

UNIVERSITY OF TWENTE.

Department of Research Methodology, Measurement and Data Analysis (OMD)

**Small and negative correlations among
clustered observations: A simulation study
about limitations of the linear mixed effects
model**

28th June 2019

Natalie Nielsen

1819194

First supervisor: Jean-Paul Fox

Second supervisor: Wouter A. C. Smink

ABSTRACT

Usually, the linear mixed effects model is used for analysing clustered data. The model assumes that observations in clusters are positively correlated. When there is a true negative correlation within groups, the model assumes incorrectly that observations are independently distributed. A simulation study was conducted to report about the biases that occur when ignoring negative correlation within groups. The results show that, when ignoring the negative correlation, the Type-I error is deflated, and standard errors and p-values are overestimated. This resulted in an underestimation of the difference between group means and the population mean. These findings indicate that researchers need to be aware that negative correlation between observations in clustered data may occur. Also, it is important to watch out for biases when ignoring negative correlation within groups and incorrectly assuming that observations are independently distributed. Altogether, this study shows that the linear mixed effects model is an inappropriate tool for analysing clustered data containing negative correlation within groups.

Keywords: linear mixed effects model, negative correlation, negative ICC, Type-I error

INTRODUCTION

Within-group and between-group correlations are of importance when analysing clustered data. The intraclass correlation (ICC) coefficient is an indicator of the correlation that exists due to a grouping structure in a sample (Hox, 2010). In the literature, it is interpreted as the proportion of variance that is explained by the grouping of observations in clusters, which is computed as the ratio of the between-group variance to the total variance (Giberson, Resick, & Dickson, 2005; Kenny, Mannetti, Pierro, Livi, & Kashy, 2002). When modelling this correlation as variance of independent group means the range of possible correlation values is restricted to a lower bound of zero because variances cannot be negative. However, a negative ICC can occur when for instance a fixed resource is divided among group members, at non-random sampling when dissimilar groups are sampled by intention, or when there is competitive social interaction (Kenny & Judd, 1986; Kenny et al., 2002).

Researchers use different strategies to control for correlations in clustered observations. On the one hand, a linear mixed effects (LME) model is used to model positive correlation, and it does not allow negative correlation (Eldridge, Ukoumunne, & Carlin, 2009). The LME model assumes independently distributed observations given a random (clustering) effect. Further, it assumes positively correlated data within clusters. When observations within a group are negatively correlated, the LME model will ignore the correlation and assume independently distributed observations. If not done by the model, some researchers fix negative ICC estimates to zero and therefore, ignore the negative correlation within a group (Baldwin, Stice, & Rohde, 2008; Maas & Hox, 2005). On the other hand, researchers tend to give alternative explanations for negative ICC estimates. Most researchers explain the estimate based on the fact that greater within-group variance than between-group variance leads to negative ICC values (Giberson et al., 2005; Krannitz, Grandey, Liu, & Almeida, 2015; Langfred, 2007). Other researchers, for example Edridge et al. (2009), relate negative ICC values to sampling error. In all cases, negative ICC estimates are ignored, so that effects of negative clustering are not included in the analysis.

In general, ignoring a small positive or (small) negative clustering effect can lead to violations of the LME model assumptions. When assuming incorrectly that the observations are independently distributed, Type-I errors will be inaccurate, which is might increase the risk of accepting the

alternative hypothesis when the null hypothesis is true (Clarke, 2008). There are two conditions that lead to an inaccurate Type-I error: (1) ignoring (small) positive intra-class correlations, and (2) ignoring (small) negative intra-class correlations.

Regarding the first condition, Barcikowski (1981) examined the effects of ignoring small positive correlations in clustered observations in a two-level study design including a group level and an individual level. He showed that, when having ten observations per group, even the ignorance of an ICC of 0.01 can lead to a slight inflation of making a Type-I error. This means that when ignoring a small positive clustering effect, a fixed effect will be assumed to be significant with a significance level of 5% although the true significance level equals 6% (Barcikowski, 1981). Other researchers have also demonstrated that ignoring a positive ICC leads to an inflation of the Type-I error and confidence intervals (CI) that are too narrow (Clarke, 2008; Rosner & Grove, 1999).

Looking at the effects in more detail, Barcikowski (1981) showed that a Type-I error increases for increasing values of an ICC. For example, for an ICC of .05 the Type-I error rate is .11 and for an ICC of .40 the Type-I error is .46. Moreover, by increasing the number of observations per group the Type-I error is even more inflated (Barcikowski, 1981; Dorman, 2008). Also, Dorman (2008) showed that a larger number of clusters lead to a smaller inflation of a Type-I error. The inflation of Type-I error can be explained by the variance inflation factor (VIF), also called design effect. The standard error (SE) of a fixed effect will increase when the VIF increases (Baldwin, Murray, & Shadish, 2005). The VIF is computed as $VIF = 1 + (\text{average cluster size} - 1) * ICC$ and is positively related to the variance of the sample mean (Norton, Bieler, Ennett, & Zarkin, 1996). When ignoring a small positive ICC, the VIF is underestimated and so is the SE of a fixed effect. Therefore, the effect size is incorrect as it is calculated from group means and their SEs (Hox, 2010; Kenny et al., 2002). Further, the underestimation of the VIF leads to an inflation of the Type-I error.

Regarding a negative ICC, the effect on the Type-I error is reversed. Researchers agree that when fixing a negative ICC to zero, the Type-I error is deflated (Baldwin et al., 2008, Rosner & Grove, 1999). Also in this case, the design effect is used to explain the deflation. When the true negative ICC is fixed to zero, the design effect is overestimated, so that SEs are also overestimated, and the Type-I error is deflated (Norton et al., 1996).

When incorrectly assuming that observations are independently distributed (using a lower bound of zero for the ICC), while they are negatively correlated, then a larger SE of a fixed effect and a deflation of the Type-I error are obtained.

