

RESEARCH ARTICLE

Open Access



How large are the consequences of covariate imbalance in cluster randomized trials: a simulation study with a continuous outcome and a binary covariate at the cluster level

Mirjam Moerbeek*  and Sander van Schie

Abstract

Background: The number of clusters in a cluster randomized trial is often low. It is therefore likely random assignment of clusters to treatment conditions results in covariate imbalance. There are no studies that quantify the consequences of covariate imbalance in cluster randomized trials on parameter and standard error bias and on power to detect treatment effects.

Methods: The consequences of covariance imbalance in unadjusted and adjusted linear mixed models are investigated by means of a simulation study. The factors in this study are the degree of imbalance, the covariate effect size, the cluster size and the intraclass correlation coefficient. The covariate is binary and measured at the cluster level; the outcome is continuous and measured at the individual level.

Results: The results show covariate imbalance results in negligible parameter bias and small standard error bias in adjusted linear mixed models. Ignoring the possibility of covariate imbalance while calculating the sample size at the cluster level may result in a loss in power of at most 25 % in the adjusted linear mixed model. The results are more severe for the unadjusted linear mixed model: parameter biases up to 100 % and standard error biases up to 200 % may be observed. Power levels based on the unadjusted linear mixed model are often too low. The consequences are most severe for large clusters and/or small intraclass correlation coefficients since then the required number of clusters to achieve a desired power level is smallest.

Conclusions: The possibility of covariate imbalance should be taken into account while calculating the sample size of a cluster randomized trial. Otherwise more sophisticated methods to randomize clusters to treatments should be used, such as stratification or balance algorithms. All relevant covariates should be carefully identified, be actually measured and included in the statistical model to avoid severe levels of parameter and standard error bias and insufficient power levels.

Keywords: Cluster randomization, Covariate imbalance, Unadjusted linear mixed model, Adjusted linear mixed model, Simulation study

* Correspondence: M.Moerbeek@UU.NL
Department of Methodology and Statistics, Utrecht University, P.O. Box
80140, 3508 TC Utrecht, The Netherlands



Background

Randomized controlled trials are considered the gold standard for evaluating the effect of a new treatment relative to an old treatment, a placebo or no treatment at all. A necessary (but not sufficient) condition for the estimate of the treatment effect to be unbiased is that subjects in a given treatment group are not contaminated by those in another. This condition is often not fulfilled in trials that are offered in naturally existing groups, such as schools, general practices or communities, especially so when the treatments rely on interpersonal communication in risk reduction sessions and peer pressure groups [1]. For this reason, cluster randomization is often preferred above individual randomization. With cluster randomized trials complete clusters are randomized to treatment conditions and all subjects within a cluster receive the same treatment. So, cluster randomized trials may reduce logistical and administrative costs. Community interventions, such as mass media interventions using TV, are directed towards the whole cluster and cannot be delivered at the individual level. In such trials there is no alternative to cluster randomization. Cluster randomization has become popular in the medical, health and behavioural sciences over the past decades and relevant textbooks on this research design are [2–6].

The number of clusters in a cluster randomized trial is often low [7]. As a consequence, random assignment of clusters to treatment conditions does not ensure the treatment groups are comparable at baseline with respect to all variables at the subject and cluster level that have an effect on the outcome variable. In other words, it is likely there is covariance imbalance at baseline. It is therefore good practice to identify such variables based on findings in the literature or experts' expectations, to actually measure them and to include them as covariates in an adjusted linear mixed model. As such, an unbiased estimate of the effect of treatment relative to the control can be obtained. In addition to that, part of the residual variance at the cluster and/or subject level is explained by the covariate, which has an advantageous effect on precision and power [8], especially so when the covariate has a strong effect on the outcome. The drawback, however, is that this comes at the cost of degrees of freedom, which may have a decreasing effect on the power of the test on treatment effect, especially when the number of clusters is small. Furthermore, it is known that less efficient estimates of the treatment effect, and hence lower power levels, are obtained with larger degrees of covariate imbalance [9, 10]. Finally, covariate imbalance should not only be taken into account while analysing the data but also while performing an a priori power analysis to calculate the required number of clusters to achieve a desired power level. Ignoring the possibility of covariate imbalance results in an underestimate of the required

number of clusters and hence insufficient power, even when the covariates are adjusted for in an adjusted linear mixed model.

In practice it often occurs not all relevant covariates are known and/or not all of them have been measured. As a consequence, type I or type II error rates may be inflated. The degree of inflation depends on the precision of the treatment effect estimator, which is affected by the strength of the covariate effect and degree of covariate imbalance. A recent study quantified the effect of covariate imbalance on bias and precision of the treatment effect estimator and the power of the test on treatment effect for simple randomized trials [11]. With such trials subjects are randomly assigned to treatment conditions and there is no nesting of subjects within clusters. One continuous covariate was considered and the adjusted linear model was compared to the unadjusted linear model, which does not include the covariate. The conclusion was that, while the adjusted linear model produces unbiased estimates of the treatment effect, the unadjusted linear model is subject to bias. The unadjusted linear model was found to be less precise than the adjusted linear model. Finally it was found that the power of the unadjusted linear model can be much larger than the anticipated level at certain levels of imbalance.

