Does it decay? Obtaining decaying correlation parameter values from previously analysed cluster randomised trials

Jessica Kasza[1] (iD), Rhys Bowden[1] (iD), Yongdong Ouyang[2,3] (iD), Monica Taljaard[2,3] and Andrew B Forbes[1]

Abstract

A frequently applied assumption in the analysis of data from cluster randomised trials is that the outcomes from all participants within a cluster are equally correlated. That is, the intracluster correlation, which describes the degree of dependence between outcomes from participants in the same cluster, is the same for each pair of participants in a cluster. However, recent work has discussed the importance of allowing for this correlation to decay as the time between the measurement of participants in a cluster increases. Incorrect omission of such a decay can lead to under-powered studies, and confidence intervals for estimated treatment effects can be too narrow or too wide, depending on the characteristics of the design. When planning studies, researchers often rely on previously reported analyses of trials to inform their choice of intracluster correlation. However, most reported analyses of clustered data do not incorporate a correlation decay. Thus, often all that is available are estimates of intracluster correlations obtained under the potentially incorrect assumption of no decay. In this article, we show that it is possible to use intracluster correlation values obtained from models that incorrectly omit a decay to inform plausible choices of decaying correlations. Our focus is on intracluster correlation estimates for continuous outcomes obtained by fitting linear mixed models with exchangeable or block-exchangeable correlation structures. We describe how plausible values for decaying correlations may be obtained given these estimated intracluster correlations. An online app is presented that allows users to obtain plausible values of the decay, which can be used at the trial planning stage to assess the sensitivity of sample size and power calculations to decaying correlation structures.

Keywords

Cluster autocorrelation, hierarchical models, intracluster correlation, sample size calculation, stepped wedge, within-cluster correlation structure

1 Introduction

Longitudinal cluster randomised trials are cluster randomised trials where clusters are followed up over multiple trial periods. This class of designs includes parallel arm cluster randomised trials, stepped wedge trials, cluster randomised crossover trials, and associated variants such as incomplete stepped wedge designs.[1] Participants within clusters may provide data in one or more trial periods, although our focus is on the setting where each participant provides a single measurement, and

[1]School of Public Health and Preventive Medicine, Monash University, Melbourne, VIC, Australia
[2]Clinical Epidemiology Program, Ottawa Hospital Research Institute, Ottawa, Canada
[3]School of Epidemiology and Public Health, University of Ottawa, Ottawa, Canada

Corresponding author:
Jessica Kasza, School of Public Health and Preventive Medicine, Monash University, 553 St Kilda Road, Melbourne, VIC 3004, Australia.
Email: jessica.kasza@monash.edu

we limit attention to a continuous outcome. When designing or analysing cluster randomised trials it is vitally important to account for the similarity between pairs of outcomes of participants in the same cluster. The simplest assumption is that the outcomes from participants in the same cluster are all equally correlated. In the stepped wedge literature, the model corresponding to this assumption is often referred to as the exchangeable model, or the Hussey and Hughes model.[2] However, more recently researchers have recognised the need to be able to allow for correlations between outcomes to vary based upon how far apart in time that they are measured. In many situations it is most reasonable to assume that outcomes recorded in the same time period will be more similar than those in differing time periods. There are two popular models for expressing this: the block-exchangeable model, where the correlation between a pair of participants in the same cluster only depends on whether they were measured in the same or different study periods[3,4]; and the discrete time decay model, which allows for the correlation between a pair of participants to decay exponentially as a function of the number of study periods between their recruitment.[5]

When planning longitudinal cluster randomised trials, estimates of intracluster correlations and associated parameters are required for sample size/power calculations.[6] When the exchangeable model is assumed, researchers need only specify a single intracluster correlation, which describes the similarity in outcomes for participants in the same cluster regardless of time period. When the block exchangeable model is assumed, researchers must specify two parameters. The first is an intracluster correlation that describes the similarity of outcomes for participants in the same cluster recruited during the same time period (a within-period intracluster correlation). The second parameter is known as the cluster autocorrelation, which describes how this within-period intracluster correlation changes when a pair of participants in the same cluster but differing time periods is considered. When the discrete time decay model is assumed, values for the within-period intracluster correlation and cluster autocorrelation must once again be specified; however, the cluster autocorrelation now describes how the intracluster correlation changes when a pair of participants in the same cluster but adjacent periods is considered. We will refer to an intracluster correlation that quantifies the similarity of participants in a cluster regardless of timing of measurement (i.e. that assumed under the exchangeable model) as an *aggregate* intracluster correlation, so-called because rather than separating measurements into distinct trial periods, all measurements are aggregated together in the estimation of this intracluster correlation.

The discrete time decay model was introduced relatively recently.[5] The Shiny CRT app allows for the discrete time decay model in sample size calculations,[1] allowing researchers to input values for the intracluster correlation and the cluster autocorrelation. A repository of estimates of these discrete time decay model parameters was recently made available,[7] but to date there have been few longitudinal cluster randomised trials that have been designed assuming a discrete time decay correlation structure, and fewer still that have been analysed assuming this structure. The discrete time decay model can be fit in SAS, using the commercial ASReml-R package in R[8]; it also appears that the free R package glmmTMB can be used to fit this model.[9] However, with the exception of the repository by Korevaar et al.,[7] few estimates of the correlation parameters associated with the discrete time decay model are available; researchers may thus be unsure what values of these parameters are reasonable in their context.