Overall, the common procedure to ignore small positive and negative ICC leads to invalid SEs, Type-I errors, confidence intervals, and effect sizes (Kenny et al., 2002). The purpose of this study is to confirm biases that occur when ignoring small positive ICC. Additionally, this study reports about the magnitude of bias of p-values and standard errors, the deflation of Type-I errors, and bias in the intra-group correlation estimates, when ignoring a true negative ICC.

METHOD

Statistical models

In this simulation study the linear mixed effects (LME) model was used to describe clustered data. Data was generated under a LME model and under a multivariate model. As the LME model assumes that the observations are conditionally independently distributed, we call it conditional LME (conLME). The multivariate model is referred to as a marginal LME (marLME) because it models the dependence through a covariance matrix, thereby allowing small positive as well as (small) negative correlation within groups. The conLME model is represented as:

$$Y_{ij} = \beta_0 + X_{ij}\beta_1 + \beta_j + e_{ij} \quad (1a)$$

$$\text{where } \beta_j \sim N(0, \tau), \quad (1b)$$

$$e_{ij} \sim N(0, \sigma^2) \quad (1c)$$

In this model, Y_{ij} is the dependent variable. The X_{ij} is the independent variable and its effect β_1 is the regression coefficient that represents the effect of the independent variable. The intercept is represented by β_0 and β_j , whereby the random part of the intercept is β_j , which is normally distributed with a mean of zero and a variance τ . The β_j refers to the variability in the model due to a clustering effect. The variance of β_j is equal to τ , and models the correlation within groups, so it is related to the ICC because $ICC = \frac{\tau}{\tau + \sigma^2}$. As τ is a variance its value range is restricted to at least zero. The e_{ij} is the residual error at the individual level.

In contrast, the marLME allows the correlation within groups to be negative (Giberson, Resick, & Dickson, 2005). The marLME model is given by:

$$Z_j = X_j\beta_1 + E_j \quad (2a)$$

$$\text{where } E_j \sim N(0, \Sigma_j) \quad (2b)$$

$$\Sigma_j = \begin{bmatrix} \sigma^2 + \tau & \tau & \tau & \dots \\ \tau & \sigma^2 + \tau & \tau & \dots \\ \tau & \tau & \sigma^2 + \tau & \dots \\ \dots & \dots & \dots & \dots \end{bmatrix} \quad (2c)$$

In the formula 2a, the independent and dependent variables are the same as in conLME (formula 1a), but here the outcome of each group is assumed to be multivariate normally distributed. Also, the regression coefficient and intercept refer to the same variables. Instead of modelling the correlation in the clusters through a random intercept, the correlation is described in the covariance matrix of the multivariate distributed errors E_{ij} , that are normally distributed with a mean of 0 and a covariance matrix of Σ_j . Like in conLME, also in marLME, τ represents the correlation within groups and σ^2 is the residual error at the individual level. As τ is included in the variance-covariance matrix and as it is a covariance parameter, a negative ICC is allowed in marLME.

Each set of generated data was analysed with a conLME model allowing only positive clustering effects, and with a linear model (LM) that does not account for clustering effects. Therefore, in RStudio the function `lme4` was used for the conLME model of analysis which is in accordance with formula 1a. For LM as the model of analysis the function `lm` was used. The LM was represented by:

$$Y_i = \beta_0 + X_i\beta_1 + e_i \quad (3)$$

Hereby, the parameters represent the same effects as in the conLME model. In contrast to conLME (formula 1a), the LM does not include a clustering effect and therefore, does not include the random effect parameter β_j .

Summarising: (1) data generated under conLME was analysed with the conLME model, (2) data generated under marLME was analysed with the conLME model, (3) data generated under conLME was analysed with LM and (4) data generated under marLME was analysed with LM. The models 1 and 3 allowed positive but no negative true values of τ because the data was generated under conLME, where τ was the variance of the parameter β_j . In contrast, data generated under marLME did

allow positive and negative true values of τ because it was a covariance parameter; this applies to models 2 and 4. Regarding the estimates, models 3 and 4 did not account for correlation at all, so values of τ could not be estimated. The models 1 and 2 allowed only positive estimates of τ .

Simulation study design

This simulation study had in total $3 \times 2 \times$ (levels of τ) conditions. According to Dorman (2008), a large number of clusters can compensate for biases that occur due to the ignorance of small and negative ICC, therefore the number of clusters was fixed to $a = 10$. Barcikowski (1981) showed that the number of observations per cluster influences the magnitude of a Type-I error. Therefore, in this study, the number of observations per cluster took three values ($m = \{10; 15; 30\}$). Moreover, the effect of correlated data on the fixed effect was investigated by setting the regression coefficient in one condition to 0 ($\beta_0 = 0$) and in a second condition to $\beta_1 = 0.1$.

Parameter τ , the variance in the model due to clustering, was set as a function of m and σ^2 as it was not possible to have more negative correlation within groups than $\tau_{min} = \frac{-(\sigma^2)}{m}$ (Fox, Mulder, & Sinharay, 2017). Thereby, σ^2 was fixed to 1. Thus, τ took values from $\frac{-1}{m}$ to 0 in steps of 0.02 and from 0 to 2 in steps of 0.05. The steps for $\tau < 0$ were set to 0.005 for $m = 30$, so that at least two negative values of τ were used for the data generation. Negative values of τ were only allowed by marLME so that negative true τ values were set to 0 for generating data under conLME. Therefore, the amount of conditions changed depending on m . For each condition, 1000 simulations were conducted with a seed sampled out of 1 to 100000.