The number of clusters in a cluster randomized trial is typically much lower than the number of subjects in a simple randomized trial, hence covariance imbalance is more likely to occur in cluster randomized trials than in simple randomized trials. For this reason it is worthwhile to extend the study in reference [11] to cluster randomized trials. We study to what extent a loss of power is achieved when the power calculation is done under the assumption of covariate balance, while the randomization process results in covariate imbalance. In other words, we study the loss of power when the unadjusted linear mixed model is assumed in an a priori power analysis but the adjusted linear mixed model is used for data analysis. In addition to that, we investigate the effects of ignoring relevant covariates while analysing the data. Stated differently, we study the size of the parameter and standard error bias when the unadjusted linear mixed model is used to analyse the data while the data are generated by an adjusted linear mixed model. Also, we study the power levels achieved with the unadjusted linear mixed model.

The focus is on a covariate at the cluster-level. Such a covariate may be a cluster-level characteristic, such as school-type (e.g. public versus private) in school-based intervention studies, or an aggregate, such as a school's mean socio-economic status. Associations between the covariate and outcome tend to be much stronger between than within clusters [12, 13]. Furthermore, it is often less expensive and less time-consuming to measure

cluster-level covariates than subject-level covariates, simply because the number of clusters is always (much) smaller than the number of subjects.

The data obtained from cluster randomized trials have a multilevel data structure with subjects nested within clusters. Such data have to be analysed by using the mixed model. Ignoring the multilevel data structure results in too liberal tests with respect to the effect of treatment [14]. Our approach to study the effects of covariate imbalance is a simulation study in which various sizes of the covariate effect and degrees of covariance imbalance are considered. Cluster size and the intraclass correlation coefficient (i.e. the proportion variance at the cluster level) are known to have an effect on precision and power in cluster randomized trials [15], hence various values of these two factors are considered as well.

Mixed model for cluster randomized trials

To put things in context a school-based intervention study to evaluate the effect of a new type of smoking prevention intervention on smoking behaviour of secondary school students is used as an example. Secondary schools are randomly assigned to treatment conditions and all students within a school receive the same condition. The covariate we consider is educational level; it is binary and measured at the level of the secondary school: vocational schools train their students for specific vocations while high schools prepare their students for follow-up education.

The outcome we consider is continuous with higher scores related to higher levels of smoking abstinence. It is very likely scores of students within the same school are correlated, for instance as a result of mutual influence and shared norms towards smoking. Ignoring such correlation results in overoptimistic conclusions with respect to the effect of intervention [14]. To account for such dependencies the adjusted linear mixed model should be used. The score y_{ij} of student i in school j is given by

$$y_{ij} = \beta_0 + \beta_1 x_j + \beta_2 z_j + u_j + e_{ij}. \tag{1}$$

Treatment condition is denoted x_j and has the value 0 for the control and 1 for the intervention. The covariate is denoted z_j and has the value 0 for vocational schools and 1 for high schools. Note that both x_j and z_j have a subscript j but not i since they vary between but not within schools. The regression weights β_0 , β_1 and β_2 are the intercept and the effect of the intervention and covariate, respectively. The random school-level effect u_j is the discrepancy of the mean in school j and the mean score in its treatment condition and school type, and the random subject-level error term e_{ij} is the discrepancy of the score of student i and the mean score of its school j .

Both random effects are assumed to be independent of each other and to follow a normal distribution with zero mean and variances σ_u^2 and σ_e^2 . The total error variance $\sigma^2 = \sigma_u^2 + \sigma_e^2$ is the sum of these two variance components.

On the basis of the adjusted linear mixed model (1) the treatment effect is the difference in expected outcomes across the two treatments, corrected for the covariate:

$$\beta_1 = \mu_{y_i} - \mu_{y_c} - \beta_2(\mu_{z_i} - \mu_{z_c}). \tag{2}$$

Here, μ_{y_i} and μ_{y_c} are the expected outcome scores in the intervention and control condition, and μ_{z_i} and μ_{z_c} are the expected covariate scores in the intervention and control conditions. The expectations μ can be replaced by their observed means as calculated from a sample to obtain an estimate of the treatment effect. The standard error of this estimator is

$$se(\hat{\beta}_1) = \sqrt{\frac{\sigma_e^2 + n_1 \sigma_u^2}{n_1 n_2 (1 - \rho_{xz}^2)}}, \tag{3}$$

with n_1 and n_2 the common school size and number of schools, and ρ_{xz} the correlation between treatment condition and the covariate. The larger this correlation, the larger the degree of covariate imbalance. Eq. (3) shows that the standard error $se(\hat{\beta}_1)$ becomes larger when ρ_{xz}^2 increases. In other words, covariate imbalance results in a loss of efficiency. The factor $1/(1 - \rho_{xz}^2)$ is the variance inflation factor and it is the same equation as for designs without nesting of subjects in clusters [9]. The standard error (3) is estimated by replacing the variance components and the correlation between treatment and the covariate by their estimates $\hat{\sigma}_e^2$ and $\hat{\sigma}_u^2$ and r_{xz} .

The test statistic for the test on treatment effect is calculated as $t = \hat{\beta}_1 / se(\hat{\beta}_1)$ and follows a central t distribution with $n_2 - 3$ degrees of freedom under the null hypothesis of no treatment effect $H_0: \beta_1 = 0$. The statistical power is the probability of rejecting the null hypothesis when it is false. The number of clusters to achieve a power level of at least $1 - \beta$ in a test with a two-sided alternative hypothesis and type I error rate α is calculated as

$$n_2 = 4 \frac{\sigma_e^2 + n_1 \sigma_u^2}{n_1 (1 - \rho_{xz}^2)} \left(\frac{z_{1-\alpha/2} + z_{1-\beta}}{\beta_1} \right)^2. \tag{4}$$

This will always give a non-integer value which has to be rounded upwards to the nearest integer to achieve the desired power level.