In contrast, several repositories of aggregate intracluster correlation estimates, estimated assuming the exchangeable model, are available.[10,11] Martin et al.[11] also included correlation parameter estimates for the block exchangeable model, while Korevaar et al.[7] included estimates for the exchangeable and block-exchangeable models for a bank of datasets. The exchangeable and block exchangeable models can be fit using any standard statistical software, for example using the free lme4 package in R,[12] in Stata, in SPSS, or in SAS, and estimates of the corresponding correlation parameters for individual trial outcomes are often available in published trial reports. An important question is whether the estimates of aggregate intracluster correlations that are available in such repositories and reports can tell us anything about plausible values of the parameters associated with a decaying correlation structure. That is, can we transform the estimates of intracluster correlation parameters obtained by fitting exchangeable or block exchangeable models to get estimates of the discrete time decay parameters? The answer is yes, such transformations are possible: using the expressions for estimated correlation components when an exchangeable or block exchangeable model have been incorrectly fitted instead of the discrete time decay model given by Kasza and Forbes,[13] we show how to obtain such transformed values. Such a transformation was previously applied by Kasza et al.[14] to obtain values of the discrete time decay cluster autocorrelation from published estimates of correlation parameters for the block exchangeable model. For an aggregate intracluster correlation obtained via the exchangeable model, a range of values of the discrete time decay correlation parameters will be consistent with this aggregate intracluster correlation.

In Section 2, we present the exchangeable, block exchangeable and discrete time decay models. In Section 3, we present the equations to be solved to transform correlation estimates obtained using the exchangeable or block exchangeable models to correlation estimates compatible with the discrete time decay model. We also include the equation for obtaining plausible values of the block exchangeable model's correlation parameters from the estimate of the intracluster correlation obtained from the exchangeable model. In Section 4, we apply the proposed method to two examples, demonstrating the use of our

Table 1. Notation for the intracluster correlations, between-period intracluster correlations, and cluster autocorrelations for each of the three models considered.

Model	Within-period intracluster correlation $\mathrm{corr}\left(Y_{kti}, Y_{ktj}\right)$	Between-period intracluster correlation $\mathrm{corr}\left(Y_{kti}, Y_{ksj}\right)$	Cluster autocorrelation $\frac{\mathrm{corr}\left(Y_{kti}, Y_{k(t+1)j}\right)}{\mathrm{corr}\left(Y_{kti}, Y_{ktj}\right)}$
Exchangeable	ρ_e	ρ_e	1
Block exchangeable	ρ_b	$r_b \rho_b$	r_b
Discrete time decay	ρ_d	$r_d^{|t-s|} \rho_d$	r_d

Y_{kti} is the observation for participant i in period t in cluster k. We refer to ρ_e as the aggregate intracluster correlation.

freely available online app (https://monash-biostat.shinyapps.io/ConsistentCACICC/). Using estimates by Korevaar et al.,[7] in Section 5, we empirically evaluate our method by comparing the values for the cluster autocorrelation obtained using our method to those obtained when the discrete time decay model is directly fitted to a set of 16 datasets. In Section 6, we conclude with a short discussion.

2 The exchangeable, block exchangeable and discrete time decay models

We now present the three models that we will consider for continuous outcomes Y_{kti} for participant $i = 1, \ldots, m$ in period $t = 1, \ldots, T$ in cluster $k = 1, \ldots, K$. In each model, β_t is the fixed effect corresponding to period t; X_{kt} is the treatment effect indicator for cluster k in period t, with $X_{kt} = 1$ if cluster k implements the intervention in period t and 0 otherwise; θ is the treatment effect of interest; and ϵ_{kti} is the error term for participant i in period t in cluster k. The three models differ only in the included random effect terms. Table 1 summarises the correlation parameters for these three models.

The three models we present in this section have previously appeared in the literature. For example, in the context of the stepped wedge design, the exchangeable model was originally presented by Hussey and Hughes,[2] the block exchangeable model was presented by Girling and Hemming[3] and Hooper et al.,[4] and the discrete time decay model was originally presented by Kasza et al.[5] For convenience, we present these models here.

- The exchangeable model has the form:

$$Y_{kti} = \beta_t + \theta X_{kt} + \alpha_k + \epsilon_{kti}, \quad \alpha_k \sim N\left(0, \sigma_{1\alpha}^2\right), \ \epsilon_{kti} \sim N\left(0, \sigma_{1\epsilon}^2\right) \tag{1}$$

This model includes a random effect for cluster, α_k. The aggregate intracluster correlation for the exchangeable model is given by $\rho_e = \mathrm{corr}\left(Y_{kti}, Y_{ksj}\right) = \sigma_{1\alpha}^2/(\sigma_{1\alpha}^2 + \sigma_{1\epsilon}^2)$.

- The block exchangeable model (sometimes referred to as the nested exchangeable model[15]) has the form:

$$Y_{kti} = \beta_t + \theta X_{kt} + \alpha_k + \gamma_{kt} + \epsilon_{kti}, \quad \alpha_k \sim N(0, \sigma_{2\alpha}^2), \ \gamma_{kt} \sim N(0, \sigma_\gamma^2), \ \epsilon_{kti} \sim N(0, \sigma_{2\epsilon}^2) \tag{2}$$

In addition to the random effect for cluster, α_k, this model also includes a random effect for cluster-period, γ_{kt}. The within-period intracluster correlation for the block exchangeable model is given by $\rho_b = \mathrm{corr}(Y_{kti}, Y_{ktj}) = (\sigma_{2\alpha}^2 + \sigma_\gamma^2)/(\sigma_{2\alpha}^2 + \sigma_\gamma^2 + \sigma_{2\epsilon}^2)$. The between-period intracluster correlation for this model is given by $\mathrm{corr}(Y_{kti}, Y_{ksj}) = \sigma_{2\alpha}^2/(\sigma_{2\alpha}^2 + \sigma_\gamma^2 + \sigma_{2\epsilon}^2)$. The cluster autocorrelation for the block exchangeable model is $r_b = \mathrm{corr}\left(Y_{kti}, Y_{ksj}\right)/\rho_b = \sigma_{2\alpha}^2/(\sigma_{2\alpha}^2 + \sigma_\gamma^2)$.