Simulation study procedure

To evaluate the accuracy of the conLME model as an analysing tool for clustered data the simulation was conducted under each condition 1000 times. First, it was checked if the ignorance of the correlation within groups had an influence on the estimation of the fixed effect. Therefore, data was generated under conLME for two conditions each: only-intercept condition ($\beta_0 = 0$) and the intercept-slope condition ($\beta_1 = 0.1$). The model for the only-intercept condition was formulated as:

$$Y_{ij} = \beta_0 + \beta_j + e_{ij} \quad (4)$$

Moreover, model for the intercept-slope condition was formulated as:

$$Y_{ij} = \beta_0 + X_{ij} * \beta_1 + \beta_j + e_{ij} \quad (5)$$

The parameters of the models refer to the same parameters that are mentioned in formula 1a.

The Type-I error was estimated for each model under each condition. Therefore, for the only-intercept condition, when the population mean was 0, it was examined if the ignorance of small positive and (small) negative correlation within groups affected the intercept estimate. Moreover, for the intercept-slope condition, where the population slope was equal to 0.1, it was checked if ignoring correlation within groups had an effect on the slope estimate.

To compute the Type-I error, a coverage rate (CR) was computed by computing the significance of either the intercept or the correlation coefficient ($\beta_0 = 0$ or $\beta_1 = 0.1$) for each condition and model variant. Under the only-intercept condition, we generated data with an intercept equal to 0. Under the intercept-slope condition, we generated data with a correlation coefficient of 0.1. When the true value was within the range of the 95% CI of a simulation, the value 1 was assigned, if not, the assigned value was 0. Afterwards, the amount of 1s was divided by the total amount of simulations ($n = 1000$) for each condition and model variant. So, the p-values represented the significance level. With this, the Type-I error could be estimated by subtracting the CR from 1. So, it was checked if the ignorance of the correlation within groups had an influence on the intercept or on the slope of the fixed effect.

Second, the bias of τ was estimated for all conditions and each model variant by comparing the true τ used in the data generation with the estimated τ . And third, the p-value and SE of the fixed effect were evaluated for different values of τ . Moreover, the true ICC was computed with the true values of τ and σ^2 to see how large the bias was regarding the p-value and SE of the fixed effect. The formula was given by:

$$ICC = \frac{\tau}{\tau + \sigma^2} \quad (6)$$

RESULTS

Coverage rate and Type-I error

The CR was computed for the only-intercept ($\beta_0 = 0$) and the intercept-slope condition ($\beta_1 = 0.1$). Only model variants 2 and 4 allowed true values of τ to be negative, as data generated under marLME could have true negative τ values. For model variants 1 and 3, data was generated under conLME, so that true τ values were confined to at least 0. For that reason, the CR for model variants 1 and 3 was only computed for $\tau \geq 0$. For all model variants the CR of the slope ($\beta_1 = 0.1$) was accurately estimated at about .95 for data generated under the intercept-slope condition (see Table A1).

In contrast, CRs of the intercept under the only-intercept condition ($\beta_0 = 0$) were biased (see Table A2). When the true value of τ was smaller than 0 the CR was larger than .95 for data generated under marLME, which means that the Type-I error was reduced. For all cluster sizes used in this study the CR was at least .97 for very small negative ICC ($\tau \leq -.01$). This effect was the same for model variants 2 and 4 as they were alike for $\tau < 0$. This is because the conLME model could not estimate negative values of τ , so that when the true τ was negative, τ and β_0 were estimated close to 0, leading LM and the conLME model to be similar models of analysis for true $\tau < 0$.

For true $\tau > 0$ the conLME model as model of analysis gave approximately the same CR values for different true values of τ . The CRs were ranging from .90 (model variant 1; $m = 30$; true $\tau = .20$) to .95 (model variant 1; $m = 15$; true $\tau = .05$). Regarding LM as the model of analysis, the CRs were highly underestimated for true $\tau > 0$; thereby, the underestimation was larger for $m = 30$ than for $m = 10$. This is because LM ignored correlation within groups so that the model assumed more information within the data than there actually was. Therefore, the Type-I error was inflated.

In total, there were biases in the CRs of the intercept when the population mean was 0, but the CRs of the slope were accurately estimated at about .95 when the population slope was 0.1. From this, it was concluded that the ignorance of the correlation within groups did not have an influence on the Type-I error of the slope. As the slope was unaffected by the ignorance, results were only reported for data generated under the only-intercept condition.

Estimates of τ

The τ values could only be estimated by the conLME model of analysis because, in contrast to the LM model, the conLME model included a parameter for correlation on group level. Moreover, true τ values generated under conLME were restricted to a lower bound of 0, so that τ estimates for true negative τ values could only be simulated for data generated under marLME. However, the model of analysis (LME) was restricted to a non-negative parameter space for τ estimates. This was evident in the τ estimates for true negative τ values as they were biased upwards due to the lower bound of 0 (for all τ estimates see Table A3).

For the same reason, τ estimates were also biased upwards for data generated under conLME and marLME when the true value of τ was equal to 0. As can be seen in Table 1, the bias was smaller for larger values of m because with an increasing sample size there is less sampling error. This led to less variability in the data, which means that more information about the fixed effect was included in the data.

Regarding true positive τ values, estimates of τ were not biased for model variants 1 and 2 when true values of τ were greater than or equal to .05. This effect was equal for all values of m used in this study, which was because the conLME model gave correct estimates of τ for clustered data with positive ICC.

Table 1

τ estimates under different conditions for m for data analysed with the conLME model for true $\tau = 0$

Numbers of observations per cluster (m)	Model variant	
	Data generated under conLME is analysed with the conLME model (1)	Data generated under marLME is analysed with the conLME model (2)
10	.019	.018
15	.012	.012
30	.006	.007

p-values

Generally, when testing the effect of the independent variable $\beta_1 = 0$ on the dependent variable and the true value is 0, then p-values are expected to be uniformly distributed, which means that they are centred at .50. This happens when 5% of the p-values are smaller than or equal to .05. If the p-value is biased upwards, less p-values take a value of at most .05, which means that Type-I error is deflated. For p-values that are biased downwards, the effect is vice-versa.

