Worked Example for Closed Cohort SWGRT with Discrete-Time Decay Correlation Structure

This calculator will use the parameter estimates you have provided to calculate the required sample size of groups or clusters per treatment sequence for the intervention effect and the level of power that you specified.

In addition, the calculator will use the other parameter estimates to calculate the detectable difference Δ that you specified as a function of the number of groups or clusters g in each treatment sequence and the number of members m in each group or cluster as described below.

The models and expressions described below are based on Hooper et al. (2016), Kasza, Hemming, et al. (2019), Kasza, Taljaard, and Forbes (2019), Kasza et al. (2020), and Hemming et al. (2020).

1 Group-Period Model

Let T be the total number of time periods including baseline. Let Y_{ikt} be a continuous outcome for the ith member ($i = 1, \dots, m$) nested within the kth group ($k = 1, \dots, g$) at time t ($t = 1, \dots, T$). In the closed cohort discrete-time decay setting, a group-period model for Y_{ikt} is

$$Y_{ikt} = \beta_t + \theta X_{kt} + GP_{kt} + MP_{ikt} \qquad (1)$$

where G_{kt} is a group-period random effect for group k at time t and MP_{ikt} is a random effect for member i in group k at time t. Let $GP_k = (GP_{k1}, \dots, GP_{kT})'$ and $MP_{ik} = (MP_{ik1}, \dots, MP_{ikT})'$ be the $T \times 1$ vectors of group-period and member-level random effects, respectively. GP_k and MP_{ik} are assumed to be independent and distributed as follows:

$$\begin{aligned} GP_k &\sim N(\mathbf{0}, \sigma^2_{G,D}\mathrm{D}_G) \\ MP_{ik} &\sim N(\mathbf{0}, \sigma^2_{M,D}\mathrm{D}_M) \end{aligned}$$

where $\mathbf{0}$ is a $T \times 1$ vector of zeroes, D_G and D_M are symmetric, $T \times T$ matrices with diagonal elements equal to 1, $\sigma^2_{G,D}$ is the variation attributable to group at time t, and $\sigma^2_{M,D}$ is the variation attributable to member at time t.

Let $\mathrm{D}_G[t,u]$ and $\mathrm{D}_M[t,u]$ denote the (t,u) element from matrices D_G and D_M, respectively. In the discrete-time decay setting, for $t \neq u$ we have $\mathrm{D}_G[t,u] = \pi_D^{|t-u|}$ and $\mathrm{D}_M[t,u] = \tau_D^{|t-u|}$, where π_D and τ_D are cluster autocorrelation (CAC) and individual autocorrelation (IAC) parameters subject to discrete-time decay, respectively.

Let $\overline{Y}_{\cdot kt} = \sum_{i=1}^{m} Y_{ikt} / m$ be the mean outcome for group k at time t. Then the group-period mean model is given by

$$\overline{Y}_{\cdot kt} = \beta_t + \theta X_{kt} + GP_{kt} + \overline{MP}_{\cdot kt} \qquad (2)$$

where $\overline{MP}_{\cdot kt} = \sum_{i=1}^{m} M\, P_{ikt}/m$ is the mean member-level effect for group k at time t. GP_k and $\overline{MP}_{\cdot kt}$ are assumed to be independent and $\overline{MP}_{\cdot kt} \sim N(\mathbf{0}, \frac{\sigma_{M,D}^2}{m}\mathrm{D}_M)$.