- The discrete time decay model has the form:

$$\begin{aligned} Y_{kti} &= \beta_t + \theta X_{kt} + \alpha_{kt} + \epsilon_{kti}, \quad \epsilon_{kti} \sim N(0, \sigma_{3\epsilon}^2), \\ \alpha_k &= (\alpha_{k1}, \ldots, \alpha_{kT})^T \sim N(\mathbf{0}, \sigma_{3\alpha}^2 \Sigma) \end{aligned} \tag{3}$$

where Σ is a $T \times T$ correlation matrix with entries $\Sigma[t, s] = r_d^{|t-s|}$. This model now includes dependent random effects for cluster-period, α_{kt}. The within-period intracluster correlation for the discrete time decay model is given by $\rho_d = \mathrm{corr}\left(Y_{kti}, Y_{ktj}\right) = \sigma_{3\alpha}^2/(\sigma_{3\alpha}^2 + \sigma_{3\epsilon}^2)$. The between-period intracluster correlation for this model is given by $\mathrm{corr}\left(Y_{kti}, Y_{ksj}\right) =$

$r_d^{|t-s|} \sigma_{3\alpha}^2 / (\sigma_{3\alpha}^2 + \sigma_{3\epsilon}^2) = r_d^{|t-s|} \rho_d$, so that the cluster autocorrelation for the discrete time decay model is given by r_d (the ratio of $\mathrm{corr}(Y_{kti}, Y_{k(t+1)j})$ to $\mathrm{corr}(Y_{kti}, Y_{ktj})$). When $\rho_b = \rho_d$ and $r_d = r_b = 1$, all three models are equivalent. When $T = 2$, the block exchangeble and discrete time decay models are equivalent.

3 Obtaining consistent values for the discrete time decay intracluster correlation and cluster autocorrelation

We now consider three different scenarios. In the first scenario, estimates are available from the block exchangeable model, and need to be applied to a planned design under a discrete time decay model. In the second scenario, an estimate of the aggregate intracluster correlation is available from an exchangeable model and needs to be applied to a planned design under a discrete time decay model. Third, we provide a method for obtaining estimates for the block exchangeable correlation parameters from an estimate of the aggregate intracluster correlation.

We note that these methods apply to any type of longitudinal cluster randomised trial, including but not limited to the cluster crossover and stepped wedge designs. These results extend the work that was presented by Kasza and Forbes,[13] where the focus was on the impact of misspecification of the within-cluster correlation structure on inference for the treatment effect. Here we use the equations presented in that previous paper in a novel way to obtain values of correlation parameters associated with the discrete time decay and block exchangeable models that would have led to the observed exchangeable or block exchangeable correlation estimates. These conversion formulas have not been published previously. The derivation of results for our proposed approach is provided in the Supplemental Material.

3.1 When estimates are available from the block exchangeable model

If the block exchangeable model has been fit to clustered data, estimates of the within-period intracluster correlation $(\hat{\rho}_b)$ and cluster autocorrelation $(\hat{r}_b)$ will be available. An investigator planning a trial under an assumed discrete time decay model could nevertheless use these estimates to obtain relevant sample size parameter values that are consistent with those obtained under the block exchangeable model. Given these values, it is possible to identify values of the intracluster correlation and cluster autocorrelation from the discrete time decay model (ρ_d and r_d) that are consistent with $\hat{\rho}_b$ and $\hat{r}_b$. Put another way, we can determine values of ρ_d and r_d that could have generated data that resulted in estimates $\hat{\rho}_b$ and $\hat{r}_b$ after analysis with the incorrect block exchangeable model. The r_d term obtained by solving the polynomial below is the discrete time decay cluster autocorrelation that corresponds to periods of the same length as were used in the original calculation of $\hat{r}_b$.

When estimates are available from the block exchangeable model, two equations with two unknowns must be solved. Hence, unique values for ρ_d and r_d are obtained. In the Supplemental Material, we show that applying the results in Table 2 of Kasza and Forbes[13] indicates that $\rho_d = \hat{\rho}_b$ and

$$\frac{1}{T(T-1)} \sum_{t=1}^{T} \sum_{s=1}^{T} r_d^{|t-s|} - \frac{1}{T-1} = \hat{r}_b \tag{4}$$

This equation states that the estimated cluster autocorrelation for the block exchangeable model is the average of all values of the within-cluster correlation matrix of cluster-period means, divided by ρ_d. This implies that to obtain a value of r_d that is compatible with the estimate obtained by fitting the (incorrect) block exchangeable model, the roots of the following polynomial in r_d must be found:

$$f(r_d) = \frac{1}{T(T-1)} \sum_{t=1}^{T} \sum_{s=1}^{T} r_d^{|t-s|} - \frac{1}{T-1} - \hat{r}_b \tag{5}$$

The double sum over the r_d terms can be written as

$$f(r_d) = (T-1) r_d + (T-2) r_d^2 + (T-3) r_d^3 + \cdots + 2r_d^{T-2} + r_d^{T-1} - \frac{T(T-1)}{2} \hat{r}_b \tag{6}$$

The roots of this polynomial can be found in R using the functions in the package 'polynom',[16] or in Stata using the Mata command 'polyroots', for example. We provide an online app to find these roots, and demonstrate its use in Section 4.

3.2   When estimates are available from the exchangeable model

The situation is more complicated when an estimate of the aggregate intracluster correlation is obtained from the exchangeable model: as we show in the Supplemental Material, there is now a single equation in two unknown parameters (r_d and ρ_d) to be solved. If $\hat{\rho}_e$ is the estimated aggregate intracluster correlation from the exchangeable model, the following equation must be solved for r_d and ρ_d:

$$
\begin{aligned}
&(T-1)r_d + (T-2)r_d^2 + (T-3)r_d^3 + \cdots + 2r_d^{T-2} + r_d^{T-1} \\
&+ \frac{T}{2}\left(1 - \frac{\hat{\rho}_e}{\rho_d}\frac{KTm - K - T + 1}{Km - 1 - \hat{\rho}_e m + \hat{\rho}_e} - \frac{(K - 1 + \hat{\rho}_e K + \hat{\rho}_e T - \hat{\rho}_e Tm)}{Km - 1 - \hat{\rho}_e m + \hat{\rho}_e}\right) = 0
\end{aligned}
\tag{7}
$$

To solve this equation, we fix values of r_d ranging from 1 to 0, and then solve this equation for ρ_d.