As can be seen in Figure 1 and Table A4, when the conLME model was used for the analysis of data generated under conLME and marLME, the p-value was about .50 for true values of τ greater than 0 for all values of m . This confirmed that the conLME model gave correct estimates of p-values when observations within clusters were positively correlated.

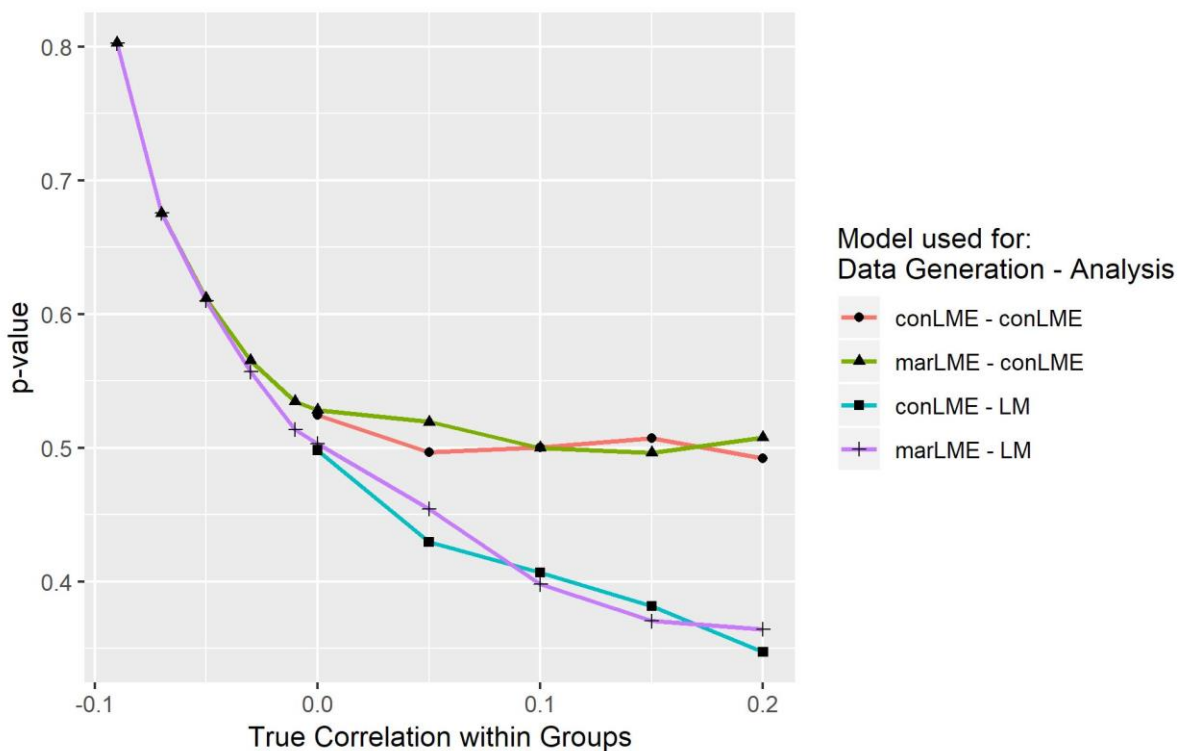


Figure 1. p-values of the intercept for different values of true τ shown for each model variant. The figure shows the standard error under the condition $m = 10$.

For $\tau > 0$ the LM model as model of analysis gave underestimated p-values as the model did not account for the ICC. Therefore, the model assumed that there was more information within the

data than there actually was, which led to a higher percentage of significant p-values than 5%.

Therefore, the p-values were underestimated. With increasing τ that was ignored, the concomitant greater variability in the data was assumed to be because of the dependent variable, leading to a greater underestimation of p-values.

For true τ values smaller than .05 the p-value was always biased upwards for data analysed with the conLME model because estimates of τ were upwardly biased when the true τ was negative or close to 0. This was due to the lower bound of τ estimates for the conLME model as model of analysis. Moreover, the more negative the true value of τ was, the more similar were the p-values for the analysis models conLME and LM. For LM, correlation within groups was always ignored, so that the p-value was only dependent on the σ^2 estimate. However, for the conLME model, the p-value depended on σ^2 estimates and τ estimates when the true τ takes smaller values than .05. For larger true negative τ values the τ estimate was smallest so that the p-value of LM and the conLME model both depended only on the σ^2 estimate. Therefore, the p-value was similarly biased for both analysing models when $\tau < .05$.

Overall, the bias of the p-value was smaller for larger values of m , which was due to the increased sample size and, therefore, smaller sampling error and more information about the effect of the independent variable in the data. However, in case of a large number of observations per cluster, already a small negative correlation within clusters could lead to large bias in p-value. For example, for $m = 30$ the correct p-value of 0.50 was increased by .06 for true $\tau = -.008$; and for true $\tau = -.023$ the p-value was overestimated by .22 ($p = .77$). For smaller values of m the bias for small true negative values of τ were smaller. However, for smaller m the lower boundary of τ was also smaller which means that more negative correlation within groups could be in the data which could lead to large bias of the p-value.

Standard errors

In general, as can be seen in Figure 2, SE values were larger when the conLME model was used for the analysis than when LM was used. This is because the conLME model assumed more variability in

the data and accounted for a clustering effect. The LM assumed that there was more information about the fixed effect in the data, so that SE values were smaller than in the conLME model of analysis. All true values of τ smaller than 0 were only generated under marLME, so that the SE for true negative τ values could only be computed for model variants 2 and 4.

For all model variants, SE takes the smallest value when τ is smallest because then there was less variability in the data, so that the data contained most information about the effect of the dependent variable. When τ increased the variability in the data increased as well so that the information about the effect of the fixed effect was reduced leading to larger SE values. For LM as the model of analysis, SE values were rather small in comparison to SE values from the analysis with the conLME model. This reflects that LM as a model of analysis did not account for variability due to correlation within groups. Therefore, the slight increase of SE values for model variants 3 and 4 when τ got larger was due to an increase of σ^2 . For the same reason, the SE was linearly related to the true τ value for model variant 4 when the true value of τ was negative.

