Technical Appendix: Details on the Power Calculations

Kristoffer Magnusson
Department of Clinical Neuroscience
Karolinska Institutet

2017-09-08

In standard multilevel notation the fully nested three-level model is

$$
\begin{aligned}
&\text{Level 1} \\
&\quad Y_{ijk} = \beta_{0jk} + \beta_{1jk} t_{ijk} + R_{ijk} \\
&\text{Level 2} \\
&\quad \beta_{0jk} = \gamma_{00k} + U_{0jk} \\
&\quad \beta_{1jk} = \gamma_{10k} + U_{1jk} \\
&\text{Level 3} \\
&\quad \gamma_{00k} = \delta_{000} + \delta_{001} TX_k + V_{0k} \\
&\quad \gamma_{10k} = \delta_{100} + \delta_{101} TX_k + V_{1k}
\end{aligned}
$$

The parameter of interest is δ_{101}, i.e. the mean difference in slopes between the two treatment groups. However, we can simplify the calculations by calculation the variance of the slope-coefficient separately for each treatment group. Since the slope in the treatment and control group are independent, the variance of the interaction-term is simply

$$\mathrm{V}(\delta_{101}) = \mathrm{V}(\delta_{100[tx]} - \delta_{100[c]}) = \mathrm{V}(\delta_{100[tx]}) + \mathrm{V}(\delta_{100[c]})$$

where $\delta_{100[tx]}$ and $\delta_{100[c]}$ is fixed time effect in the treatment and control group respectively. In order to calculate the variances we begin by formulation the three-level model for the complete data from a single treatment arm.

$$\mathbf{Y} = \mathbf{XZW}\boldsymbol{\beta} + \mathbf{Xu} + \mathbf{XZv} + \boldsymbol{\epsilon} \tag{1}$$

where $\mathbf{Y}$ is the $N_1 \times 1$ outcome vector containing the for all observations from all the subjects in the treatment arm, $\mathbf{X}$ is a $N_1 \times 2N_2$ matrix containing co-variate information for all $i = 1, \ldots, N_2$ subjects in the treatment arm, $\mathbf{X}$ is also used as the design matrix for the second-level random effects. $\mathbf{Z}$ is a $2n_3 \times 2N_2$ matrix containing the level-three random effects design matrices for each $j = 1, \ldots, n_3$ cluster in the treatment arm. $\mathbf{W}$ is a $2n_3 \times 2$ matrix relating the third-level to the overall effects β, and here β is simply a 2×1 vector with the population values for the fixed intercept and slope effects. Lastly, $\mathbf{u}$ is a $2N_2 \times 1$ vector with the level two random effects, $\mathbf{v}$ is a $2n_3 \times 1$ vector with the third-level random effects, and $\boldsymbol{\epsilon}$ a $N_1 \times 1$ vector with the level one residuals.

The random effects and residuals being distributed

$$
\begin{aligned}
\mathbf{u} &\sim \mathcal{N}(\mathbf{0}, \boldsymbol{\Psi}_2), \\
\mathbf{v} &\sim \mathcal{N}(\mathbf{0}, \boldsymbol{\Psi}_3), \\
\boldsymbol{\epsilon} &\sim \mathcal{N}(\mathbf{0}, \sigma^2 \mathbf{I}_{N_1}).
\end{aligned}
$$

With the second-level variance components being

$$\mathbf{\Psi}_2 = \mathbf{I}_{N_2} \otimes \begin{pmatrix} u_0^2 & u_{01} \\ u_{01} & u_1^2 \end{pmatrix}$$

and the third-level variance components being

$$\mathbf{\Psi}_3 = \mathbf{I}_{n_3} \otimes \begin{pmatrix} v_0^2 & v_{01} \\ v_{01} & v_1^2 \end{pmatrix}$$

$\mathbf{X}$ is a block-diagonal matrix containing a sub-matrix $\mathbf{X}_{ij}$ for each subject (level-two unit).

$$\mathbf{X} = \begin{pmatrix} \mathbf{X}_1 & 0 & \cdots & 0 \\ 0 & \mathbf{X}_2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \mathbf{X}_N \end{pmatrix}$$

