Bayesian Sample Size Determination for Analysis of Covariance in Randomised Controlled Trials Incorporating Historical Data

David SWQ Jones, Haiyan Zheng

BAYES 2026, 21-23 October 2026







Why adjust for covariates?

Analysis of covariance is shown to increase precision for parameters of interest.

Provides smaller confidence intervals resulting in more powerful tests under the frequentish framework.

For sample size determination matched to this controlling size and power, the total sample size is reduced by a factor of $(1 - \rho^2)$. Borm et al. [2007]

Our Set Up

Consider an RCT measuring outcome Y and prognostic baseline covariate X, let

$$\boldsymbol{Z}_{ij} = \begin{pmatrix} X_{ij} \\ Y_{ij} \end{pmatrix} \Bigg| \boldsymbol{\mu}_i, \boldsymbol{\Sigma} \sim \mathcal{N}_2(\boldsymbol{\mu}_i, \boldsymbol{\Sigma}) \tag{1}$$

i denotes treatment arm, $i = E, C$, j denotes patient in treatment arm, $j = 1, ..., n_i$.
Set prior

$$\boldsymbol{\mu}_i \sim \mathcal{N}_2\Big(\boldsymbol{m}_{0i}, \boldsymbol{\Lambda}_{0i} = \boldsymbol{\kappa}_{0i}^{-1/2} \boldsymbol{\Sigma} (\boldsymbol{\kappa}_{0i}^{T})^{-1/2}\Big) \tag{2}$$

When the covariance matrix is known

$$\boldsymbol{\Sigma} = \begin{pmatrix} \tau^2 & \rho\sigma\tau \\ \rho\sigma\tau & \sigma^2 \end{pmatrix} \tag{3}$$

When covariance is unknown

$$\boldsymbol{\Sigma} \sim \text{InvWishart}_2\left(\boldsymbol{\Psi}_0 = \begin{pmatrix} \tau^2 & \rho\sigma\tau \\ \rho\sigma\tau & \sigma^2 \end{pmatrix}, \nu_0\right) \tag{4}$$

Decision Making Framework

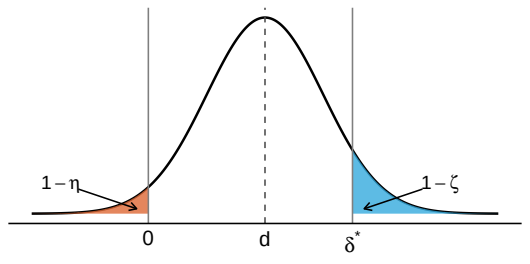


Figure 1: Posterior tail probabilities

Adjusting the framework from Whitehead et al. [2008]

$$\mathbb{P}(\delta_{Y|X} > 0 \mid \text{Covariate, Data}) \geq \eta \qquad \text{(Efficacy)} \qquad (5)$$

$$\mathbb{P}(\delta_{Y|X} < \delta^{\star} \mid \text{Covariate, Data}) \geq \zeta \qquad \text{(Futility)} \qquad (6)$$

Sample Size Formulae

Known:

$$\frac{(\kappa_{0E_Y} + n_E)(\kappa_{0C_Y} + n_C)}{\kappa_{0E_Y} + \kappa_{0C_Y} + N} \geq (1 - \rho^2)\sigma^2 \left(\frac{z_\eta + z_\zeta}{\delta^\star} \right)^2 \tag{7}$$

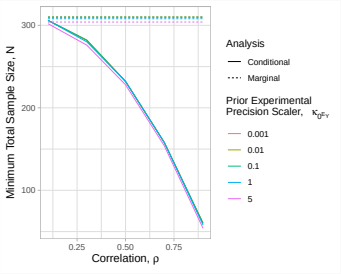
Unknown:

$$(\nu_0 + 3)\left(\frac{(\kappa_{0E_Y} + n_E)(\kappa_{0C_Y} + n_C)}{\kappa_{0E_Y} + \kappa_{0C_Y} + N} \right)\left(1 - \text{Beta}_{2,\left(\frac{\nu_0 - 1}{2}\right),\xi}\right) \geq (1 - \rho^2)\sigma^2 \left(\frac{t_{2\alpha_1,\eta} + t_{2\alpha_1,\zeta}}{\delta^\star} \right)^2 \tag{8}$$

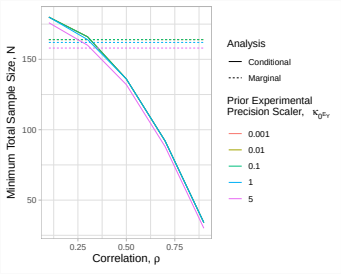
where

$$\alpha_1 = \left(\frac{\nu_0 - 1}{2} \right) + 2 \tag{9}$$

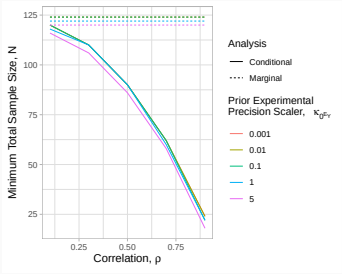
What does this mean?



(a) Known covariance matrix sample size



(b) Unknown scenario with 6 degrees of freedom



(c) Unknown scenario with 15 degrees of freedom

Thank you for your attention

References

George F Borm, Jaap Fransen, and Wim AJG Lemmens. A Simple Sample Size Formula for Analysis of Covariance in Randomized Clinical Trials. *Journal of Clinical Epidemiology*, 60(12):1234–1238, 2007. DOI: https://doi.org/10.1016/j.jclinepi.2007.02.006.

John Whitehead, Elsa Valdés-Márquez, Patrick Johnson, and Gordon Graham. Bayesian Sample Size for Exploratory Clinical Trials Incorporating Historical Data. *Statistics in Medicine*, 27(13):2307–2327, 2008. DOI: https://doi.org/10.1002/sim.3140.