Let $S = T - 1$ denote the total number of treatment sequences. The variance of the treatment effect estimator σ_Δ^2 can be calculated using the following equation (Kasza, Taljaard, and Forbes 2019; Kasza et al. 2020).

$$\sigma_\Delta^2 = \frac{1}{g}\left\{\sum_{j=1}^{S} \mathbf{X}_j' \mathbf{V}^{-1}\mathbf{X}_j - \frac{1}{S}\left(\sum_{j=1}^{S} \mathbf{X}_j' \mathbf{V}^{-1}\right)\mathbf{V}\left(\sum_{j=1}^{S} \mathbf{V}^{-1}\mathbf{X}_j\right)\right\}^{-1} \qquad (3)$$

Here $\mathbf{X}_j$ is a $T \times 1$ vector indicating treatment assignments for treatment sequence j ($j = 1, \ldots, S$), $\mathbf{X}_j'$ denotes the transpose of $\mathbf{X}_j$, and $\mathbf{V}$ is a $T \times T$ covariance matrix of the $T \times 1$ vector of outcome means at each time period for a given group $\overline{\mathbf{Y}}_{\cdot k} = (\overline{Y}_{\cdot k1}, \ldots, \overline{Y}_{\cdot kT})'$. We assume $\mathbf{V}$ is the same for all groups.

Let $\sigma_Y^2 = \sigma_{G,D}^2 + \sigma_{M,D}^2$ be the total variance of the outcome at time period t ($t = 1, \ldots, T$) not adjusted for covariates. Let R_G^2 be the proportion of variance explained by group-level covariates and R_M^2 be the proportion of variance explained by member-level covariates. Then $\sigma_{G|X,D}^2 = \rho\sigma_Y^2(1 - R_G^2)$ and $\sigma_{M|X,D}^2 = (1-\rho)\sigma_Y^2(1 - R_M^2)$ are the group- and member-level components of variation adjusted for covariates. Let $\mathbf{V}[t,u]$ be the (t,u) element of $\mathbf{V}$. Then, when $t = u$, this element represents the variance in the outcome mean for a given group at time period t and is given by

$$\mathbf{V}[t,t] = Var(\overline{Y}_{\cdot kt}) = \sigma_{G|X}^2 + \frac{\sigma_{M|X}^2}{m} \qquad (4)$$

When $t \neq t'$, this element represents the covariance between the outcome means at time periods t and t' and is given by

$$\mathbf{V}[t,u] = Cov(\overline{Y}_{\cdot kt}, \overline{Y}_{\cdot ku}) = \sigma_{G|X,D}^2 \mathbf{D}_G[t,u] + \frac{\sigma_{M|X,D}^2}{m}\mathbf{D}_M[t,u] \qquad (5)$$

Finally, after calculating σ_Δ^2, we can calculate the detectable difference Δ using the equation below.

$$\Delta = \sqrt{\sigma_\Delta^2 (t_{\alpha/2} + t_\beta)^2}$$

Here $t_{\alpha/2}$ is the t-value selected based on the two-tailed level of significance and available degrees of freedom, and t_β is the t-value selected based on the desired power and available degrees of freedom. This calculator defines the degrees of freedom as the number of groups minus number of time periods minus 1 (Hemming et al. 2020). In addition, as per Murray (1998) we also subtract the degrees of freedom used for group-level covariates.

2 Example: Discrete-Time Decay Structure

As an example, consider the following set of parameter estimates:

Let $m = 10$, $g = 10$, $T = 4$, $S = 3$, and $df_g = 1$, where df_g are the degrees of freedom used for group-level covariates. Then

$$df = g * S - T - 1 - df_g = 10 * 3 - 4 - 1 - 1 = 24$$

For a two-tailed type 1 error rate of 5% and power of 80%

$$t_{\alpha/2} = 2.0639 \text{ and } t_{\beta} = 0.8569$$

Let σ^2 = 1.0, $\rho = 0.05$.

Let $R_G^2 = 0$ and $R_M^2 = 0.30$.

Let $\pi_D = 0.50$ and $\tau_D = 0.30$ with a discrete-time decay correlation structure.