Alternatively, ρ_d could be fixed and the polynomial above solved for r_d. When taking this approach, some values of ρ_d may lead to negative values of r_d; while negative cluster autocorrelations are theoretically possible, they may not be plausible in practice. Negative cluster autocorrelations indicate that the correlations between pairs of observations oscillate between being positive or negative, depending on the number of periods between them. This oscillation does not appear to be plausible, and therefore, we recommend that researchers discard values of ρ_d that lead to negative values of r_d if fixing ρ_d and solving for r_d.

In addition to the above, we propose an approximation to equation (7) which can be used when K and m are unknown:

$$
(T-1)\,r_d + (T-2)\,r_d^2 + (T-3)\,r_d^3 + \cdots + 2r_d^{T-2} + r_d^{T-1} + \frac{T}{2}\left(1 - \frac{\hat{\rho}_e}{\rho_d}T\right) = 0
\tag{8}
$$

This approximation is equivalent to taking the average of all values in the correlation matrix of the cluster-period means for a single cluster and setting this equal to $\hat{\rho}_e$. We show in the Supplemental Material that this approximation will be valid when $K \times m$ is large relative to K, $\rho_d \times m$ and $\hat{\rho}_e \times m$. We demonstrate how our online app can be used to obtain solutions of this equation in Section 4.

3.3   Obtaining values for the block exchangeable correlation parameters from estimates from the exchangeable model

We also consider the situation when researchers wish to obtain estimates of the block exchangeable correlation parameters from an estimate of the aggregate intracluster correlation. As was the case for obtaining estimates for discrete time decay correlation parameters from an estimate of the aggregate intracluster correlation, there is a single equation in two unknowns that must be solved. If $\hat{\rho}_e$ is the estimated aggregate intracluster correlation from the exchangeable model, the following equation must be solved for r_b and ρ_b:

$$
r_b + \frac{\hat{\rho}_e}{\rho_b}\frac{KTm - K - T + 1}{m(K + T - 1 - KT)} + \frac{(m-1)(1 - T - K)}{m(K + T - 1 - KT)} = 0
\tag{9}
$$

An option in our online app allows researchers to investigate the solutions of this equation for specific choices of $\hat{\rho}_e$, K, T and m.

## 4   Use of this approach in practice

4.1   Application to the diabetes data of Martin et al.[11]

Martin et al.[11] provided a repository of intracluster correlations for the exchangeable and block exchangeable models for data from The Health Improvement Network (THIN) database. A range of outcomes from patients aged 18 years or over with a type-2 diabetes diagnosis, treated in one of 430 general practices in the United Kingdom were considered. Here we consider the HbA1c outcome, measured in patients between 1 January 2007 and 31 December 2007 (i.e. a period of 12 months). The estimated aggregate intracluster correlation for HbA1c over this 12-month period was 0.032. We assume 241 patients per practice in each 12-month period.

We suppose that interest is in planning a stepped wedge trial with periods of length three months each, and we wish to obtain values of the discrete time decay model intracluster correlation and cluster autocorrelation that are consistent with the value of 0.032 that was obtained by fitting the exchangeable model, although it is unlikely that period effects

Does it decay? Obtaining decaying correlation parameter values from previously analysed longitudinal cluster randomised trials
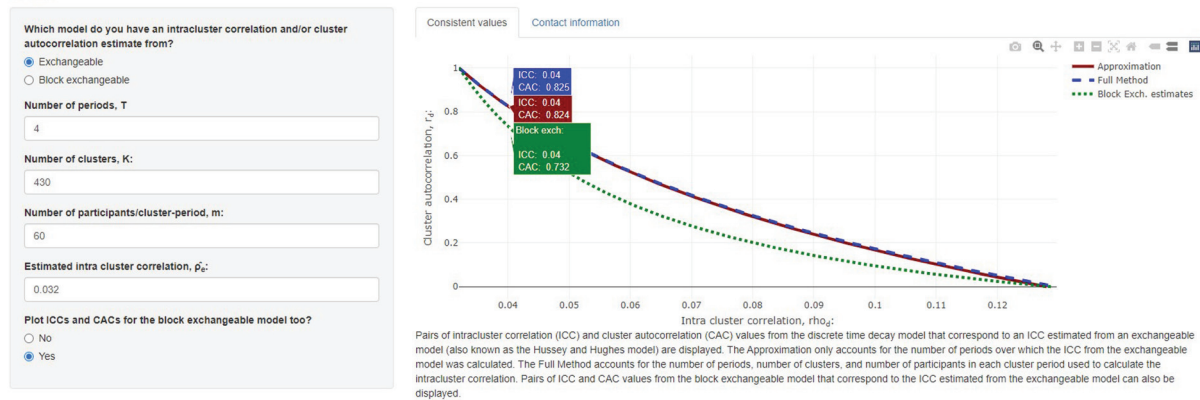
Figure 1. The pairs of intracluster correlation and cluster autocorrelation values that are aligned with the intracluster correlation obtained for the Hb1Ac outcome in. One particular set of values is highlighted: $\rho_d = 0.04$ and $r_d = 0.825$ are consistent with $\hat{\rho}_e = 0.032$ when $T = 4$, $K = 430$ and $m = 60$ (using equation (10)). This is very close to the pair of values obtained using the approximation in equation (11): $\rho_d = 0.04$ and $r_d = 0.824$ are consistent with $\hat{\rho}_e = 0.032$ when $T = 4$. This plot also displays the block exchangeable correlation parameters that correspond to the estimated aggregate intracluster correlation.

were included in this model. We assume minimal period effects here. The 12-month period that was used to calculate this intracluster correlation thus corresponds to $T = 4$ three-month periods. Were researchers instead interested in planning studies with periods of length six months, for example, then for this example, the number of periods $T = 2$ would be input into the calculations, since 12 months of data was used to calculate $\hat{\rho}_e = 0.032$. For three-month periods, the following equation needs to be solved:

$$3r_d + 2r_d^2 + r_d^3 + \frac{4}{2}\left(1 - \frac{0.032}{\rho_d}\frac{4Km - K - 4 + 1}{Km - 1 - 0.032m + 0.032} - \frac{(K - 1 + 0.032K + 0.032 \times 4 - 0.032 \times 4 \times m)}{Km - 1 - 0.032m + 0.032}\right) = 0 \quad (10)$$

