Contents

1 [HH07]

1.1 Introduction

- In cluster randomized trials (CRTs), randomization of interventions are done at the level of groups rather than individuals. This is most useful when intervention can only be administered at the community-level, when contamination may be an issue, or for other reasons. The inidividual units within a cluster are correlated.

- In parallel CRTs, clusters are all assigned an intervention at a single time point. If cluster sizes are equal, t-tests/ANOVA may be used to compare cluster-level mean responses. Clusters may also be matched for a paired setup. When cluster sizes vary, individual level analyses such as generalized estimating equations (GEE) or random effects models may be used.

- In crossover CRTs, each interventions is applied to a cluster at different time points (with a possible "washout" period in between). The order of the interventions is randomized for each cluster. Crossover designs are less commonly used in CRTs due to extending the period of the study. Crossover CRTs usually use paired t-tests to make within-cluster comparisons.

- In a stepped wedge CRT, clusters cross from one intervention to another at different time points (typically all starting from the control intervention and all ending with the treatment intervention). The time at which each cluster crosses over is randomized. The stepped wedge CRT is useful when there are limited resources for applying the intervention. The key features of the stepped wedge CRT is that the crossover is unidirectional and that the intervention is never removed once implemented. The unidirectionality does complicate the analysis as the treatment effect cannot be estimated from only within-cluster comparisons.

1.2 Model

- For a design with I clusters, T time points, N individuals sampled per cluster per time interval, define the cluster means

  $$\mu_{ij} = \mu + \alpha_i + \beta_j + \theta X_{ij}$$

  where

  - $\alpha_i \sim N(0, \tau^2)$ is a random effect for cluster $i \in \{1, \dots, I\}$
  - β_j is a fixed effect for time interval $j \in \{1, \dots, T-1\}$ (with $\beta_T = 0$ for identifiability)
  - X_{ij} is the intervention indicator in cluster i at time j (with 1 denoting intervention)
  - θ is the treatment effect.

- Individual level responses are modelled as

  $$Y_{ijk} = \mu_{ij} + \epsilon_{ijk}$$

  where $\epsilon_{ijk} \overset{iid}{\sim} N(0, \sigma^2)$. Cluster means are modelled as

  $$\bar{Y}_{ij.} = \mu_{ij} + \bar{\epsilon}_{ij.}$$

  where $\bar{\epsilon}_{ij.} = \frac{\sum_k \epsilon_{ijk}}{N}$. Assume that ϵ_{ijk} are independent of α_i.

- The variance of the individual-level response is

  $$\text{Var}(Y_{ijk}) = \tau^2 + \sigma^2$$

  The variance of the cluster-level response is

  $$\text{Var}(\bar{Y}_{ij.}) = \tau^2 + \frac{\sigma^2}{N} = \left(\frac{\tau^2 + \sigma^2}{N}\right)(1 + (N-1)\rho)$$

where $(1+(N-1)\rho$ is the "variance inflation factor" and $\rho = \frac{\tau^2}{\tau^2+\sigma^2}$ is the intraclass correlation. Note:

$$\begin{aligned}
\left(\frac{\tau^2+\sigma^2}{N}\right)(1+(N-1)\rho) &= \left(\frac{\tau^2+\sigma^2}{N}\right)\left(1+\frac{(N-1)\tau^2}{\tau^2+\sigma^2}\right) \\
&= \frac{\tau^2+\sigma^2}{N} + \frac{N\tau^2(\tau^2+\sigma^2)}{N(\tau^2+\sigma^2)} - \frac{(\tau^2+\sigma^2)\tau^2}{N(\tau^2+\sigma^2)} \\
&= \frac{\tau^2+\sigma^2}{N} + \tau^2 - \frac{\tau^2}{N} \\
&= \tau^2 + \frac{\sigma^2}{N}
\end{aligned}$$

- Some characterize the cluster effect on the variance using the coefficient of variation (CV) $\frac{\tau}{\mu}$.
- If the individual level responses are binary, then the cluster-level response $\bar{Y}_{ij.}$ is a proportion and it is assumed $\sigma^2 = \mu(1-\mu)$.
- For varying-sized clusters, replace N with N_{ij}.