In practice it often occurs not all relevant covariates are measured, hence they cannot be included into the

regression model. The unadjusted linear mixed model ignores the covariate z_j :

$$y_{ij} = \beta_0^* + \beta_1^* x_j + u_j^* + e_{ij}^* \tag{5}$$

The model parameters are now indicated by an asterisk to distinguish them from their counterparts in the adjusted linear mixed model (1). The error variances also have an asterisk: σ_e^{*2} and σ_u^{*2} and again the errors are assumed normally distributed with zero mean. As z_j varies between but not within schools, it only explains part of the error variance at the between-school level, hence $\sigma_e^{*2} = \sigma_e^2$, $\sigma_u^{*2} > \sigma_u^2$ and $\sigma_u^{*2} + \sigma_e^{*2} > \sigma_u^2 + \sigma_e^2$.

On the basis of the unadjusted linear mixed model the treatment effect is the difference in expected outcomes across the treatments:

$$\beta_1^* = \mu_{y_I} - \mu_{y_C} \tag{6}$$

Here, μ_{y_I} and μ_{y_C} are the expected outcomes in the intervention and control condition. Replacing the expectations μ by their observed means as calculated from a sample provides an estimate $\hat{\beta}_1^*$ of the treatment effect. A comparison of Eqs. (2) and (6) shows that ignoring the covariate results in a biased estimate of the effect of treatment, unless the covariate has no effect on the outcome (i.e. $\beta_2 = 0$) or it is balanced across treatment conditions (i.e. $\mu_{z_I} = \mu_{z_C}$). The standard error of the estimator $\hat{\beta}_1^*$ is

$$se(\hat{\beta}_1^*) = \sqrt{\frac{\sigma_e^{*2} + n_1 \sigma_u^{*2}}{n_1 n_2}} \tag{7}$$

and it is estimated by replacing the variance components by their estimates $\hat{\sigma}_e^{*2}$ and $\hat{\sigma}_u^{*2}$. The test statistic for the test on treatment effect is calculated as $t = \hat{\beta}_1^* / se(\hat{\beta}_1^*)$ and follows a central t distribution with $n_2 - 2$ degrees of freedom under the null hypothesis of no treatment effect $H_0: \beta_1^* = 0$. Note that there is one more degree of freedom than in the adjusted linear mixed model since the effect of the covariate is not estimated.

For simplicity it is assumed that half of the schools are vocational schools and the others are high schools. Furthermore, half of the schools are randomly assigned to receive the intervention, which implies the other half receives the control. Let us assume the intervention results in higher outcome scores (i.e. $\beta_1 > 0$) and that higher outcome scores are observed in high schools than in vocational schools (i.e. $\beta_2 > 0$). Let us define positive imbalance as a situation where, by chance, the random assignment results in more than half of the schools in the intervention condition being high schools. This would imply ignoring the covariate in the statistical model

results in an overestimated treatment effect. Negative imbalance is the opposite situation and the treatment effect is underestimated.

The number of high schools in the intervention condition follows from a hypergeometric distribution with parameters n_2 (the total number of schools), $\frac{1}{2}n_2$ (the number of high schools) and $\frac{1}{2}n_2$ (the number of schools that are randomized to receive the intervention). The expectation of the hypergeometric distribution is $\frac{1}{4}n_2$, which implies that both treatment groups consist of as many high schools as vocational schools. The degree of covariate imbalance is expressed by quantiles from the hypergeometric distribution rather than the percentage high schools in the intervention condition. As such we adopt the approach by Egbewale, Lewis and Sim [11] to express the degree of covariate imbalance. The percentage high schools in the intervention condition for a given quantile depends on the total number of schools. For the 95th quantile, for instance, this percentage is equal to 70 when the total number of schools is 20, and it is equal to 56 when the total number of schools is 200. This illustrates large percentage deviations from the expectation are more likely to occur for a small number of clusters than for a large number of clusters.

Methods

A simulation study was conducted to gain further insight in the size of the effects of covariate imbalance in cluster randomized trials. Data generation and parameter estimation were done in R [16]. Parameter estimation was done using the function `lme` from the package `nlme` for linear and non-linear mixed effect models [17]. Estimation was done by means of restricted maximum likelihood since it produces less biased estimates of the variance components than full information maximum likelihood, especially when the number of clusters is small.

Four factors are considered in the simulation study: cluster size, intraclass correlation coefficient, covariate effect size and degree of covariate imbalance. For cluster size the values 5, 30 and 50 were used. These values are common in respectively households, school classes and departments in hospitals or companies, and were also used by Maas and Hox [18]. Note that the cluster size did not vary across clusters; the implication of varying cluster sizes is discussed later on. The intraclass correlation coefficient ρ measures the proportion of the variance that is located at the cluster level. It is calculated as $\rho = \sigma_u^2 / (\sigma_u^2 + \sigma_e^2)$, and the values 0.01, 0.05, and 0.1 were used in this simulation study. These values cover a wide range of plausible values, with small

values common in large clusters, such as hospitals, and large values common in small clusters, such as households. Without loss of generality the total error variance was fixed to $\sigma^2 = \sigma_u^2 + \sigma_e^2 = 1$. The covariate effect size β_2 had levels 0.2, 0.5, and 0.8, which correspond to small, medium and large effect sizes [19]. We do not consider the case where the covariate has no effect (i.e. $\beta_2 = 0$) because the consequences of ignoring covariate imbalance are expected to be negligible. The degree of covariate imbalance is expressed as quantiles of the hypergeometric distribution and the following values were used: 0.025, 0.05, 0.1, 0.5, 0.9, 0.95, and 0.975. Values above 0.5 imply positive imbalance (i.e. too many high schools in the intervention condition). The value 0.5 implies perfect balance of the covariate across the treatments and is also considered.

