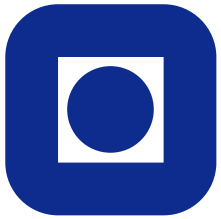

NTNU
Norwegian University of
Science and Technology

KLMED8008 Analysis of repeated measurements: Power and sample size: Simulation

Turid Follestad

Sample size for clustered data

Sample size for comparing two groups (treatment/control):

$$
\begin{aligned}
m_{\text{cluster}} &= c \times k = \text{total number of subjects in each group} \\
c &= \text{number of clusters} \\
k &= \text{number of subjects in each cluster (i.e. cluster size)}
\end{aligned}
$$

Three problems:

1. Given n in each group from usual calculations (disregarding the clustering), how large should m_{cluster} be?
2. Given m_{cluster} and k, how many clusters (c) are required?
3. Given n (from usual calculations) and c, how large should the cluster size (k) be?

Clustered data: Can compute the required numbers based on the ICC (intra-class correlation)

Sample size for longitudinal data

- Often more complex within-cluster correlation structures (auto-correlations) are required
- Analytic calculation of sample size only for simplest models
- Can use stochastic simulation for power and sample size calculations

Sample size calculations for LMM by stochastic simulation

Example:
Random intercept model for a longitudinal RCI comparing two treatments (active, placebo), with time as factor:

$$\begin{aligned} y_{i,j} &= \beta_{tr(j)} + \alpha_i + \zeta_j + \epsilon_{i,j} \\ \beta_{tr} &= \text{effect of treatment} \\ \alpha_i &= \text{effect of time, } i = 1, \ldots, n_t \end{aligned}$$

- $\zeta_j \sim N(0, \psi)$
- $\epsilon_{\mathbf{j}} = (\epsilon_{1,j}, \epsilon_{2,j}, \ldots, \epsilon_{n_t,j})$ multivariate normal with mean zero and covariance matrix Σ.
- Aim: Compare post-treatment means between the groups, what is required sample size?

Sample size calculations for LMM by stochastic simulation

Given number of time points (n_t), what is required number of subjects (n)?

- Specify the model for the fixed part of the LMM, including hypothesized values for the regression coefficients
- Specify the model for the random part of the LMM
  - Marginal model: Covariance structure for the total residuals, including hypothesized values of variances and covariances (correlations)
  - Multi-level models: Covariance structure for the random effects and the (level-1) residuals, including hypothesized values of variances and covariances (correlations)
- Select significance level α

Sample size calculations for LMM by stochastic simulation (cont.)

- Select candidate values of the sample size n
- Select the number of iterations n_{it} for each value of n
- For each n (sample size): Loop over all iterations $1, \ldots, n_{it}$
  - Generate artificial data set by stochastic simulation from the model
  - Estimate parameters for the simulated data by LMM
  - Check whether p-value for comparison of interest is $< \alpha$
- Power = fraction of tests with p-value $< \alpha$
- Repeate for increasing n until power is > required power (80%, 90%)