1.3 Data analysis

1.3.1 τ^2 and σ^2 known

If the variances τ^2 and σ^2 are known, then estimates of the fixed effects can be estimated used weighted least squares (WLS) at the cluster-level. Let $\mathbf{X}$ be the $IT \times (T+1)$ design matrix of cluster-time means corresponding to parameter vector $\eta = (\mu, \beta_1, \ldots, \beta_{T-1}, \theta)$. Let $\mathbf{W}$ be a $IT \times IT$ block diagonal matrix where each $T \times T$ block of $\mathbf{W}$ describes the correlation structure between the repeated cluster means over time and has the structure

$$\begin{bmatrix}
\tau^2+\frac{\sigma^2}{N} & \tau^2 & \ldots & \tau^2 \\
\tau^2 & \ddots & & \vdots \\
\tau^2 & & \ddots & \tau^2 \\
\tau^2 & \ldots & \tau^2 & \tau^2+\frac{\sigma^2}{N}
\end{bmatrix}$$

Then the fixed effect estimates are given by

$$\hat{\eta} = (\mathbf{X}^T\mathbf{W}^{-1}\mathbf{X})^{-1}\mathbf{X}^T\mathbf{W}^{-1}\mathbf{y}$$

1.3.2 τ^2 and σ^2 unknown

- When the response is continuous and normally distributed, an empirical Bayes approach at the cluster-level is possible to estimate the fixed effects in the LMM. This approach also works for non-normal individual-level data when the cluster sizes are approximately equal.
- If the responses are non-normal and the cluster sizes vary, then analysis at the individual-level using GLMM or GEE is preferred.

1.3.3 Within-cluster analysis

If there are no temporal effects on the outcome (i.e., $\beta_j = 0$ for all j), then a within-clluster analysis can be used to estimate the treatment effect. Let t_i be the last time point at which cluster i receives the control. Then a within-cluster estimator of θ is given by

$$\tilde{\theta} = \frac{1}{I}\sum_i \left(\frac{\sum_{j>t_i}\bar{Y}_{ij.}}{T-t_i} - \frac{\sum_{j\le t_i}\bar{Y}_{ij.}}{t_i}\right)$$

and the variance is given by

$$\text{Var}(\tilde{\theta}) = \frac{\sigma^2}{NI^2} \sum_i \left(\frac{1}{t_i} + \frac{1}{T - t_i} \right)$$

A paired t-test is appropriate for testing the treatment effect in this case.

If the time effects are non-trivial, then the estimator is biased. The bias is

$$\text{bias}(\tilde{\theta}, \theta) = \frac{1}{I} \sum_i \left(\frac{\sum_{j > t_i} \beta_j}{T - t_i} - \frac{\sum_{j \leq t_i} \beta_j}{t_i} \right) = \sum_j \beta_j \sum_i \frac{t_i - T(1 - X_{ij})}{It_i(T - t_i)}$$

Note that the bias is independent of the true value θ. The bias is also a linear combination of the time effects where the weights can be calculated once the treatment schedule is determined. Understanding the contribution of the time effects can be done during the design phase of the trial.

1.4 Power analysis

Consider testing the hypothesis $H_0 : \theta = 0$ versus $H_a : \theta = \theta_a$. A Wald test may be based on $Z = \frac{\theta}{\sqrt{\text{Var}(\theta)}}$ where $\hat{\theta}$ is from $\hat{\eta}$. The approximate power for a two-tailed test of size α is

$$\text{power} = \Phi \left(\frac{\theta_a}{\sqrt{\text{Var}(\hat{\theta})}} - Z_{1-\alpha/2} \right)$$