In total $3 \times 3 \times 3 \times 7 = 189$ combinations of factor levels were considered and these are called simulation conditions henceforward. For each simulation condition 5000 data sets were generated from the adjusted linear mixed model and analysed by means of the adjusted linear mixed model and unadjusted linear mixed model. For each combination of the cluster size and intraclass correlation coefficient the required number of clusters to achieve a power level $1 - \beta = 0.8$ to detect a medium treatment effect $\beta_1 = 0.5$ in a test with a two-sided alternative hypothesis and $\alpha = 0.05$ was calculated from (4). This gives a non-integer value which is rounded upwards such that an even number of clusters is available per treatment condition. Only when an even number of clusters per treatment is used balance of the binary covariate at the cluster level can be achieved. This is important because we want to include perfect balance of the covariate across the treatments in our simulation study (see above). Due to rounding upwards to integer values the nominal power levels were somewhat larger than 0.8, and they also varied across the simulation conditions. In practice the possibility of covariate imbalance is often ignored while calculating the required number of clusters. The required number of clusters is therefore calculated under the assumption of covariate balance (i.e. $\rho_{xz} = 0$), but data are generated under various degrees of imbalance and covariate effect size to study the effect of covariate imbalance on parameter estimates and empirical power.

Various criteria were used to evaluate the effects of covariate imbalance. The primary focus was on the treatment effect, since this is the parameter of main interest in intervention research. The percentage parameter bias is defined as

$$parameter\ bias = 100\% * \frac{\overline{\hat{\beta}_1} - \beta_1}{\beta_1}, \tag{8}$$

where $\overline{\hat{\beta}_1}$ is the estimated treatment effect averaged across the 5000 generated data sets and the population value of β_1 is 0.5 in this simulation study. For the unadjusted linear mixed model $\overline{\hat{\beta}_1}$ is replaced by $\hat{\beta}_1^*$. Positive biases indicate the treatment effect is overestimated, which may result in too liberal tests on treatment effect and hence inflated type I error rates. Acceptable biases are below 10 % in absolute value [20].

The second criterion for evaluation is the percentage standard error bias:

$$standard\ error\ bias = 100\% * \frac{\overline{se(\hat{\beta}_1)} - sd(\hat{\beta}_1)}{sd(\hat{\beta}_1)}. \tag{9}$$

Here the mean standard error of the treatment effect estimates is compared to the standard deviation of the 5000 treatment effect estimates. Positive bias implies the standard errors are overestimated, which may result in a loss of power. Acceptable biases are below 5 % in absolute value [20].

The third criterion is the empirical power level. For each simulation condition it is approximated as the percentage of data sets for which the null hypothesis of no treatment effect was rejected in a test with a two-sided alternative hypothesis and type I error rate $\alpha = 0.05$.

Results

For all simulation conditions the parameter bias of the adjusted linear mixed model was negligible: the largest bias was -1.1 % (figures not shown). For the unadjusted linear mixed model the parameter biases are much larger because it ignores the covariate and in the worst cases these biases are slightly over 100 %, as is depicted in Fig. 1. The lines cut when the covariate is balanced across treatments (i.e. when covariate imbalance = 0.5). For negative covariate imbalance the bias is negative, and vice versa for positive covariate imbalance. In terms of the example: the effect of treatment is underestimated when there are too many vocational schools in the intervention condition, and overestimated when there are too many high schools in the intervention condition. In general the biases increase with increasing degree of covariate imbalance and covariate effect size. This also follows from a comparison of Eqs. (2) and (6). The bias is $\beta_2(\mu_{z_1} - \mu_{z_c})$ and increases with covariate effect size β_2 and the difference between expected covariate scores μ_{z_1} and μ_{z_c} in both treatment conditions. Furthermore, the bias increases with increasing cluster size. The effect of covariate imbalance becomes slightly less pronounced when the intraclass correlation increases.

For the adjusted linear mixed model the standard error biases were small and the largest bias was -5.1 % (figures not shown). The standard error biases are much