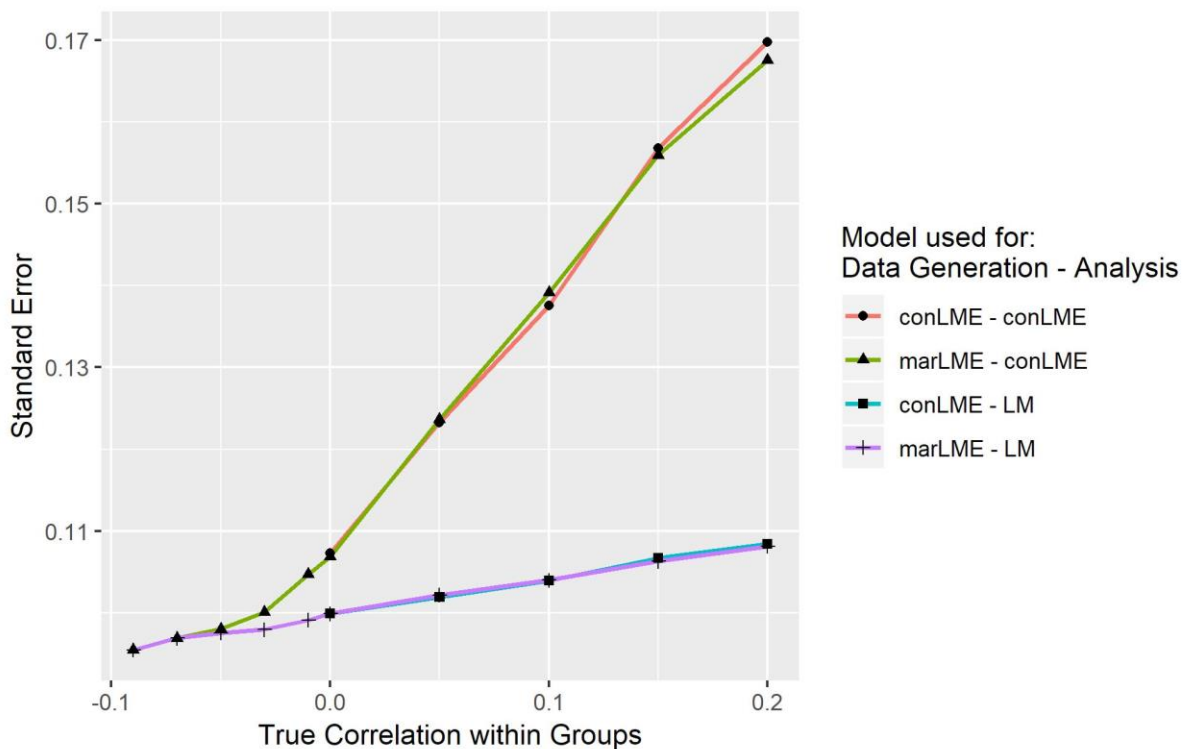


Figure 2. Standard error of the intercept for different values of true τ shown for each model variant.

The figure shows the standard error under the condition $m = 10$.

When the conLME model was used for the analysis, SE values were also positively related to true τ values. For true $\tau \geq 0$ the slope was steeper than for true $\tau < 0$. For true $\tau \geq 0$ the conLME model as a model of analysis could estimate the correlation within groups accurately. Thus, the conLME model assumed correctly that there was less information about the effect of the independent variable in the data, leading to an increase of SE values with growing τ values. In contrast, the SE values for true negative τ were only dependent on σ^2 estimates as negative ICC could not be estimated by the conLME model in the analysis and, therefore, was estimated close to 0. As clustered data with negative correlation within groups is more informative than data gathered under independent sampling, the SEs were overestimated when negative correlation within groups was ignored.

Regarding the number of observations per cluster, the SE got smaller the larger m was which was due to the smaller sampling error with increasing sample size. Thus, more information about the effect of the independent variable was included in the data so that the SE got smaller. For all values of SE see Table A5.

Intra-class correlation

As in this study σ^2 was set to the value 1, the value of the true ICC was very close to the value of τ . For example, when the true τ was .05, the true ICC was 0.048, almost equal to $\frac{0.05}{0.05+1}$. Therefore, the results were mostly reported for τ and only sparsely for the ICC. However, in more general situations, when σ^2 is not equal to 1, the ICC will be more interesting to look at.

Regarding the p-value, the ignorance of negative ICC led to an overestimation of the p-value in an analysis with the conLME model for data generated under marLME. The effects on the bias for different conditions of m can be found under ‘p-values’. Moreover, for a true ICC greater than 0 the p-values were correctly estimated at about .50 by the conLME as model of analysis for data generated under conLME and marLME. See ‘p-values’ for more details.

Similarly, the relation between true ICC and SEs was corresponding to the relation between τ and SEs. The smaller the true ICC was the smaller was the SE for data analysed with conLME. When data was generated under conLME, this was true for true $ICC \geq 0$, whereas for data generated

marLME it was true for all possible true values of ICC. For a more detailed description, see ‘Standard errors’.

DISCUSSION

The purpose of this study was to confirm that ignoring (small) positive ICC leads to biased estimates of the LME model. Further, this study reported about the magnitude of bias of p-values and SEs, deflation of Type-I errors, and bias in the intra-group correlation estimates when ignoring (small) true negative ICC. The results show that, in general, only the intercept and not the slope of the fixed effect was affected by ignoring positive or negative correlation within groups. As expected, the estimates of the intra-group correlation were biased upwards for true small positive and negative values. Further, the results of this study confirm that the ignorance of (small) positive correlation within groups leads to an increase of Type-I error, p-values that are biased downwards and CIs that are too narrow. Regarding the ignorance of (small) negative correlation within groups, the results showed the opposite of ignoring positive correlation within groups: a deflation of Type-I errors, p-values that are biased upwards and overestimated SEs.