where Φ is the cumulative standard normal distribution and $Z_{1-\alpha/2}$ is the $(1 - \alpha/2)$-th quantile of the standard normal distribution. $\text{Var}(\hat{\theta})$ is an element of $(\mathbf{X}^T \mathbf{W}^{-1} \mathbf{X})^{-1}$ but may be possible to express in the closed form

$$\text{Var}(\hat{\theta}) = \frac{I\sigma^2 \left(\frac{\sigma^2}{N} + T\tau^2 \right)}{(IU - W)\sigma^2 + N(U^2 + ITU - TW - IV)\tau^2}$$

where

$$U = \sum_{ij} X_{ij}$$

$$W = \sum_j \left(\sum_i X_{ij} \right)^2$$

$$V = \sum_i \left(\sum_j X_{ij} \right)^2$$

Note: under H_a,

$$Z - \frac{\theta_a}{\sqrt{\text{Var}(\hat{\theta})}} \sim N(0, 1) \rightarrow Z \sim N \left(\frac{\theta_a}{\sqrt{\text{Var}(\hat{\theta})}}, 1 \right)$$

and so

$$\begin{aligned}
\text{power} &= P\left(Z > Z_{1-\alpha/2}\right) \\
&= P\left(Z - \frac{\theta_a}{\sqrt{\operatorname{Var}(\hat{\theta})}} > Z_{1-\alpha/2} - \frac{\theta_a}{\sqrt{\operatorname{Var}(\hat{\theta})}}\right) \\
&= 1 - \Phi\left(Z_{1-\alpha/2} - \frac{\theta_a}{\sqrt{\operatorname{Var}(\hat{\theta})}}\right) \\
&= \Phi\left(\frac{\theta_a}{\sqrt{\operatorname{Var}(\hat{\theta})}} - Z_{1-\alpha/2}\right)
\end{aligned}$$

1.5 Effect of number of steps

Optimal power is achieved when a single cluster crosses over to the intervention at each time. The loss of power is primarily due to the loss of measurement times rather than due to the loss of randomization times. The loss of power is also relatively independent of the CV.

1.6 Efficacy of WLS relative to within-cluster analysis

The relative efficacy of the WLS estimator $\hat{\theta}$ versus the within-cluster estimator $\tilde{\theta}$ is given by the inverse ratio of the variances. If there are no time effects, the ratio is

$$\text{efficacy}(\hat{\theta}, \tilde{\theta}) = \frac{\sum_i \left(\frac{1}{t_i} + \frac{1}{T-t_i}\right)\left(ITU - U^2)\frac{\sigma^2}{N} + IT(TU-V)\tau^2\right)}{I^3\left(\frac{\sigma^2}{N} + T\tau^2\right)}$$

When there are no time effects, the WLS estimator is more efficient than the within-cluster estimate unless $\tau^2 = 0$. If there are time effects, the WLS estimator is less efficient but the within-cluster estimator is likely biased.
Note: $\operatorname{Var}(\hat{\theta})$ is not the one given above as it is assumed there are no time effects.

$$\begin{aligned}
\text{efficacy}(\hat{\theta}, \tilde{\theta}) &= \frac{\operatorname{Var}(\tilde{\theta})}{\operatorname{Var}(\hat{\theta})} \\
&= \frac{\left(\frac{\sigma^2}{NI^2}\sum_i\left(\frac{1}{t_i} + \frac{1}{T-t_i}\right)\right)}{\left(\frac{I\sigma^2\left(\frac{\sigma^2}{N} + T\tau^2\right)}{(IU-W)\sigma^2 + N(U^2 + ITU - TW - IV)\tau^2}\right)} \\
&= \frac{\sum_i\left(\frac{1}{t_i} + \frac{1}{T-t_i}\right)\left((IU-W)\sigma^2 + N(U^2 + ITU - TW - IV)\tau^2\right)}{NI^3\left(\frac{\sigma^2}{N} + T\tau^2\right)} \\
&= \frac{\sum_i\left(\frac{1}{t_i} + \frac{1}{T-t_i}\right)\left((IU-W)\frac{\sigma^2}{N} + (U^2 + ITU - TW - IV)\tau^2\right)}{I^3\left(\frac{\sigma^2}{N} + T\tau^2\right)}
\end{aligned}$$

1.7 Delayed treatment effect

If effect of intervention only reaches full effect θ after some time, then power is reduced. The delay may be modelled by allowing X_{ij} to be fractional (although the given $\operatorname{Var}(\hat{\theta})$ is not valid in this case). Power is increased by adding additional measurement periods at the end of the trial or by increasing the time intervals.

