mathbf{L}\mathbf{z}$

**Problem**: Inverting or decomposing a 26×26 matrix has:

- Condition number $\kappa(\Sigma) \approx 1000$ (ill-conditioned)
- Computational cost $O(n^3) = O(26^3) \approx 17,576$ operations
- Numerical instability from floating-point accumulation

The solution: **Partition the problem and use conditional sampling.**

## 9.2 Block Matrix Partitioning

Reorder the variables as:

- $\mathbf{X}_1 = [\text{BR}_1, \ldots, \text{BR}_8, \text{ER}_1, \ldots, \text{ER}_8, \text{TR}_1, \ldots, \text{TR}_8]^\top$ (24-dimensional: all responses)
- $\mathbf{X}_2 = [\text{Biomarker}, \text{Baseline}]^\top$ (2-dimensional: baseline variables)

Then partition the covariance matrix:

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

where:

$$\begin{aligned} &\Sigma_{11} \in \mathbb{R}^{24\times 24} \quad \text{(covariance of responses)} \\ &\Sigma_{22} \in \mathbb{R}^{2\times 2} \quad \text{(covariance of baseline)} \\ &\Sigma_{12} \in \mathbb{R}^{24\times 2} \quad \text{(cross-covariance: responses vs baseline)} \\ &\Sigma_{21} = \Sigma_{12}^\top \in \mathbb{R}^{2\times 24} \quad \text{(by symmetry)} \end{aligned}$$

### 9.2.1 Explicit Block Structure

$$\Sigma_{26\times 26} = \begin{pmatrix} \Sigma_{BR,BR}^{8\times 8} & \Sigma_{BR,ER}^{8\times 8} & \Sigma_{BR,TR}^{8\times 8} & \Sigma_{BR,BM}^{8\times 1} & \Sigma_{BR,BL}^{8\times 1} \\ \Sigma_{ER,BR}^{8\times 8} & \Sigma_{ER,ER}^{8\times 8} & \Sigma_{ER,TR}^{8\times 8} & \Sigma_{ER,BM}^{8\times 1} & \Sigma_{ER,BL}^{8\times 1} \\ \Sigma_{TR,BR}^{8\times 8} & \Sigma_{TR,ER}^{8\times 8} & \Sigma_{TR,TR}^{8\times 8} & \Sigma_{TR,BM}^{8\times 1} & \Sigma_{TR,BL}^{8\times 1} \\ \Sigma_{BM,BR}^{1\times 8} & \Sigma_{BM,ER}^{1\times 8} & \Sigma_{BM,TR}^{1\times 8} & \Sigma_{BM,BM}^{1\times 1} & \Sigma_{BM,BL}^{1\times 1} \\ \Sigma_{BL,BR}^{1\times 8} & \Sigma_{BL,ER}^{1\times 8} & \Sigma_{BL,TR}^{1\times 8} & \Sigma_{BL,BM}^{1\times 1} & \Sigma_{BL,BL}^{1\times 1} \end{pmatrix}$$

In block form:

$$\Sigma = \begin{pmatrix} \Sigma_{11}^{24\times 24} & \Sigma_{12}^{24\times 2} \\ \Sigma_{21}^{2\times 24} & \Sigma_{22}^{2\times 2} \end{pmatrix}$$

## 9.3 Construction of $\Sigma_{22}$ (2×2)

The baseline covariance matrix:

$$\Sigma_{22} = \begin{pmatrix} \sigma_{\text{BM}}^2 & \rho_{\text{BM,BL}}\sigma_{\text{BM}}\sigma_{\text{BL}} \\ \rho_{\text{BM,BL}}\sigma_{\text{BM}}\sigma_{\text{BL}} & \sigma_{\text{BL}}^2 \end{pmatrix}$$

where:

- $\sigma_{\text{BM}}$ = standard deviation of biomarker
- $\sigma_{\text{BL}}$ = standard deviation of baseline
- $\rho_{\text{BM,BL}}$ = correlation between biomarker and baseline (typically $\approx 0.3$)

**Key property**: $\Sigma_{22}$ is a **small, well-conditioned 2×2 matrix** with:

- Condition number $\kappa(\Sigma_{22}) \approx 10$ (well-conditioned)
- Inversion cost: $O(2^3) = O(8)$ (trivial)
- Determinant: $|\Sigma_{22}| = \sigma_{\text{BM}}^2\sigma_{\text{BL}}^2(1 - \rho_{\text{BM,BL}}^2) > 0$ (always PD)