In line with earlier research, the slope parameter was unaffected by the ignorance of any correlation within groups in this study. According to Kenny et al. (2002), the correlation within groups does not have an influence on estimates of the independent variables, whereby they refer to the variation that is explained by the slope variable of the LME model. However, other researchers pointed out that estimates of a fixed effect are influenced when ignoring intra-group correlation (Baldwin et al., 2005; Barcikowski, 1981). Thereby, they refer to estimates of a fixed effect in general, so they include the intercept and slope parameter and do not restrict the influence on the estimate of the slope of the fixed effect. This is in accordance with our finding that bias occurred only when testing if group means were significantly different from the population mean, referring to the intercept. However, there was no bias found when testing if the slopes of the groups were significantly different from the population slope, referring to the random effect of an independent variable.

Regarding the estimates of the correlation within groups, the results of this study confirm the restriction of the LME model that correlation within groups is confined to at least zero (Eldridge et al.,

2009). Already for small positive correlation within groups ($\tau < .05$) the estimate of the correlation was overestimated. When the true correlation was smaller than that or negative, the overestimation was even larger. However, for a positive correlation within groups that was larger than .05, the conLME model gave correct estimates, which validates the accuracy of the LME model for clustered data with positively correlated observations.

The results of this study for ignoring true positive correlation within clusters validate earlier research. The results show that ignoring positive ICC leads to significance levels that are too low. In line with this, in this study CIs were too narrow, and the Type-I error was inflated. Similar results were found by Barcikowski (1981), Clarke (2008), and Rosner and Grove (1999). Thereby, the value of the Type-I error of this study was the same as the value found by Barcikowski (1981) for ten observations per cluster and an ignored ICC of .05, which is adding to the validity of this study. Additionally, we found out that the SE is overestimated when ignoring positive correlation within groups, which is in line with the other results.

Regarding different conditions when ignoring positive correlation within groups, the results of this study are also confirming earlier research. Like Barcikowski (1981) and Dorman (2008), we found the least bias in Type-I error for a smaller number of observations per group. When the number of observations per group increased, the inflation of the Type-I error also got larger. Moreover, when increasing the correlation that was ignored, the Type-I error was increasing as well. The same was reported by Barcikowski in 1981.

Also, for the ignorance of negative correlation within groups the results are in accordance with earlier research. However, this study reports in more detail about the effects. The deflation of Type-I error that occurred when ignoring negative correlation within groups was also reported by other researchers (Baldwin et al., 2008; Rosner & Grove, 1999). Also, the overestimation of SEs was reported elsewhere (Norton et al., 1996). This finding indicates that when having data with negative correlation within groups, the data is more informative than it would be under independent sampling. This may result in underestimating the difference between group means and the population mean.

This study adds that smaller biases of the Type-I error and the SEs occur when the number of observations within a group gets larger. This can be explained by the increase in sample size, leading

to smaller sampling error and more precise estimates. In addition, this study specifies the bias of the Type-I error: A deflation of the Type-I error by 2% can already occur for the small true negative ICC of -.01. Thereby, the deflation was the same for all numbers of observations per cluster tested.

Moreover, the results indicate that p-values were biased upwards when negative correlation within groups was ignored. Again, due to the smaller sampling error, a greater number of observations per cluster lead to smaller bias for the same ignored true negative correlation. Nevertheless, for a larger number of observations per cluster, even the ignorance of very small negative correlation within groups can lead to large bias of the p-value.

With respect to the results of the current study and findings of other researchers, it is recommended to be aware of the fact that negative correlation within groups, which leads to a negative ICC, may occur. We advise to be very cautious with clustered data that possibly contains negative correlation between observations within groups. Researchers should know that the LME model estimates the correlation between groups close to zero, when the true correlation is negative. This can lead to a deflation of Type-I error, p-values that are biased upwards and SEs that are overestimated. All of this results in an underestimation of the difference between group means and the population mean, and therefore, to false results.

In addition, it is suggested to consider the number of observations per cluster, as a large number of observations per cluster does not compensate for bias. It is important to notice that at 30 observations per cluster already a very small true negative correlation of -.008 within clusters can lead to an inflation of the p-value ($p = .55$). Likewise, researchers should be cautious that, when having a smaller number of observations per cluster, the correlation can be more negative than at a larger number of observations. In this case, the bias of the p-value is smaller when the true negative correlation between observations within groups is small, but as the correlation can take larger true negative values, the bias can be large as well.

The current study focused on bias that occur when ignoring (small) positive or (small) negative correlation within groups. The results of this study are limited by fixing the variance of the residual

error at the individual level to 1. Other researchers found that the ICC and VIF are affected by the ignorance of correlation within groups (Baldwin et al., 2005; Norton et al., 1996). In this study, we fixed the variance of the residual error to 1, so the true ICC was very similar to the true correlation within groups. Thus, all parameter estimates of the conLME model were similar for the true ICC and its corresponding true correlation. Therefore, it is suggested for future research to look at how the ICC and VIF are influenced by the ignorance of correlation within groups in more general situations when the variance of the random error is not equal to 1.

Furthermore, this study looked at different values of correlation within groups and at different cluster sizes. Dorman (2008) also mentioned that the number of clusters has an influence on the bias as well, so for future research it is recommended to inspect different conditions of this parameter. By this it will be possible to specify the bias found in this study.

Moreover, the findings of this study are limited in their application because, until now, there is no proper tool known for detecting negative correlation within groups. So, in case researchers are not aware that there might be negative correlation within groups when having clustered data, the LME model automatically assumes true negative correlation to be close to zero. This leads to bias that may not be recognised by the researchers. Therefore, this paper can only advise researchers to be aware that negatively correlated observations within groups may exist even at a large number of observations per cluster, which can lead to bias in the analysis. In future research it is essential to find a tool or strategy to detect negative correlation within groups. In addition, future research should aim at developing a model that can account for negative correlation within groups.