1.8 Simulation results

Compared power of test for LMM, GEE and GLMM for varying levels of relative risk. In equal cluster size case, LMM $>$ GEE $>$ GLMM. In unequal cluster size case, GEE $\approx$ GLMM $>$ LMM. When cluster sizes vary significantly, it is suggested to do individual-level analyses. A jackknife estimate of the variance is suggested to maintain the size of the test in GEE/GLMM analyses.

2 (TODO) Contamination

One benefit of CRTs is said to minimize contamination risks. In some situations, contamination of individuals in a cluster may still be possible. Can we model potential contamination directly in the model? Can estimation of the treatment effect still be done? How may it influence the power of the study?

2.1 Contamination references

- Contamination in trials: is cluster randomisation the answer?
- Reducing contamination risk in cluster-randomized infectious disease-intervention trials
- The Implications of “Contamination” for Experimental Design in Education (Rhoads, 2011): compared contamination in CRTs and RBDs based on several measures (power, accuracy, MSE, etc.).
- Analysis of contamination in cluster randomized trials of malaria interventions
- A scoping review of the problems and solutions associated with contamination in trials of complex interventions in mental health

2.2 Possible methods references

- Repeated measures regression mixture models
- Example

3 Hemming et al. [Hem+15]

This paper provides a high-level review of the motivation, design, analysis, and reporting considerations of a stepped wedge cluster randomized trial.

- Stepped wedge CRT is a pragmatic study design compared to competing designs in that its design includes built-in logistical and ethical (if there is some evidence that the intervention is beneficial) considerations.

- Methods for sample size and power calculations have only been described for cross-sectional stepped wedge designs. Power calculations and efficiency depend on ICC, number of clusters, number of observations in each cluster, and the structure of the design. Methods that determine power for trials of fixed size are implemented in Stata (and R).

- Care should be taken to mitigate the risk of participants varying systematically across exposed and unexposed observation periods. Ideally, participants are recruited before allocation or blind to exposure status.

- Clusters act as their own controls, but calendar time is a potential confounder associated with both exposure to intervention and possibly outcome.

- Directions of further research: design and analysis of cohort stepped wedge trials, clusters within clusters, trials with more than two arms, restricted randomization (e.g., pairing), effect of varying cluster sizes and varying step sizes, and hybrid designs.

4 Ouyang et al. [Ouy+20]

This paper investigated the variation in the attained power (as opposed to expected power) of a cross-sectional stepped wedge design with unequal cluster sizes. The cross-sectional design refers to the design where participants receive the treatment delivered by their cluster at the time of study entry and contributes a single measurement. The motivation is because the power varies according to the random allocation of the design when cluster sizes are unequal, and the actual power may be lower than what is expected. To investigate the problem, the researchers examine the power distribution of a hypothetical trial across allocations and across values of relelvant parameters, such as sample sizes, ICC, and CV. Sample sizes were simplified to be categorical. Configurations of parameters were investigated through a factorial design experiment. In addition to investigating the power distribution, the researchers also attempted to explain the variation in attained power through a logistic regression model on allocation characteristics TTC (treatment-vs-time period correlation) and TGI (treatment group imbalance).

The researchers found that risks of low power generally decreased with increasing ICC and decreasing CV. TTC was found to be the dominating factor in explaining the variation in power.