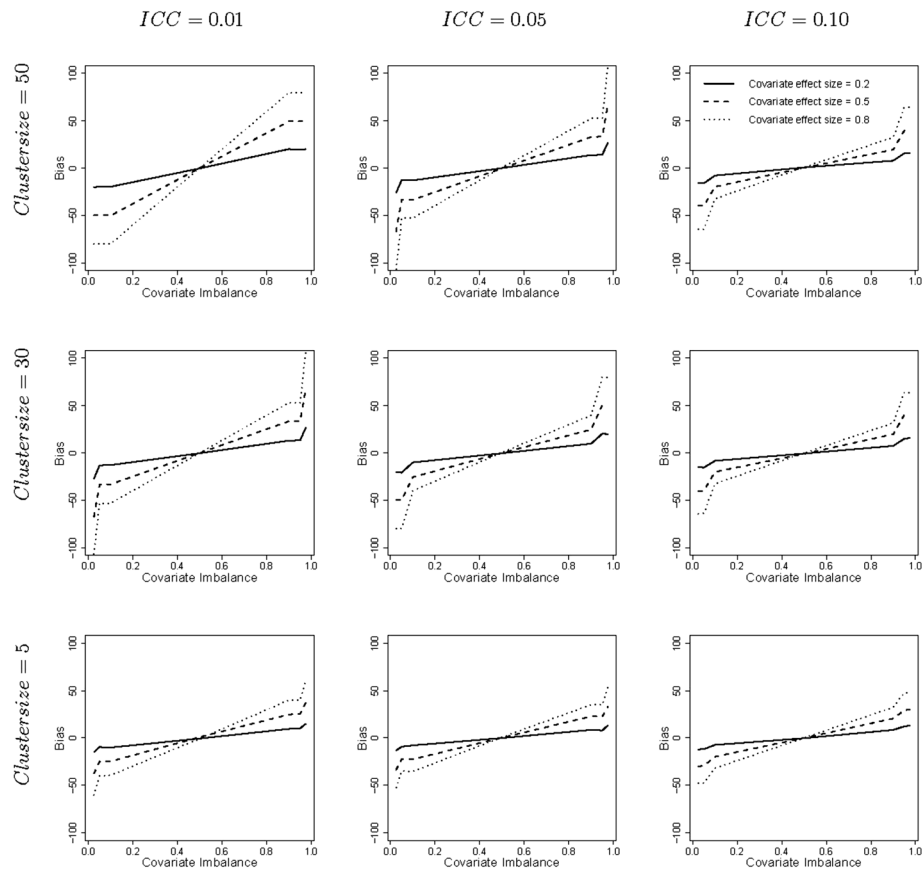


Fig. 1 Parameter bias for the linear mixed model. Legend: The degree of covariate imbalance is given on the horizontal axes and expressed as quantiles of the hypergeometric distribution. Values below 0.5 imply negative covariate imbalance, values above 0.5 imply positive covariate imbalance and the value 0.5 implies covariate balance

larger for the unadjusted linear mixed model and up to almost 200 %, see Fig. 2. The standard error bias is always positive since the unadjusted linear mixed model overestimates the variance component at the cluster level, and hence it overestimates the standard error of the treatment effect estimate (see Eq. (7)). The standard error bias increases with increasing covariate effect size and increasing cluster size. These effects were also found for the parameter bias. For large cluster size $n_1 = 50$ the standard error bias decreases with increasing intraclass correlation coefficient but this effect becomes weaker when the cluster size decreases. Surprisingly the standard error bias is most often largest when the covariate is uncorrelated with treatment.

Figure 3 shows the empirical power levels as achieved by the adjusted linear mixed model. Again note that in this simulation study the number of clusters to achieve 80 % power was calculated under the assumption of covariate balance while the data were generated for various levels of covariate imbalance. The number of clusters as calculated from Eq. (4) is often a non-integer and has to

be rounded upwards to ensure a power of at least 80 %. As a consequence the nominal power is somewhat above 80 %, especially in the cases where the cluster size is large and/or the intraclass correlation coefficient is small. The nominal power is shown by the horizontal line within each subplot. The largest empirical power levels are observed in case of covariate balance and these are very close to the nominal power levels. Power decreases as the degree of covariate imbalance increases because the standard error of the treatment effect increases with increasing degree of covariate imbalance (see Eq. (3)). This decrease is largest for large cluster size and small intraclass correlation coefficient. In the worst case a loss of power of almost 25 % is observed. The covariate effect size has a negligible effect on empirical power levels.

Figure 4 shows the power levels for the unadjusted linear mixed model are much different from those for the adjusted linear mixed model. For negative imbalance the power can be much below the 80 % level. This can be explained by an underestimation of the treatment effect (see Fig. 1) and an overestimation of the related standard error (see Fig. 2). In these cases the largest