With $K = 430$ and 241 patients in each practice in a year, then $m = 241/4 \approx 60$, and the equation to be solved is given by

$$3r_d + 2r_d^2 + r_d^3 + \frac{4}{2}\left(1 - \frac{0.032}{\rho_d}3.98 - 0.017\right) = 0$$

Since $K \times m = 25,800$ is large relative to K and $\hat{\rho}_e \times m$, this equation is very similar to the proposed approximation

$$3r_d + 2r_d^2 + r_d^3 + \frac{4}{2}\left(1 - \frac{0.032}{\rho_d}4\right) = 0 \quad (11)$$

These equations could be solved in Stata or R. Alternatively, the estimated intracluster correlation, the number of periods, number of clusters, and number of participants per cluster-period can be inserted into the online app available at https://monash-biostat.shinyapps.io/ConsistentCACICC/, developed using R Shiny.[17] Figure 1 displays a screenshot of the online app, and indicates that the solutions of equations (10) and (11) are very similar. A wide range of values of ρ_d and r_d are consistent with $\hat{\rho}_e = 0.032$, from $(\rho_d, r_d) = (0.032, 1)$ to $(0.127, 0.005)$. However, it would seem appropriate to place further bounds on these values; for example, it may be believed that an intracluster correlation of greater than around 0.05 is unlikely, thus limiting r_d to be greater than around 0.66.

The range of plausible values can then be used to inform sample size and power calculations, with researchers considering the sensitivity of their calculations to various choices of ρ_d and r_d consistent with the estimate $\hat{\rho}_e = 0.032$. We consider an example, where a three-sequence stepped wedge trial is planned with a cross-sectional sampling structure and 60 patients in each cluster-period, aiming to detect a standardised effect size of 0.1 with a two-sided significance level of 0.05. We use the Shiny CRT sample size calculator[1] to determine how many clusters must be randomised to each sequence of a stepped wedge design to ensure at least 80% power. We consider the originally estimated intracluster correlation corresponding to the exchangeable model and three sets of values for the discrete time decay model that are consistent with the originally estimated $\hat{\rho}_e = 0.032$, given in Table 2. If an exchangeable within-cluster correlation structure was assumed, with

Table 2. The number of clusters required to be randomised to each sequence of a three-sequence stepped wedge design to detect a standardised effect size of 0.1 with 80% power and two-sided significance level of 0.05.

ρ_d	r_d	Number of clusters per sequence
0.032	1	15
0.034	0.95	16
0.040	0.83	21
0.050	0.66	29

The values of ρ_d and r_d selected are consistent with the estimate of $\hat{\rho}_e = 0.032$ estimated in by applying the exchangeable model to the HbIAIc outcomes in The Health Improvement Network (THIN) dataset.

intracluster correlation $\rho_e = 0.032$, then randomising 15 clusters to each sequence (for a total of 45 clusters) would give at least 80% power to detect this effect. However, if values of the discrete time decay model parameters that are consistent with this estimate of 0.032 are considered, the results in Table 2 indicate that the study may require more clusters to detect the effect of interest with the same level of power. When planning a study such as this, researchers may wish to ensure that a sufficient number of clusters are randomised to ensure that the power of the study will be robust to reasonable deviations from an exchangeable correlation structure. Researchers may also wish to validate theoretical power calculations via simulation, particularly in settings with smaller numbers of clusters.

4.2 The rapid atrial fibrillation and flutter 3 (RAFF-3) trial in Canadian Emergency Departments

Acute atrial fibrillation (AF) and flutter (AFL) are common conditions seen in the Emergency Department (ED). In 2018, the Canadian Association of Emergency Physicians (CAEP) endorsed the acute AF/AFL best practices checklist which provides specific guidance for treatment in the ED. The RAFF-3 trial was a cross-sectional stepped-wedge cluster randomised design to evaluate implementation of the checklist at 11 large community and academic hospital EDs in Canada.[18] All 11 EDs started in the control condition (usual care). One hospital then crossed over to the intervention condition each month. There was a 2-month transition period to allow for implementation of the intervention. The primary outcome was length of stay in the ED in minutes from time of arrival to the documented time of discharge or hospital admission. The trial was designed to achieve at least 80% power to detect a 100-min reduction in ED length of stay, assuming an average of 10 patients per site per month, a standard deviation of 250 min, a within-period intracluster correlation of $\rho_d = 0.05$ and a cluster autocorrelation coefficient of $r_d = 0.8$ in a discrete time decay model and using a two-sided test at the 5% significance level. At the time of trial planning, the correlation estimates were based on commonly reported rules of thumb,[4,19] and, using the Shiny CRT calculator,[1] achieved 86% power to detect the target difference.

To illustrate a more careful power calculation for the RAFF-3 trial, we consider routinely collected data on length of stay for a similar patient population obtained from 15 EDs with an average of 240 patients per ED over a period of one year. For privacy reasons, patient-level data could not be obtained but the aggregate intracluster correlation was provided as $\hat{\rho}_e = 0.05$. The first step is to obtain plausible values for the within-period intracluster correlation corresponding to periods of length one month and the monthly rate of decay consistent with an intracluster correlation of 0.05 over 12 months' duration. After inserting $K = 15$ clusters, $T = 12$ one-month periods and an average of $m = 20$ patients per ED per month into the Shiny app, we obtain a wide range of values of ρ_d and r_d that are consistent with $\hat{\rho}_e = 0.05$, from $(\rho_d, r_d) = (0.05, 1)$ to $(0.598, 0.008)$. By assuming a plausible upper bound of 0.2 for the within-period intracluster correlation and allowing for a small amount of decay in the strength of the correlation of at least 5% (i.e. the cluster autocorrelation $r_d < 0.95$), a reasonable range of values to assume in a sensitivity analysis might be $(0.061, 0.949)$ to $(0.2, 0.552)$. Figure 2 displays the pairs of plausible values.