5 Zhan et al. [Zha+21]

The paper examined the effects of non-informative and informative priors for the fixed time effects in a Bayesian stepped wedge CRT model on the power of the trial. Their simulation results showed that non-informative priors generally returned a minimum required sample size similar to that calculated under the Frequentist framework and that informative, well-specified priors generally reduced the minimum required sample size. Under modest misspecification, the bias in the intervention effect estimate remains relatively low. However, if the mean is greatly mis-specified and the prior is made very precise, the bias increases which may lead to overestimation of the intervention effect.

6 Bowden, Forbes, and Kasza [BFK21]

The paper examined the effects of misspecifying the model in crossover and stepped wedge CRTs when the treatment effects are assumed to homogeneous when they are actually heterogeneous across clusters. In particular, they estimate the (approximated) variance of the treatment effect estimator under the misspecified model. Their findings show that the variance of the estimator is generally underestimated when treatment effect heterogeneity is ignored, and that the problem is amplified when the number of clusters or the number of time periods increases.

7 Teerenstra et al. [Tee+19]

The paper investigates the effect of using a multilevel cluster model in a cross-sectional stepped wedge CRT on the variance of the treatment effect estimator, and consequently, the sample size and power. Their results show that the variance of the estimator is inflated in a multilevel design by factors defined as the correlation of units within a higher level unit. For a multilevel design, increasing the number of clusters, increasing the number of times at which a cluster crosses over, and increasing the sample size at any level all increase the power of the study.

8 Wang et al. [Wan+21]

The paper proposes a method for calculating power and sample size in closed-cohort and cross-sectional CRTs with binary outcomes. The method is based on GEE and can take into account missing data. The method is evaluated based on simulated studies and a real study.

References

[BFK21] Rhys Bowden, Andrew B Forbes, and Jessica Kasza. “Inference for the treatment effect in longitudinal cluster randomized trials when treatment effect heterogeneity is ignored”. In: *Statistical Methods in Medical Research* (2021). PMID: 34569853, p. 09622802211041754. DOI: 10.1177/09622802211041754. eprint: https://doi.org/10.1177/09622802211041754. URL: https://doi.org/10.1177/09622802211041754.

[Hem+15] K Hemming et al. “The stepped wedge cluster randomised trial: rationale, design, analysis, and reporting”. In: *BMJ* 350 (2015). DOI: 10.1136/bmj.h391. eprint: https://www.bmj.com/content/350/bmj.h391.full.pdf. URL: https://www.bmj.com/content/350/bmj.h391.

[HH07] Michael A. Hussey and James P. Hughes. “Design and analysis of stepped wedge cluster randomized trials”. English. In: *Contemporary clinical trials* 28.2 (2007), pp. 182–191.

[Ouy+20] Yongdong Ouyang et al. “Explaining the variation in the attained power of a stepped-wedge trial with unequal cluster sizes”. In: *BMC medical research methodology* 20.1 (2020), pp. 1–14.

[Tee+19] Steven Teerenstra et al. “Sample size calculation for stepped-wedge cluster-randomized trials with more than two levels of clustering”. In: *Clinical Trials* 16.3 (2019). PMID: 31018678, pp. 225–236. DOI: 10.1177/1740774519829053. eprint: https://doi.org/10.1177/1740774519829053. URL: https://doi.org/10.1177/1740774519829053.

[Wan+21] Jijia Wang et al. “Sample size and power analysis for stepped wedge cluster randomised trials with binary outcomes”. In: *Statistical Theory and Related Fields* 5.2 (2021), pp. 162–169. DOI: 10.1080/24754269.2021.1904094. eprint: https://doi.org/10.1080/24754269.2021.1904094. URL: https://doi.org/10.1080/24754269.2021.1904094%5C.

[Zha+21] Denghuang Zhan et al. “Improving efficiency in the stepped-wedge trial design via Bayesian modeling with an informative prior for the time effects”. In: *Clinical Trials* 18.3 (2021). PMID: 33813906, pp. 295–302. DOI: 10.1177/1740774520980052. eprint: https://doi.org/10.1177/1740774520980052. URL: https://doi.org/10.1177/1740774520980052.