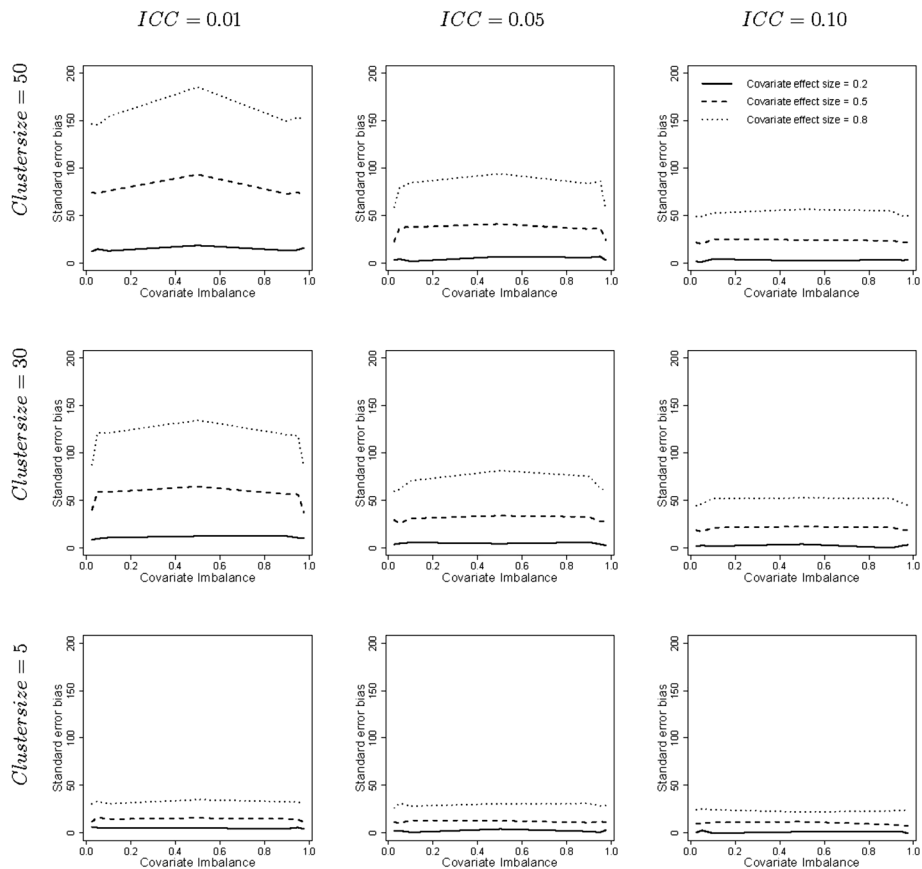


Fig. 2 Standard error bias for the linear mixed model. Legend: The degree of covariate imbalance is given on the horizontal axes and expressed as quantiles of the hypergeometric distribution. Values below 0.5 imply negative covariate imbalance, values above 0.5 imply positive covariate imbalance and the value 0.5 implies covariate balance

power is achieved for the smallest covariate effect size. For large values of positive imbalance, the overestimate of the treatment effect compensates the overestimate of the related standard error and the power level is above 80 %. Furthermore it should be noted that an absence of covariate imbalance does not imply the power level is 80 %. Finally, it should be noted that, as in the previous figures, the strongest effects of covariate imbalance are generally observed for large cluster size and small intra-class correlation coefficient.

Discussion

In this simulation study we focussed on cluster randomized trials with a continuous outcome at the individual level and a binary outcome at the cluster level. Our findings are in accordance with those for simple randomized trials with a continuous outcome [11]. Covariate imbalance has a negligible effect on parameter bias and a small effect on standard error bias when the adjusted linear mixed model is used. Ignoring covariate balance while calculating the required number of clusters may

result in a loss of power up to 25 % for the adjusted linear mixed model. It is therefore good practice to account for covariate imbalance while calculating the number of clusters or to use more sophisticated methods to randomize clusters to treatment conditions, for instance stratification or balance algorithms, such as the one in [21].

Covariate imbalance has much more severe consequences when the covariate is ignored while analysing the data. For the unadjusted linear mixed model the treatment effect was overestimated or underestimated by at most 100 % and its standard error was overestimated by at most 200 % in our simulation study. In most cases the power levels were below 80 % but for very large levels of positive imbalance they were above 80 %. This stresses the need to carefully consider which covariates are important in the study at hand, to actually measure them and to include them in the statistical model. A new tool based on propensity scores is useful to detect covariate imbalances [22].

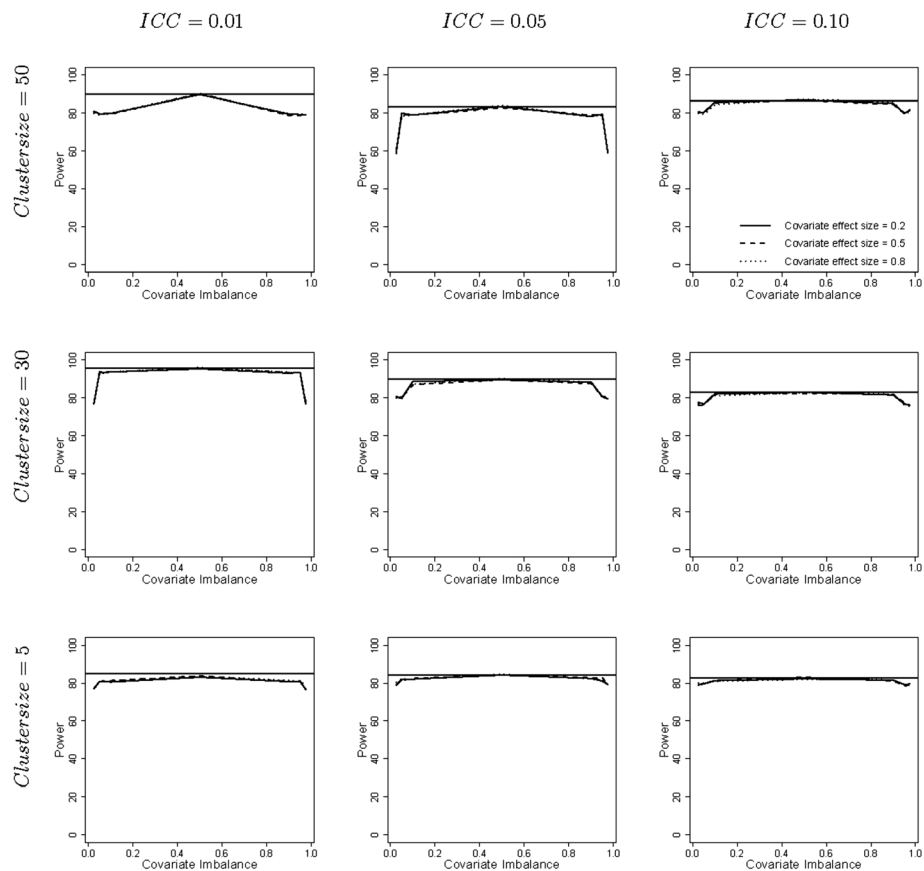


Fig. 3 Empirical power levels for the adjusted linear mixed model. Legend: The degree of covariate imbalance is given on the horizontal axes and expressed as quantiles of the hypergeometric distribution. Values below 0.5 imply negative covariate imbalance, values above 0.5 imply positive covariate imbalance and the value 0.5 implies covariate balance. The horizontal solid line represents the nominal power level

For both models and for any of the three criteria of evaluation it was observed that the strongest results occurred when the cluster size was large and the intraclass correlation coefficient was small. This is not surprising because in these cases the number of clusters to achieve 80 % power is smallest. The smaller the number of clusters the more likely the percentage high schools in the intervention condition deviates from 50 %. So one should take extra care in trials with large clusters and small degree of dependence among the subjects within a cluster. For instance, covariate imbalance can result in more severe consequences when the cluster randomized trial has general practices as unit of randomization rather than households.