Using the Shiny CRT sample size calculator, we enter a stepped wedge cross-sectional design with 11-sequences, one cluster per sequence, a total duration of 14-months with a 2-month implementation period and 10 patients in each cluster-period, aiming to detect a standardised effect size of 0.4 with a two-sided significance level of 0.05. The power that can be achieved with the reasonable combinations of within-period intracluster correlation and cluster autocorrelations is shown in Table 3. If an exchangeable model were assumed, this design will achieve 96.2% power. However, if the plausible values under the discrete time decay model are considered, power will drop substantially using the same design and could be as low as 54.7%.

We demonstrate that the issue of the selection of a suitable within-cluster correlation structure and associated parameters is often much less problematic for parallel cluster randomised trials. We consider a parallel cluster randomised trial variant of the RAFF-3 stepped wedge, with five clusters randomised to each arm of a 12-period parallel cluster randomised trial,

Does it decay? Obtaining decaying correlation parameter values from previously analysed longitudinal cluster randomised trials
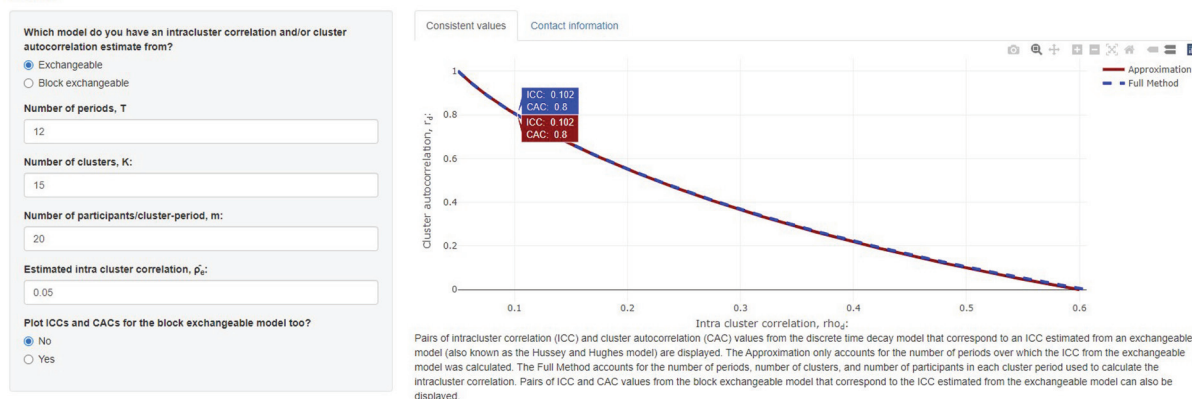
Figure 2. The pairs of intracluster correlation and cluster autocorrelation values that are consistent with the intracluster correlation obtained for the length of ED stay in RAFF-3 trial. One particular set of values using full method is highlighted: $\rho_d = 0.102$ and $r_d = 0.8$ are consistent with $\hat{\rho}_e = 0.05$ when $T = 12$, $K = 15$ and $m = 20$. In this case, this is identical to the pair of values obtained using the approximation.
RAFF-3: rapid atrial fibrillation and flutter 3; ED: Emergency Department.

Table 3. The power that can be achieved by the rapid atrial fibrillation and flutter 3 (RAFF-3) trial to detect a standardised effect size of 0.4 with two-sided significance level of 0.05.

ρ_d	r_d	Power
0.050	1.000	0.962
0.061	0.949	0.905
0.102	0.800	0.714
0.200	0.552	0.547

The values of ρ_d and r_d are consistent with $\hat{\rho}_e = 0.05$ by applying the exchangeable model to the length of stay outcomes in the RAFF-3 trial.

Table 4. The power that can be achieved by a parallel variant of the rapid atrial fibrillation and flutter 3 (RAFF-3) trial to detect a standardised effect size of 0.4 with two-sided significance level of 0.05.

ρ_d	r_d	Power
0.050	1.000	0.748
0.061	0.949	0.751
0.102	0.800	0.765
0.200	0.552	0.768

The values of ρ_d and r_d are consistent with $\hat{\rho}_e = 0.05$ by applying the exchangeable model to the length of stay outcomes in the RAFF-3 trial.

with 10 participants in each cluster in each period. Using the same set of correlation parameters as given in Table 3, we determined the power of this parallel variant of the RAFF-3 trial to detect an effect size of 0.4, with the results presented in Table 4. Although the power of the design does change depending on the correlation parameters, this example demonstrates that the parallel design is much less sensitive to the precise choice of correlation parameters than the stepped wedge design.

5 Application to datasets in the CLOUD Bank repository

In Korevaar et al.,[7] several longitudinal clustered datasets (known as the CLustered OUtcomes Dataset Bank) were analysed using all three of the models presented in Section 2, and we consider the 12 studies where each participant provided only one measurement of each outcome during the course of the study. One study included two outcomes, and another included four, for a total of 16 datasets. The name of each study, the design of each study, the number of clusters (K), time periods (T), and the average number of participants in each cluster-period (m) are provided in Table 5. In addition, $\hat{\rho}_b$ and $\hat{r}_b$ (obtained by fitting the block exchangeable model to each dataset) and $\hat{\rho}_d$ and $\hat{r}_d$ (obtained by fitting the discrete time decay model to each dataset) are included. These estimates are available directly from the CLOUD Bank online

Table 5. Estimated correlation parameters obtained for CLOUD Bank datasets, obtained by fitting the block exchangeable and discrete time decay models, and the discrete time decay cluster autocorrelation obtained by applying equation (6).