## 9.4 Construction of $\Sigma_{11}$ (24×24)

The response covariance matrix consists of three 8×8 blocks:

$$\Sigma_{11} = \begin{pmatrix} \Sigma_{\text{BR}}^{8\times 8} & \Sigma_{\text{BR,ER}}^{8\times 8} & \Sigma_{\text{BR,TR}}^{8\times 8} \\ \Sigma_{\text{ER,BR}}^{8\times 8} & \Sigma_{\text{ER}}^{8\times 8} & \Sigma_{\text{ER,TR}}^{8\times 8} \\ \Sigma_{\text{TR,BR}}^{8\times 8} & \Sigma_{\text{TR,ER}}^{8\times 8} & \Sigma_{\text{TR}}^{8\times 8} \end{pmatrix}$$

### 9.4.1 Within-Factor Blocks (AR(1) Structure)

Each diagonal block (e.g., $\Sigma_{\text{BR}}$) is an autoregressive correlation with temporal lag:

$$\Sigma_{\text{BR}}[i,j] = \sigma_{\text{within}}^2 \cdot \rho_{\text{BR}}^{|t_i - t_j|}$$

where:

- $\rho_{\text{BR}} = 0.8$ (autocorrelation from Hendrickson parameters)

- $|t_i - t_j|$ = absolute time difference in weeks
- For example, with weeks $[4, 8, 9, 10, 11, 12, 16, 20]$:
  - Timepoints 1 and 2: lag = 4 weeks, $\rho = 0.8^4 = 0.4096$
  - Timepoints 2 and 3: lag = 1 week, $\rho = 0.8^1 = 0.8$

### 9.4.2 Cross-Factor Blocks

Off-diagonal blocks represent correlations between different response factors at the same or different timepoints:

$$\Sigma_{\text{BR,ER}}[i,j] = \begin{cases} \rho_{\text{cf1t}}\sigma^2_{\text{within}} & \text{if } i = j \text{ (same timepoint)} \\ \rho_{\text{cfct}}\sigma^2_{\text{within}} \cdot (0.9)^{|t_i - t_j|} & \text{if } i \neq j \text{ (different timepoints)} \end{cases}$$

where $\rho_{\text{cf1t}} = 0.2$ (same-time cross-correlation) and $\rho_{\text{cfct}} = 0.1$ (different-time cross-correlation).

## 9.5 Construction of $\Sigma_{12}$ (24×2)

The cross-covariance between responses and baseline:

$$\Sigma_{12} = \begin{pmatrix} \Sigma^{8\times1}_{\text{BR,BM}} & \Sigma^{8\times1}_{\text{BR,BL}} \\ \Sigma^{8\times1}_{\text{ER,BM}} & \Sigma^{8\times1}_{\text{ER,BL}} \\ \Sigma^{8\times1}_{\text{TR,BM}} & \Sigma^{8\times1}_{\text{TR,BL}} \end{pmatrix}$$

Each row specifies how one response at one timepoint correlates with baseline variables:

$$\begin{aligned} \Sigma_{\text{BR,BM}}[i] &= \rho_{\text{BM}}\sigma_{\text{within}}\sigma_{\text{BM}} \quad \text{(full biomarker correlation)} \\ \Sigma_{\text{ER,BM}}[i] &= 0.5 \cdot \rho_{\text{BM}}\sigma_{\text{within}}\sigma_{\text{BM}} \quad \text{(reduced for expectancy effect)} \\ \Sigma_{\text{TR,BM}}[i] &= 0.5 \cdot \rho_{\text{BM}}\sigma_{\text{within}}\sigma_{\text{BM}} \quad \text{(reduced for time-variant)} \end{aligned}$$

where $\rho_{\text{BM}} \in [0, 0.4]$ is the biomarker moderation strength (parameter being swept).

## 9.6 The Conditional Normal Identity

**Theorem 3** (Conditional Multivariate Normal)**.** Given a bivariate normal:

$$\begin{pmatrix} \mathbf{X}_1 \\ \mathbf{X}_2 \end{pmatrix} \sim \mathcal{N}\left( \begin{pmatrix} \boldsymbol{\mu}_1 \\ \boldsymbol{\mu}_2 \end{pmatrix}, \begin{pmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{21} & \Sigma_{22} \end{pmatrix} \right)$$

The conditional distribution of $\mathbf{X}_1$ given $\mathbf{X}_2 = \mathbf{x}_2$ is:

$$\mathbf{X}_1 | \mathbf{X}_2 = \mathbf{x}_2 \sim \mathcal{N}(\boldsymbol{\mu}_{1|2}, \Sigma_{1|2})$$

where:

$$\boldsymbol{\mu}_{1|2} = \boldsymbol{\mu}_1 + \Sigma_{12}\Sigma_{22}^{-1}(\mathbf{x}_2 - \boldsymbol{\mu}_2) \tag{1}$$