In our simulation we assumed all clusters were of equal size. In practice cluster sizes vary and this can be compensated by adding 11 % more clusters [23]. As we argued above, a larger number of clusters results in weaker effects of cluster imbalance.

We acknowledge the setup of our simulation study is somewhat limited. We considered a continuous outcome variable, a single binary covariate at the

cluster level, both levels of the covariate were equally represented, and both treatment groups were of equal size. It would also be interesting to consider an individual-level covariate, such as a pre-treatment score or a socio-demographic variable. We expect the potential implications of imbalance of an individual-level covariate to depend on the between- and within-level variability of such a covariate. The higher the between-level variability as compared to the within-level variability, the more the implications of covariate imbalance will correspond to the implications of covariate imbalance of a covariate at cluster level. On the other extreme, covariate imbalance cannot occur when the individual-level covariate only varies within clusters but not between clusters because treatment condition is a cluster level variable. In that case both the unadjusted linear mixed model and adjusted linear mixed model will yield an unbiased estimate of the treatment effect size, but the associated standard error will be smaller for the adjusted linear mixed model because part of the within-cluster variance is explained by the covariate. The implications of imbalance of an individual-

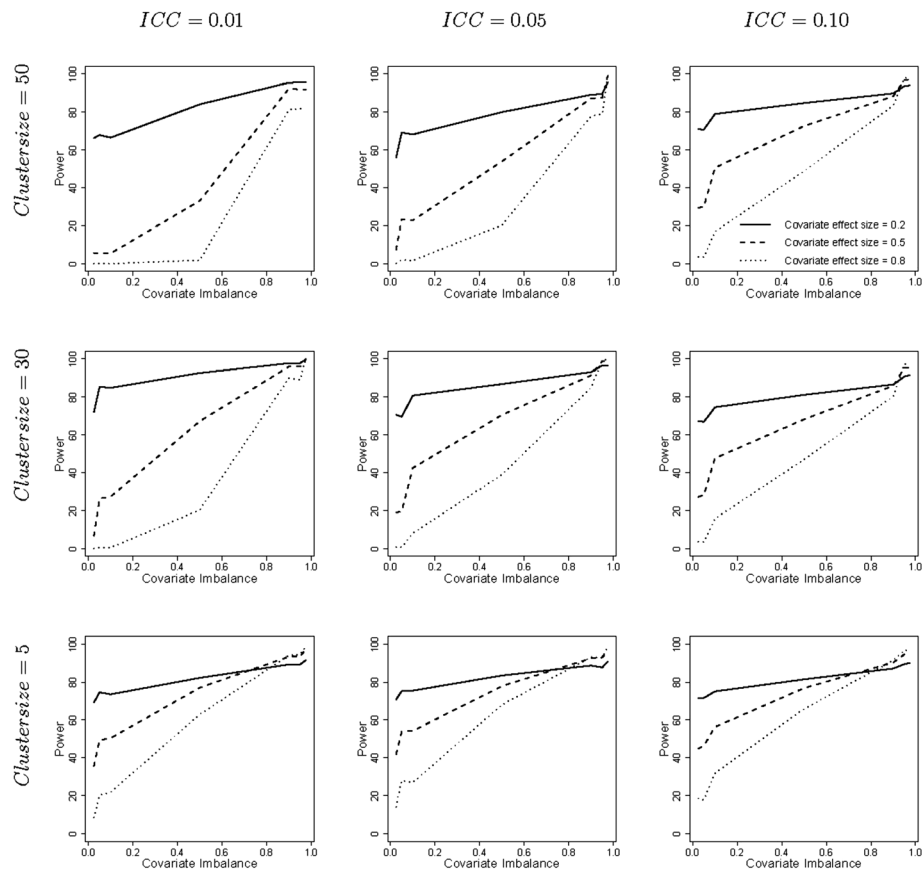


Fig. 4 Empirical power levels for the linear mixed model. Legend: The degree of covariate imbalance is given on the horizontal axes and expressed as quantiles of the hypergeometric distribution. Values below 0.5 imply negative covariate imbalance, values above 0.5 imply positive covariate imbalance and the value 0.5 implies covariate balance

level covariate that varies both between and within clusters will be between these two extremes and will have to be studied in future research. Future research could also focus on trials with more covariates which can be measured at the subject and/or cluster level and which can be measured as categorical or scale variables. In any case we expect a loss of power in adjusted linear mixed models and parameter and standard error biases in unadjusted linear mixed models. As for now, it remains unclear how large the effects may be.

Conclusion

We can conclude that covariate imbalance may result in a loss of power in adjusted linear mixed models. It is therefore advocated to be aware of the possibility of covariate imbalance while calculating the number of clusters, or to apply stratified randomization or balance algorithms. In addition, ignoring relevant covariates while analysing the data may lead to severely biased estimates of the treatment effect and its standard error. It is therefore important all relevant covariates are identified, measured and included in the regression model.