Since each subject can have a different number of observations due to dropout, each $\mathbf{X}_{ij}$ will have dimension $n_{1[i]} \times 2$, where $n_{1[i]}$ is the total number of observations for subject i in cluster j.

$$\mathbf{X}_{ij} = \begin{pmatrix} 1 & T_0 \\ 1 & T_1 \\ \vdots & \vdots \\ 1 & T_{n1[i]} \end{pmatrix} \tag{2}$$

$\mathbf{Z}$ is a block diagonal matrix containing the level-three design matrices for each cluster j.

$$\mathbf{Z} = \begin{pmatrix} \mathbf{Z}_1 & 0 & \cdots & 0 \\ 0 & \mathbf{Z}_2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \mathbf{Z}_{n3} \end{pmatrix}$$

With the sub-matrices $\mathbf{Z}_k$ being stacks of 2×2 matrices for each subject in cluster j.

$$\mathbf{Z}_j = \mathbf{1}_{n2[j]} \otimes \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}$$

thus the dimension of $\mathbf{Z}_j$ will be $n_{2[j]} \times 2$, where $n_{2[j]}$ is the number of subjects in cluster j. This enables power calculations for designs with varying number of subjects per cluster.

The matrix $\mathbf{W}$, relates the cluster-level effects to the overall effects $\boldsymbol{\beta}$

$$\mathbf{W} = \mathbf{1}_{n3} \otimes \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}$$

Thus

$$\mathbf{XZW} = \mathbf{1}_N \otimes \mathbf{X}_{ij}$$

Then we can calculate the marginal variance-covariance matrix for $\mathbf{Y}$ as

$$\mathrm{V}(\mathbf{Y}) = \mathbf{X}\mathbf{\Psi}_2\mathbf{X}^\top + \mathbf{XZ}\mathbf{\Psi}_3\mathbf{Z}^\top\mathbf{X}^\top + \boldsymbol{\epsilon}^2\mathbf{I}_{N1}$$

And

$$\mathrm{V}(\boldsymbol{\beta}) = [(\mathbf{XZW})^\top \mathrm{V}(\mathbf{Y})^{-1}\mathbf{XZW}]^{-1}$$

The lower right corner of $\mathrm{V}(\boldsymbol{\beta})$ corresponds to the variance of the time-coefficient. As we noted earlier we can use the slope variances to calculate the variance of the time×treatment-interaction.

Accounting for dropout

Dropout is calculated by defining a dropout vector $\mathbf{D} = (p_1, \ldots, p_{n_1})^\top$, where p_i is the proportion of participants that have dropped out at time point i, for the $i, \ldots, n_1$ scheduled time points, and $p_0 = 0$ and $p_i \leq p_{i+1}$. The default in `powerlmm` is to treat the values in $\mathbf{D}$ as known, i.e. exactly p_i subjects will have dropped out at time i. This is done by randomly sample which $p_i N_2$ participants should drop out a time i, then adjusting their design matrices $\mathbf{X}_{ij}$ to be of size $(i-1) \times 2$, thus their last time point will be $i-1$. Since its random which subjects will dropout, the power calculations will differ slightly each time. It is also possible to treat $\mathbf{D}$ as random (using the option `deterministic_dropout = FALSE`), then dropout will be sampled from a multinomial distribution, by converting the element of $\mathbf{D}$ to the probability p_i that a subject will have exactly i measurements. This approach is similar to Galbraith, Stat, and Marschner (2002), and Verbeke and Lesaffre (1999) who presents power calculations for two-level models with missing data.

Speeding up the computation of $V(Y)^{-1}$

Doing the matrix inversion of $V(\mathbf{Y})$, which is of dimension $N_1 \times N_1$, can be extremely slow for some designs. De Leeuw and Kreft (1986) (where they credit Swamy (1971)) noted a more computationally efficient formulation, adopting it to or three-level formulation (in Equation 1), lets us write

$$V(\mathbf{Y})^{-1} = \sigma^{-2}[\mathbf{I}_{N_1} - \mathbf{X}(\mathbf{X}^\top\mathbf{X})^{-1}\mathbf{X}^\top] + \mathbf{X}(\mathbf{X}^\top\mathbf{X})^{-1}\mathbf{A}^{-1}(\mathbf{X}^\top\mathbf{X})^{-1}\mathbf{X}$$

where

$$\mathbf{A}^{-1} = [\sigma^2(\mathbf{X}^\top\mathbf{X})^{-1} + \mathbf{\Psi}_2 + \mathbf{Z}\mathbf{\Psi}_3\mathbf{Z}^\top].$$