$\mathbf{X}_1'$, $\mathbf{X}_2'$, and $\mathbf{X}_3'$ are the first, second, and third rows, respectively, of the following matrix of treatment sequences:

$$\begin{bmatrix} 0 & 1 & 1 & 1 \\ 0 & 0 & 1 & 1 \\ 0 & 0 & 0 & 1 \end{bmatrix}$$

The matrices $\mathbf{D}_G$ and $\mathbf{D}_M$ are as follows:

$$\mathbf{D}_G = \begin{bmatrix} 1 & 0.5 & 0.25 & 0.125 \\ 0.5 & 1 & 0.5 & 0.25 \\ 0.25 & 0.5 & 1 & 0.5 \\ 0.125 & 0.25 & 0.5 & 1 \end{bmatrix}$$

$$\mathbf{D}_M = \begin{bmatrix} 1 & 0.3 & 0.09 & 0.027 \\ 0.3 & 1 & 0.3 & 0.09 \\ 0.09 & 0.3 & 1 & 0.3 \\ 0.027 & 0.09 & 0.3 & 1 \end{bmatrix}$$

Using the values given above,

$$\sigma_{G|X,D}^2 = \rho\sigma^2(1 - R_G^2) = 0.05 * 1 * (1 - 0) = 0.05$$

$$\sigma_{M|X,D}^2 = (1 - \rho)\sigma^2(1 - R_M^2) = 0.95 * 1 * (1 - 0.30) = 0.665$$

Using these results with equations 4 and 5 yields the following for matrix $\mathbf{V}$:

$$\mathbf{V} = \begin{bmatrix} 0.1165 & 0.04495 & 0.01849 & 0.00805 \\ 0.04495 & 0.1165 & 0.04495 & 0.01849 \\ 0.01849 & 0.04495 & 0.1165 & 0.04495 \\ 0.00805 & 0.01849 & 0.04495 & 0.1165 \end{bmatrix}$$

Using g, S, $\mathbf{X}_j$ ($j = 1, 2, 3$), and $\mathbf{V}$ in equation 3 for σ_Δ^2 we find $\sigma_\Delta^2 = 0.0078$. Finally, the detectable difference Δ is determined to be

$$
\begin{aligned}
\Delta &= \sqrt{\sigma_\Delta^2 (t_{\alpha/2} + t_\beta)^2} \\
&= \sqrt{0.0078 * (2.0639 + 0.8569)^2} \\
&= 0.258
\end{aligned}
$$

Given the parameter estimates for the proposed SWGRT, with 4 time intervals and 3 sequences, with 10 groups or clusters in each sequence and 10 members observed in each group or cluster, an intervention effect of 0.258 standard deviation units is detectable with 80% power given an alpha level of 0.05.

References

Hemming, K., J. Kasza, R. Hooper, A. Forbes, and M. Taljaard. 2020. "A Tutorial on Sample Size Calculation for Multiple-Period Cluster Randomized Parallel, Cross-over and Stepped-Wedge Trials Using the Shiny CRT Calculator." Journal Article. *Int J Epidemiol* 49 (3): 979–95. https://doi.org/10.1093/ije/dyz237.

Hooper, R., S. Teerenstra, E. de Hoop, and S. Eldridge. 2016. "Sample Size Calculation for Stepped Wedge and Other Longitudinal Cluster Randomised Trials." Journal Article. *Stat Med* 35 (26): 4718–28. https://doi.org/10.1002/sim.7028.

Kasza, J., K. Hemming, R. Hooper, J. Matthews, and A. B. Forbes. 2019. "Impact of Non-Uniform Correlation Structure on Sample Size and Power in Multiple-Period Cluster Randomised Trials." Journal Article. *Stat Methods Med Res* 28 (3): 703–16. https://doi.org/10.1177/0962280217734981.

Kasza, J., R. Hooper, A. Copas, and A. B. Forbes. 2020. "Sample Size and Power Calculations for Open Cohort Longitudinal Cluster Randomized Trials." Journal Article. *Stat Med*. https://doi.org/10.1002/sim.8519.

Kasza, J., M. Taljaard, and A. B. Forbes. 2019. "Information Content of Stepped-Wedge Designs When Treatment Effect Heterogeneity and/or Implementation Periods Are Present." Journal Article. *Stat Med* 38 (23): 4686–4701. https://doi.org/10.1002/sim.8327.

Murray, David M. 1998. *Design and Analysis of Group-Randomized Trials*. New York, NY: Oxford University Press.