This recommendation holds for simple randomized trials, but is even more important for cluster randomized trials because for this type of trials the number of units to be randomized is often low.

Acknowledgements

None.

Funding

There are no sources of funding to be declared.

Availability of data and materials

This research is based on a simulation study. R code to generate data can be obtained from <https://surfdrive.surf.nl/files/index.php/s/Mp3IVNNju4S7DZy>.

Authors' contributions

MM conceived and designed the study, interpreted the data and drafted the manuscript. SvS performed the simulations, interpreted the data and helped to draft the manuscript. Both authors read and approved the final manuscript.

Authors' information

MM is an associate professor at Utrecht University, the Netherlands. She is an expert in statistical power analysis and optimal experimental design, mainly in the field of cluster randomized trials. SvS is a lecturer at the same university. He wrote his master thesis on the use of internal pilots in cluster randomized trials.

Competing interests

The authors declare that they have no competing interests.

Consent for publication

Not applicable.

Ethics approval and consent to participate

Not applicable.

Received: 26 May 2016 Accepted: 25 June 2016

Published online: 11 July 2016

References

- Moerbeek M. Randomization of clusters versus randomization of persons within clusters: which is preferable? *Am Stat.* 2005;59:72–8.
- Campbell MJ, Walters SJ. How to design, analyse and report cluster randomised trials in medicine and health related research. Chichester: Wiley; 2014.
- Donner A, Klar N. Design and analysis of cluster randomization trials in health research. London: Edward Arnold; 2000.
- Eldridge M, Kerry S. A practical guide to cluster randomised trials in health services research. Chichester: Wiley; 2012.
- Hayes RJ, Moulton LH. Cluster randomised trials. Boca Raton: CRC Press; 2009.
- Murray DM. Design and analysis of group-randomized trials. New York: Oxford University Press; 1998.
- Ivers NM, Taljaard M, Dixon S, Bennett C, McRae A, Taleban J, Skea Z, Brehaut JC, Boruch RF, Eccles MP, Grimshaw JM, Weijer C, Zwarenstein M, Donner A. Impact of CONSORT extension for cluster randomised trials on quality of reporting and study methodology: review of random sample of 300 trials, 2000–8. *BMJ.* 2011; 343. doi:10.1136/bmj.d5886.
- Moerbeek M. Power and money in cluster randomized trials: when is it worth measuring a covariate? *Stat Med.* 2006;25:2607–17.
- Lipsitz SR, Parzen M. Sample size calculations for non-randomized studies. *The Statistician.* 1995;44:81–90.
- Moerbeek M, Van Breukelen GJP, Berger MPF. Optimal experimental designs for multilevel models with covariates. *Commun Stat Theor M.* 2001;30:2683–97.
- Egbewale BE, Lewis M, Sim J. Bias, precision and statistical power of analysis of covariance in the analysis of randomized trials with baseline imbalance: a simulation study. *BMC Med Res Methodol.* 2012;14:49.
- Bloom H. Randomizing groups to evaluate place-based programs. In: Bloom H, editor. Learning more from social experiments. Evolving analytic approaches. Randomizing groups to evaluate place-based programs. New York: Russell Sage; 2005. p. 115–72.
- Raudenbush SW, Martinez A, Spybrook J. Strategies for improving precision in group-randomized experiments. *Educ Eval Policy An.* 2007;29:5–29.
- Moerbeek M, Van Breukelen GJP, Berger MPF. A comparison between traditional methods and multilevel regression for the analysis of multi-center intervention studies. *J Clin Epidemiol.* 2003;56:341–50.
- Moerbeek M, Van Breukelen GJP, Berger MPF. Design issues for experiments in multilevel populations. *J Educ Behav Stat.* 2000;25:271–84.
- R Core Team. R: a language and environment for statistical computing. Vienna: R Foundation for Statistical Computing; 2014. Available: <http://www.R-project.org/>. Accessed 1 May 2016.
- Pinheiro JC, Bates DM. Mixed-effects models in S and S-PLUS. New York: Springer; 2004.
- Maas C, Hox JJ. Sufficient sample sizes for multilevel modelling. *Methodology.* 2005;1:86–92.
- Cohen J. A power primer. *Psychol Bull.* 1992;112:155–9.
- Muthén LK, Muthén BO. How to use a Monte Carlo study to decide on sample size and determine power. *Struct Equ Modeling.* 2002;9:599–620.
- De Hoop E, Teerenstra S, Van Gaal BG, Moerbeek M, Borm G. The “best balance” allocation led to optimal balance in cluster-controlled trials. *J Clin Epidemiol.* 2012;65:132–7.
- Leyrat C, Caille A, Foucher Y, Giraudeau B. Propensity score to detect baseline imbalance in cluster randomized trials: the role of the c-statistic. *BMC Med Res Methodol.* 2016;16:9.
- Van Breukelen GJP, Candel MJJM, Berger MPF. Relative efficiency of unequal versus equal cluster sizes in cluster randomized and multicentre trials. *Stat Med.* 2007;26:2589–603.

Submit your next manuscript to BioMed Central and we will help you at every step:

- We accept pre-submission inquiries
- Our selector tool helps you to find the most relevant journal
- We provide round the clock customer support
- Convenient online submission
- Thorough peer review
- Inclusion in PubMed and all major indexing services
- Maximum visibility for your research

Submit your manuscript at
www.biomedcentral.com/submit