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

1. **Stage 1**: Sample baseline variables from the marginal distribution

$$\mathbf{x}_2 \sim \mathcal{N}(\boldsymbol{\mu}_2, \Sigma_{22})$$

   This requires Cholesky decomposition of the small 2×2 matrix: $\Sigma_{22} = \mathbf{L}_{22}\mathbf{L}_{22}^\top$

2. **Stage 2**: Compute the conditional mean and covariance

$$\boldsymbol{\mu}_{1|2} = \boldsymbol{\mu}_1 + \Sigma_{12}\Sigma_{22}^{-1}(\mathbf{x}_2 - \boldsymbol{\mu}_2)$$
$$\Sigma_{1|2} = \Sigma_{11} - \Sigma_{12}\Sigma_{22}^{-1}\Sigma_{21}^\top$$

3. **Stage 2 continued**: Sample responses from the conditional distribution

$$\mathbf{x}_1 \sim \mathcal{N}(\boldsymbol{\mu}_{1|2}, \Sigma_{1|2})$$

   This requires Cholesky decomposition of the 24×24 conditional covariance

4. **Combine**: Return $[\mathbf{x}_1, \mathbf{x}_2]^\top$ as the participant's data

*Sketch.* The conditional distribution formula comes from completing the square in the joint MVN density. The full derivation shows that conditioning on $\mathbf{X}_2$ introduces a linear shift in the mean (Equation 1) and reduces the covariance by the amount of variance "explained" by $\mathbf{X}_2$ (Equation 2).

The key insight: **only $\Sigma_{22}^{-1}$ appears in both expressions**—no inversion of the large matrices is needed. □

## 9.7 Computational Advantage

| Operation | Naive (26×26) | Partitioned (2×2) | Speedup |
|---|---|---|---|
| Matrix inversion | $O(26^3) \approx 17,576$ | $O(2^3) = 8$ | $\approx 2,200\times$ |
| Condition number | $\kappa \approx 1,000$ | $\kappa \approx 10$ | $\approx 100\times$ better |
| Memory storage | $26^2 = 676$ elements | $24^2 + 2^2 + 24 \times 2 = 626$ | Similar |
| Cholesky stability | High risk of failure | Highly reliable | Crucial |

Table 1: Computational savings from block partitioning

## 9.8 Two-Stage Sampling Algorithm

Given the block partition, sampling is performed in two stages:

**Remark 2** (Why This Works)**.** The two-stage approach is **mathematically equivalent** to sampling directly from the full 26×26 MVN, but computationally superior because:

1. Only the 2×2 baseline covariance needs to be inverted

2. The conditional covariance $\Sigma_{1|2}$ is built to be PD automatically (derived from the block structure)
3. Cholesky decompositions are applied to small, well-conditioned matrices
4. Numerical errors accumulate much more slowly

This is not an approximation—it is the exact conditional distribution of the multivariate normal.

# 10 Conclusion

The identity $\Sigma = \mathbf{DRD}$ is a fundamental decomposition that:

1. **Theoretically** cleanly separates correlation structure from variance scaling
2. **Computationally** enables efficient two-step calculation and validation
3. **Numerically** ensures positive definiteness when $\mathbf{R}$ is PD
4. **Practically** allows systematic parameter selection and validation before expensive simulations

In the context of N-of-1 trial simulation, this decomposition enables robust construction of high-dimensional covariance matrices while maintaining strict control over positive definiteness constraints.

# A Matrix Algebra Reference

## A.1 Outer Product

For vectors $\mathbf{u}, \mathbf{v} \in \mathbb{R}^n$:

$$\text{outer}(\mathbf{u}, \mathbf{v}) = \mathbf{u}\mathbf{v}^\top$$

## A.2 Diagonal Matrix Properties

For diagonal matrix $\mathbf{D}$ with diagonal entries $d_1, \ldots, d_n$ and any matrix $\mathbf{A}$:

$$(\mathbf{DA})_{ij} = d_i A_{ij}$$
$$(\mathbf{AD})_{ij} = A_{ij} d_j$$

## A.3 Matrix Transpose Properties

$$(\mathbf{AB})^\top = \mathbf{B}^\top \mathbf{A}^\top$$
$$(\mathbf{A}^\top)^\top = \mathbf{A}$$
$$(\mathbf{D}^\top) = \mathbf{D} \quad \text{(if } \mathbf{D} \text{ is diagonal)}$$

## A.4 Eigenvalue Properties

For positive definite matrix $\mathbf{A}$:

- All eigenvalues are strictly positive: $\lambda_i > 0$
- Condition number: $\kappa(\mathbf{A}) = \frac{\lambda_{\max}}{\lambda_{\min}}$
- Matrix is invertible: $\det(\mathbf{A}) > 0$