Dataset	Study design	K	T	m	$\hat{\rho}_b$	$\hat{r}_b$	$\hat{\rho}_d$	$\hat{r}_d$	r_d (equation (6))
APD ICU	Observational	126	7	173	0.046	0.981	0.045	0.992	0.993
Alive&Thrive (Bangladesh)	Parallel w/baseline	20	2	49.4	0.066	0.896	0.066	0.896	0.896
Alive&Thrive (Vietnam)	Parallel w/baseline	40	2	29	0.062	0.959	0.062	0.959	0.959
Dementia referral	Parallel w/baseline	22	3	11	0.176	1	0.176	1	1
Disinvestment2	Stepped wedge	12	8	133	0.085	0.946	0.077	0.973	0.981
Disinvestment	Stepped wedge	12	9	138	0.078	0.908	0.081	0.96	0.971
MORDOR	Parallel w/baseline	30	3	37	0.113	0.324	0.111	0.246	0.404
OXTEXT7	Stepped wedge	11	16	13	0.048	0.492	0.047	0.8	0.863
PITHIA	Observational	22	6	8	0.056	0.08	0.053	−0.279	0.202
PROMPT (1)	Stepped wedge	4	7	20	0.012	0	0.024	0.615	0
PROMPT (2)	Stepped wedge	4	7	20	0.162	0.174	0.031	0.249	0.407
Syncope	Observational	6	14	119	0.222	0.858	0.198	0.951	0.969
THIN (1)	Observational	430	4	60	0.021	0.854	0.021	0.889	0.908
THIN (2)	Observational	430	4	60	0.04	0.904	0.04	0.915	0.941
THIN (3)	Observational	430	4	60	0.023	0.837	0.023	0.877	0.897
THIN (4)	Observational	430	4	60	0.02	0.88	0.02	0.914	0.925

THIN: The Health Improvement Network; CLOUD Bank: CLustered OUtcomes Dataset Bank.
K is the total number of clusters, T is the total number of time periods, and m is the average number of participants in each cluster-period. Some studies included more than one continuous outcome, and this is indicated by a number in brackets after the study name.

app (https://monash-biostat.shinyapps.io/CLOUDbank/). We applied equation (6) to find values of r_d consistent with the estimates of $\hat{\rho}_b$ and $\hat{r}_b$ obtained from the block exchangeable model, and compared these to the actual observed values of $\hat{r}_d$. That is, we compared the results of our method to the estimates obtained by fitting the discrete time decay model directly to the datasets.

Table 5 shows that the use of equation (6) resulted in similar estimates to those directly obtained by fitting the model when the model-based cluster autocorrelation was high (greater than around 0.85). Twelve of the 16 datasets had estimates of r_d that were in this range. The small differences observed for these datasets are likely to be due to departures from the assumptions of equation (6); in particular, the assumption of an equal number of participants in each cluster in each period of the study.

There were quite large differences between the model-based estimate of r_d and the value obtained through application of equation (6) for four datasets: MORDOR, PITHIA, PROMPT (1) and PROMPT (2). In MORDOR and PROMPT (2), the discrete time decay cluster autocorrelation obtained by directly fitting the model is around 0.25, while that obtained by solving equation (6) is around 0.4. In PROMPT (1), the cluster autocorrelation estimate for the block exchangeable model was 0: thus, the polynomial in equation (6) has a root $r_d = 0$. For the PITHIA trial, the estimate $\hat{r}_d$ obtained by fitting the model was negative, while the solution of equation (6) was positive. These four datasets were the only ones to have model-based estimates of r_d less than 0.65. This indicates that when the value of the autocorrelation is small, the value of r_d obtained by solving equation (6) may be inflated.

6 Discussion

In this article, we have described an approach for obtaining plausible values of the correlation parameters associated with the discrete time decay when correlation parameter estimates are available from the exchangeable or block exchangeable model; and for obtaining plausible values for the correlation parameters associated with the block exchangeable model when correlation parameter estimates are available from the exchangeable model. These methods assume that the estimates of the parameters from the exchangeable or block exchangeable model have come from a balanced dataset: the same number of participants has provided data in all clusters in all periods, and each participant has only provided data in one period. Hence, these methods should not replace the actual estimation of the model parameters in R or SAS: if raw data are available to be analysed, these data should be analysed to provide estimates of the within-period intracluster correlation and cluster autocorrelation for the discrete time decay or block exchangeable model. Further, the methods we have presented assume that a linear mixed model for outcomes is appropriate, and thus their use for non-continuous outcomes may not be valid.

When planning a longitudinal CRT, researchers may inadvertently use an anticipated aggregate intracluster correlation in lieu of the within-period ICC in the sample size calculation procedure. However, this could lead to under-estimating the required sample size as the within-period ICC must be inflated to allow for the decaying cluster autocorrelation and is therefore necessarily larger than the aggregate ICC. That is, $\rho_d \geq \hat{\rho}_e$ and $\rho_b \geq \hat{\rho}_e$. The work in this article provides some guidance on just how much ρ_d and ρ_b should be inflated, for corresponding choice of cluster autocorrelations. In lieu of these proposed methods, researchers must rely on rules-of-thumb, such as simply using a cluster autocorrelation of 0.5 or 0.8, for example. Such an approach may lead to selections of values of the discrete time decay correlation parameters that are incompatible with aggregate intracluster correlations, and makes sensitivity analyses across different correlation structures complicated. To truly understand the implications of different correlation structure assumptions for planned study power, researchers need to choose compatible input parameters across these different structures.

Application of the method to the datasets in the CLOUD Bank indicates that there is often good agreement between the estimates of r_d obtained by directly fitting the discrete time decay model and those obtained by applying equation (6) to estimates from the block exchangeable model. However, when direct application of the discrete time decay model led to estimates of the cluster autocorrelation that were around 0.6 or smaller, the values obtained by applying equation (6) were not very close to these directly obtained estimates. It seems that such small estimates of the cluster autocorrelation may be rare (with only four of the 16 considered datasets having such small cluster autocorrelations). However, when there may be quite substantial amounts of decay in the intracluster correlation over time (i.e. small cluster autocorrelations), we recommend that the methods presented here be used with caution.