CONCLUSION

Negative correlation between observations within groups can occur when having clustered data. If the LME model is used for analysing data with negative correlation within groups, the model assumes the observations to be independently distributed. The results of this study indicate that this ignorance of negative correlation within groups leads to deflated Type-I errors, overestimated SEs and overestimated p-values. Therefore, the LME model is an inappropriate tool for analysing clustered data with negative correlation within groups. Researchers should be aware that, when observations

within clusters are negatively correlated, the data contains more information than when having independently distributed observations. This means that the difference between group means and the population mean might be underestimated.

REFERENCES

- Baldwin, S. A., Murray, D. M., & Shadish, W. R. (2005). Empirically supported treatments or type I errors? Problems with the analysis of data from group-administered treatments. *Journal of Consulting and Clinical Psychology, 73*(5), 924-935. doi:10.1037/0022-006X.73.5.924
- Baldwin, S. A., Stice, E., & Rohde, P. (2008). *Statistical analysis of group-administered. Psychotherapy Research, 18*(4), 365-376. doi:10.1080/10503300701796992
- Barcikowski, R. S. (1981). Statistical power with group mean as the unit of analysis. *Journal of Educational Statistics, 6*(3), 267-285. doi:10.2307/1164877
- Clarke, P. (2008). Theory and methods: When can group level clustering be ignored? Multilevel models versus single-level models with sparse data. *Journal of Epidemiology and Community Health, 62*(8), 752-758. doi:10.1136/jech.2007.060798
- Dorman, J. P. (2008). The effect of clustering on statistical tests: An illustration using classroom environmental data. *Educational Psychology, 28*(5), 583-595. doi:10.1080/01443410801954201
- Eldridge, S. M., Ukoumunne, O. C., & Carlin, J. B. (2009). The intra-cluster correlation coefficient in cluster randomized trials: A review of definitions. *International Statistical Review, 77*(3), 378-394. doi:10.1111/ij.1751-5823.2009.00092.x
- Fox, J.-P., Mulder, J., & Sinharay, S. (2017). Bayes factor covariance testing in item response models. *Psychometrika, 82*(4), 979-1006. doi:10.1007/s11336-017-9577-6
- Giberson, T. R., Resick, C. J., & Dickson, M. W. (2005). Embedding leader characteristics: An examination of homogeneity of personality and values in organisations. *Journal of Applied Psychology, 90*(5), 1002-1010. doi:10.1037/0021-9010.90.5.1002
- Hox, J. J. (2010). *Multilevel analysis* (2nd ed.). New York, NY: Routledge.
- Kenny, D. A., & Judd, C. M. (1986). Consequences of violating the independence assumption in analysis of variance. *Psychological Bulletin, 99*(3), 422-431. doi:10.1037//0033-2909.99.3.422

- Kenny, D. A., Mannetti, L., Pierro, A., Livi, S., & Kashy, D. A. (2002). The statistical analysis of data from small groups. *Journal of Personality and Social Psychology*, 83(1), 126-137. doi:10.1037//0022-3514.83.1.126
- Krannitz, M. A., Grandey, A. A., Liu, S., & Almeida, D. A. (2015). Workplace surface acting and marital partner discontent: Anxiety and exhaustion spillover mechanisms. *Journal of Occupational Health Psychology*, 20(3), 314-325. doi:10.1037/a0038763
- Langfred, C. W. (2007). The downside of self-management: A longitudinal study of the effects of conflict on trust, autonomy, and task interdependence in self-managing teams. *Academy of Management Journal*, 50(4), 885-900. doi:10.5465/amj.2007.26279196
- Maas, C. J., & Hox, J. J. (2005). Sufficient sample sizes for multilevel modeling. *Methodology*, 1(3), 86-92. doi:10.1027/1614-1881.1.3.86
- Norton, E. C., Bieler, G. S., Ennett, S. T., & Zarkin, G. A. (1996). Methodological developments - Analysis of prevention program effectiveness with clustered data. *Journal of Consulting and Clinical Psychology*, 64(5), 919-926. doi:10.1037/0022-006X.64.5.919
- Rosner, B., & Grove, D. (1999). Use of the Mann–Whitney U-test for clustered data. *Statistics in Medicine*, 18(11), 1387-1400. Retrieved on March 14, 2019, from <https://onlinelibrary.wiley.com/>

APPENDIX A

Table 1

Coverage rate for the slope $\beta_1 = 0.1$ based on its 95% confidence interval

True value of τ	Model variant			
	Data generated under conLME is analysed with the conLME model (1)	Data generated under the marLME is analysed with the conLME model (2)	Data generated under conLME is analysed with LM (3)	Data generated under marLME is analysed with LM (4)
Number of observations per cluster (m) = 10				
-0.09		.93		.94
-0.07		.94		.94
-0.05		.95		.95
-0.03		.95		.95
-0.01		.95		.96
.00	.94	.95	.94	.95
.05	.94	.94	.94	.95
.10	.95	.95	.96	.95
.15	.95	.94	.95	.95
.20	.94	.94	.95	.95
Number of observations per cluster (m) = 15				
-0.06		.96		.96
-0.04		.95		.96
-0.02		.94		.94
.00	.95	.94	.95	.94
.05	.95	.95	.95	.95
.10	.96	.95	.97	.95
.15	.94	.95	.94	.96
.20	.94	.96	.94	.95
Number of observations per cluster (m) = 30				
-0.03		.95		.96
-0.01		.96		.97
.00	.95	.94	.94	.94
.05	.96	.95	.95	.94
.10	.94	.93	.93	.93
.15	.95	.94	.94	.92
.20	.96	.95	.94	.94

Note. τ is restricted to a lower bound of $-\frac{(\sigma^2)}{m}$. For data generated under conLME true values of τ are

restricted to values of at least 0.