Here $\mathbf{A}$ of size $2N_2 \times 2N_2$. However, since $\mathbf{A}$ is block-diagonal, with each block for cluster j being of size $n_{3[j]}$, the computation done in `powerlmm`, takes advantage of the sparse matrix functions from the `Matrix`-package. By using sparse matrix algebra the speed of computing $\mathbf{A}^{-1}$ will depend greatly on the number of subjects per cluster. In most cases this solution is dramatically faster then directly solving $V(\mathbf{Y})^{-1}$. For instance, calculating $\mathrm{V}(\beta)$ for a study with $n_1 = 10$, $n_2 = 30$, $n_3 = 20$ is approximately 50 times faster using this method.

Power

The power function is

$$P(t_{\nu,\lambda} > t_{\nu,1-\alpha/2}) + P(t_{\nu,\lambda} < t_{\nu,\alpha/2})$$

Where λ is the non-centrality parameter, calculated as

$$\lambda = \delta_{101}/\sqrt{\mathrm{V}(\delta_{101})}$$

and ν is the appropriate degrees of freedom of the t distribution, for the fully nested three-level model the degrees of freedom are $N3 - 2$, where $N3$ is the total number of clusters in both treatment arms.

Partially nested designs

For the partially nested designs $\mathrm{V}(\delta_{100[tx]})$ is calculated as above, and $\mathrm{V}(\delta_{100[c]})$ by setting the cluster-level random effects to zero. Degrees of freedom for this model is trickier, and currently $2n_3 - 2$ i used, where n_3 is the number of clusters in the treatment group only. This is not exact, but works better than doing a normal approximation.

Two-level designs

For the two-level designs, $\mathrm{V}(\delta_{101})$ can be calculated using the three-level formulas with the cluster-level random effects set to zero. Deleting these terms reduces the model to the classical two-level formulation. Degrees of freedom for this model is $N_2 - 2$, where N_2 is the total number of subjects in both treatment arms.

Standardized formulation

If there's no missing data and the clusters sizes are balanced, the variance of the slope can be calculated more simply as

$$\mathrm{V}(\delta_{100}) = \frac{\sigma^2 + n_1\sigma^2_{u_1}V(T) + n_1n_2\sigma^2_{v_1}V(T)}{n_1n_2n_3V(T)}$$

With

$$\mathrm{V(T)} = \Sigma_{i=1}^{n_1}(t_i - \bar{t})^2$$

By defining the amount of slope variance at the cluster-level as $\rho_s = \sigma^2_{v_1}/(\sigma^2_{v_1} + \sigma^2_{u_1})$ and the variance ratio as $r_\tau = (\sigma^2_{v_1} + \sigma^2_{u_1})/\sigma^2_e$ we can then rewrite the formula using the relative parameters ρ_s and r_τ, which will yield the same non-centrality parameters as long as the interaction-coefficient corresponds to the same standardized value, e.g. Cohen's d.

$$\mathrm{V}(\delta_1^*) = \frac{1 + n_1\mathrm{V}(T)(n_2\rho_s r_\tau + (1-\rho_1)r_\tau)}{n_1n_2n_3\mathrm{V}(T)}$$

And see that power depends n1, n2, n3, T_end, the amount of slope variance at the third level and the variance ratio.

References

De Leeuw, J, and I Kreft. 1986. "Random coefficient models for multilevel analysis." *Journal of Educational Statistics* 11 (1): 57.

Galbraith, S, M Stat, and I C Marschner. 2002. "Guidelines for the design of clinical trials with longitudinal outcomes." *Controlled Clinical Trials* 23 (3): 257–73.

Swamy, Paravastu Ananta Venkata Bhattanadha. 1971. *Statistical Inference in Random Coefficient Regression Models*. New York: Springer Verlag.

Verbeke, Geert, and Emmanuel Lesaffre. 1999. "The Effect of Drop-Out on the Efficiency of Longitudinal Experiments." *Journal of the Royal Statistical Society* 48 (3): 363–75.