The block exchangeable and discrete time decay models that we have presented depend on the notion of trial 'periods': trials are assumed to be split into time periods of equal length. For these models, the intracluster correlation and cluster autocorrelation depend on the length of these trial periods, and thus need to be recalculated when interest is in periods of differing lengths. Recent work has suggested that in certain circumstances it would be more appropriate to consider time as a continuous phenomenon in longitudinal cluster randomised trials, especially when recruitment of participants occurs continuously throughout each study period.[20,21] A continuous time decay correlation structure that considers the precise

times of measurement rather than the period of measurement in defining the correlations between pairs of observations was introduced by Grantham et al.[22]

The methods that we have presented allow researchers to obtain some understanding of plausible values of the correlation parameters associated with the discrete time decay model, even when such a model has not been fit to the dataset. Although the paper introducing the discrete time decay model has been quite influential in the stepped wedge literature,[23] few longitudinal cluster randomised trials have yet been designed or analysed assuming a discrete time decay (with the exception of the secondary analyses of several studies that was presented by Korevaar et al.[7]). Hence, when planning longitudinal cluster randomised trials, there are few estimates of the discrete time decay correlation parameters available for researchers to rely upon. The methods we have presented allow researchers to obtain plausible values to be used in sample size and power calculations, and have been implemented in an online app available at https://monash-biostat.shinyapps.io/ConsistentCACICC/.

Declaration of conflicting interests

Funding

ORCID iDs

Jessica Kasza 🔟 https://orcid.org/0000-0002-8940-0136
Rhys Bowden 🔟 https://orcid.org/0000-0001-9880-0206
Yongdong Ouyang 🔟 https://orcid.org/0000-0002-8692-2991

Supplemental material

Supplemental material for this article is available online.

References

1. Hemming K, Kasza J, Hooper R, et al. A tutorial on sample size calculation for multiple-period cluster randomized parallel, crossover and stepped-wedge trials using the shiny CRT calculator. *Int J Epidemiol* 2020; **49**: 979–995.
2. Hussey MA and Hughes JP. Design and analysis of stepped wedge cluster randomized trials. *Contemp Clin Trials* 2007; **28**: 182–191.
3. Girling AJ and Hemming K. Statistical efficiency and optimal design for stepped cluster studies under linear mixed effects models. *Stat Med* 2016; **35**: 2149–2166.
4. Hooper R, Teerenstra S, de Hoop E, et al. Sample size calculation for stepped wedge and other longitudinal cluster randomised trials. *Stat Med* 2016; **35**: 4718–4728.
5. Kasza J, Hemming K, Hooper R, et al. Impact of non-uniform correlation structure on sample size and power in multiple-period cluster randomised trials. *Stat Methods Med Res* 2019; **28**: 703–716.
6. Ouyang Y, Li F, Preisser JS, et al. Sample size calculators for planning stepped-wedge cluster randomized trials: A review and comparison. *Int J Epidemiol* 2022; **51**: 2000–2013. dyac123.
7. Korevaar E, Kasza J, Taljaard M, et al. Intra-cluster correlations from the Clustered Outcome Dataset Bank to inform the design of longitudinal cluster trials. *Clinical Trials* 2021; **18**: 529–540.
8. Butler D, Cullis B, Gilmour A, et al. ASReml-R Reference Manual Version 4. Hemel Hempstead, HP1 1ES, UK, 2017.
9. Brooks ME, Kristensen K, van Benthem KJ, et al. glmmTMB balances speed and flexibility among packages for zero-inflated generalized linear mixed modeling. *R J* 2017; **9**: 378–400.
10. Lajos GJ, Haddad SM, Tedesco RP, et al. Intracluster correlation coefficients for the Brazilian Multicenter Study on Preterm Birth (EMIP): methodological and practical implications. *BMC Med Res Methodol* 2014; **14**: 54.
11. Martin J, Girling A, Nirantharakumar K, et al. Intra-cluster and inter-period correlation coefficients for cross-sectional cluster randomised controlled trials for type-2 diabetes in UK primary care. *Trials* 2016; **17**: 402.
12. Bates D, Mächler M, Bolker B, et al. Fitting linear mixed-effects models using lme4. *J Stat Softw* 2015; **67**: 1–48.
13. Kasza J and Forbes AB. Inference for the treatment effect in multiple-period cluster randomised trials when random effect correlation structure is misspecified. *Stat Methods Med Res* 2019; **28**: 3112–3122.
14. Kasza J, Hooper R, Copas A, et al. Sample size and power calculations for open cohort longitudinal cluster randomized trials. *Stat Med* 2020; **39**: 1871–1883.
15. Li F, Hughes JP, Hemming K, et al. Mixed-effects models for the design and analysis of stepped wedge cluster randomized trials: an overview. *Stat Methods Med Res* 2021; **30**: 612–639.

16. Venables B, Hornik K and Maechler M. polynom: A collection of functions to implement a class for univariate polynomial manip-
    ulations. https://CRAN.R-project.org/package=polynom R package version 1.4-0. S original by Bill Venables, packages for R by
    Kurt Hornik and Martin Maechler. 2019.

17. Chang W, Cheng J, Allaire JJ, et al. *shiny: Web Application Framework for R* 2021. R package version 1.7.1. URL
    https://CRAN.R-project.org/package=shiny.

18. Stiell IG, Archambault PM, Morris J, et al. RAFF-3 trial: A stepped-wedge cluster randomised trial to improve care of acute atrial
    fibrillation and flutter in the emergency department. *Can J Cardiol* 2021; **37**: 1569–1577.

19. Campbell MJ. Cluster randomized trials in general (family) practice research. *Stat Methods Med Res* 2000; **9**: 81–94.

20. Grantham KL, Kasza J, Heritier S, et al. How many times should a cluster randomized crossover trial cross over? *Stat Med* 2019;
    38: 5021–5033.

21. Hooper R and Copas A. Stepped wedge trials with continuous recruitment require new ways of thinking. *J Clin Epidemiol* 2019;
    116: 161–166.

22. Grantham KL, Kasza J, Heritier S, et al. Accounting for a decaying correlation structure in cluster randomized trials with continuous
    recruitment. *Stat Med* 2019; **38**: 1918–1934.

23. Murray DM. Influential methods reports for group-randomized trials and related designs. *Clinical Trials* 2022; **19**: 353–362.