Table 2

Coverage rate for the intercept $\beta_0 = 0$ based on its 95% confidence interval

True value of τ	Model variant			
	Data generated under conLME is analysed with the conLME model (1)	Data generated under marLME is analysed with the conLME model (2)	Data generated under conLME is analysed with LM (3)	Data generated under marLME is analysed with LM (4)
Number of observations per cluster (m) = 10				
-0.09		1.00		1.00
-0.07		1.00		1.00
-0.05		1.00		1.00
-0.03		.98		.98
-0.01		.98		.97
.00	.96	.97	.95	.95
.05	.93	.94	.89	.90
.10	.94	.94	.86	.86
.15	.94	.93	.81	.80
.20	.94	.94	.76	.80
Number of observations per cluster (m) = 15				
-0.06		1.00		1.00
-0.04		1.00		1.00
-0.02		.98		.98
.00	.95	.96	.94	.95
.05	.95	.94	.88	.88
.10	.93	.93	.81	.80
.15	.93	.92	.75	.76
.20	.91	.93	.70	.72
Number of observations per cluster (m) = 30				
-0.03		1.00		1.00
-0.01		.97		.97
.00	.97	.96	.96	.95
.05	.94	.94	.82	.79
.10	.93	.93	.69	.68
.15	.92	.94	.62	.62
.20	.90	.93	.53	.59

Note. τ is restricted to a lower bound of $-\frac{(\sigma^2)}{m}$. For data generated under conLME true values of τ are

restricted to values of at least 0.

Table 3

 τ estimates under all conditions for m for data analysed with the conLME model

True value of τ	Model variant	
	Data generated under conLME is analysed with the conLME model	Data generated under marLME is analysed with the conLME model
	(1)	(2)
number of observations per cluster (m) = 10		
-.09		.000
-.07		.000
-.05		.001
-.03		.005
-.01		.014
.00	.019	.018
.05	.059	.059
.10	.099	.103
.15	.159	.157
.20	.204	.197
number of observations per cluster (m) = 15		
-.06		.000
-.04		.000
-.02		.004
.00	.012	.012
.05	.055	.051
.10	.104	.100
.15	.151	.150
.20	.194	.201
number of observations per cluster (m) = 30		
-.03		.000
-.01		.002
.00	.006	.007
.05	.049	.050
.10	.097	.099
.15	.150	.150
.20	.196	.195

Note. The τ estimates are shown with three decimals to show the slight decrease of τ estimates for model variant 2 when the true τ is smaller than 0. τ is restricted to a lower bound of $\frac{-(\sigma^2)}{m}$. For data generated under conLME true values of τ are restricted to values of at least 0.

Table 4

p-values of the intercept under the condition $\beta_0 = 0$ for all model variants and values of m

True value of τ	Model variant			
	Data generated under conLME is analysed with the conLME model (1)	Data generated under marLME is analysed with the conLME model (2)	Data generated under conLME is analysed with LM (3)	Data generated under marLME is analysed with LM (4)
number of observations per cluster (m) = 10				
-.09		.80		.80
-.07		.68		.68
-.05		.61		.61
-.03		.57		.56
-.01		.53		.51
.00	.52	.53	.50	.50
.05	.50	.52	.43	.45
.10	.50	.50	.41	.40
.15	.51	.50	.38	.37
.20	.49	.51	.35	.36
number of observations per cluster (m) = 15				
-.06		.76		.76
-.04		.62		.62
-.02		.57		.56
.00	.52	.52	.49	.50
.05	.52	.51	.43	.42
.10	.50	.49	.37	.36
.15	.49	.50	.33	.34
.20	.49	.50	.32	.31
number of observations per cluster (m) = 30				
-.03		.77		.77
-.01		.56		.55
.00	.53	.51	.50	.49
.05	.51	.50	.38	.37
.10	.50	.48	.32	.29
.15	.49	.49	.26	.25
.20	.47	.50	.23	.25

Note. τ is restricted to a lower bound of $-\frac{(\sigma^2)}{m}$. For data generated under conLME true values of τ are

restricted to values of at least 0.

Table 5

SE values of the intercept under the condition $\beta_0 = 0$ for all model variants and values of m

True value of τ	Model variant			
	Data generated under conLME is analysed with the conLME model (1)	Data generated under marLME is analysed with the conLME model (2)	Data generated under conLME is analysed with LM (3)	Data generated under marLME is analysed with LM (4)
number of observations per cluster (m) = 10				
-0.09		0.095		0.095
-0.07		0.097		0.097
-0.05		0.098		0.098
-0.03		0.100		0.098
-0.01		0.105		0.099
.00	0.107	0.107	0.100	0.100
.05	0.123	0.124	0.102	0.102
.10	0.138	0.139	0.104	0.104
.15	0.157	0.156	0.107	0.106
.20	0.170	0.168	0.108	0.108
number of observations per cluster (m) = 15				
-0.06		0.079		0.079
-0.04		0.080		0.080
-0.02		0.083		0.081
.00	0.087	0.087	0.081	0.081
.05	0.108	0.106	0.084	0.083
.10	0.127	0.126	0.085	0.085
.15	0.144	0.143	0.087	0.087
.20	0.157	0.160	0.088	0.089
number of observations per cluster (m) = 30				
-0.03		0.057		0.057
-0.01		0.059		0.057
.00	0.062	0.062	0.058	0.058
.05	0.088	0.089	0.059	0.059
.10	0.111	0.112	0.060	0.060
.15	0.132	0.131	0.061	0.061
.20	0.147	0.147	0.062	0.063

Note. The SE values are shown with three decimals to show the slight increase of SE values with

increasing τ for all values of m and for model variants 3 and 4. τ is restricted to a lower bound of

$-\frac{(\sigma^2)}{m}$. For data generated under conLME true values of τ are restricted to values of at least 0.