Title

Intra-cluster correlations from the CLustered OUtcome Dataset (CLOUD) bank to inform the design of longitudinal cluster trials

Running heading

CLOUD bank ICC estimates

Word count: 3946

Authors

Elizabeth Korevaar[1], Jessica Kasza[1], Monica Taljaard[2,3], Karla Hemming[4], Terry Haines[5], Elizabeth L Turner[6,7], Jennifer A Thompson[8], James P Hughes[9] and Andrew B Forbes[1]

Affiliations

[1]School of Public Health and Preventive Medicine, Monash University, Melbourne, VIC, Australia
[2]Clinical Epidemiology Program, Ottawa Hospital Research Institute, Ottawa, Ontario, K1Y 4E9 Canada
[3]School of Epidemiology, Public Health and Preventive Medicine, University of Ottawa, Ottawa, Ontario, K1N 6N5, Canada
[4]Institute of Applied Health Research, University of Birmingham, Birmingham, UK.
[5]School of Primary and Allied Health Care, Monash University, Melbourne, VIC, Australia.
[6]Department of Biostatistics and Bioinformatics, Duke University, Durham, North Carolina.
[7]Duke Global Health Institute, Durham, North Carolina.
[8]Department of Infectious Disease Epidemiology, London School of Hygiene and Tropical Medicine, London, UK
[9]Department of Biostatistics, University of Washington, Seattle, WA USA

Corresponding author

Andrew B Forbes, School of Public Health and Preventive Medicine, Monash University, Level 4, 553 St Kilda Road, Melbourne, 3004, Victoria, Australia
Email: andrew.forbes@monash.edu
https://orcid.org/0000-0003-4269-914X

Grant support

Monash University Network of Excellence grant. 2018-2019. Forbes AB, Kasza J, Haines T, Hemming K, Taljaard M, Hooper R, Copas A, Hughes J, Preisser J, Turner EL. International Network for Innovative Cluster Randomised Trial Designs NOE180009.
Australian Government Research Training Program (RTP) Scholarship administered by Monash University, Australia.
Karla Hemming is funded by a NIHR Senior Research Fellowship SRF-2017-10-002.
JAT is jointly funded by the UK Medical Research Council (MRC) and the UK Department for International Development (DFID) under the MRC/DFID Concordat agreement and is also part of the EDCTP2 programme supported by the European Union MR/R010161/1.

Abstract

Background Sample size calculations for longitudinal cluster randomised trials, such as crossover and stepped-wedge trials, require estimates of the assumed correlation structure. This includes both within-period intra-cluster correlations which importantly differ from conventional intra-cluster correlations by their dependence on period), and also cluster autocorrelation coefficients to model correlation decay. There are limited resources to inform these estimates. In this paper we provide a repository of correlation estimates from a bank of real-world clustered datasets. These are provided under several assumed correlation structures namely exchangeable, block-exchangeable and discrete-time decay correlation structures.

Methods Longitudinal studies with clustered outcomes were collected to form the CLustered OUtcomes Dataset (CLOUD) bank. Forty-four available continuous outcomes from 29 datasets were obtained and analysed using each correlation structure. Patterns of within-period intra-cluster correlation coefficient (WP-ICC) and cluster autocorrelation (CAC) coefficients were explored by study characteristics.

Results The median WP-ICC for the discrete-time decay model was 0.05 (interquartile range IQR: 0.02 – 0.09) with a median CAC of 0.73 (IQR: 0.19 – 0.91). The WP-ICCs were similar for the exchangeable, block-exchangeable and discrete-time decay correlation structures. WP-ICCs and CACs were found to vary with the number of participants per cluster-period, the period-length, type of cluster (primary care, secondary care, community or school) and country income status (high income country or low- and middle-income country). The WP-ICCs tended to decrease with increasing period-length and slightly decrease with increasing cluster-period sizes, while the CACs tended to move closer to 1 with increasing cluster-period size. Using the CLOUD bank, an RShiny app has been developed for determining plausible values of correlation coefficients for use in sample size calculations.

Conclusion This study provides a repository of intra-cluster correlations and cluster autocorrelations for longitudinal cluster trials. This can help inform sample size calculations for future longitudinal cluster randomised trials.

Keywords

Intra-cluster correlation coefficient, within-period correlation, between-period correlation, cluster autocorrelation, discrete-time decay, cluster randomised trial, sample size

Introduction

Sample size calculation is an essential requirement for the design of any randomised trial. The optimum sample size is neither too large (to avoid wasting valuable resources) nor too small (to avoid an underpowered study). Importantly, in the design of cluster randomised trials (CRTs) it is well known that accounting for the correlation between observations within a cluster (e.g. village/hospital), quantified by the intra-cluster correlation coefficient (ICC), is critical for accurate determination of sample size[1,2]. Failing to account for this correlation in the design typically results in a calculated sample size that is too small, leading to an underpowered study. In the analysis, type-one error rates may be inflated[1].

Recently, longitudinal cluster randomised trials, where clusters are followed over multiple treatment periods, have become increasingly popular. Particular examples of longitudinal cluster randomised trials are stepped-wedge designs[3], cluster randomised crossover trials[4,5] and parallel cluster randomised trials with baseline and one or more multiple follow up times[6].

Sample size calculation procedures for longitudinal cluster randomised trials have been developed under three main types of correlation structures: the “exchangeable” correlation structure developed by Hussey and Hughes[7], the “block-exchangeable” correlation structure of Hooper et al. and Girling and Hemming[8,9], and the “discrete-time decay” within-cluster correlation structure introduced by Kasza et al.[10] These procedures require correlation parameters that quantify the degree of similarity between observations in each cluster and period. Under the exchangeable correlation structure, this is quantified by a single ICC[7]. The block-exchangeable and the discrete-time decay correlation structures require specification of the within-period ICC (WP-ICC) and the cluster autocorrelation coefficient (CAC), to allow the degree of similarity between observations of the same cluster but different periods to differ[2,8-10]. While these structures assume different participants are assessed in each period, they can be

extended to accommodate cohort designs[11]. These correlation structures and their underlying models are outlined in subsequent sections. Failing to appropriately account for the within-cluster correlation pattern when calculating sample sizes can lead to an under- or over-estimated sample size during trial design[10].

For conventional parallel cluster randomised trials, trialists often rely on repositories of correlation parameter estimates[12-16], estimates from similar trials, or use patterns and 'rules of thumb'. For example, Adams et al. estimated ICCs for 245 continuous variables from 31 cluster trials in primary care, identifying that larger cluster sizes and process outcomes were associated with larger ICCs[12]. There are few such resources for longitudinal correlation structures to inform sample size calculations. One example is Martin el al.[17] who estimated the ICC, WP-ICC and CAC for six continuous outcomes from a database of patients with type-2 diabetes from UK general practices. To date no consideration has been given to characteristics that may influence the longitudinal correlation parameters such as period-length, cluster sizes and type of cluster (e.g. primary vs secondary care) and we are unaware of any such repositories for the discrete-time decay correlation structure. Repositories exist for estimates of exchangeable[12,14,16,17] and block-exchangeable correlation parameters[17], derived from the analysis of the outcomes from single, large studies with clustered data[14,16,17], or outcomes across a collection of studies[12,16].

This study used an accumulated bank of cluster trial datasets with a variety of study designs and characteristics (hereafter referred to as the CLustered OUtcomes Dataset, or CLOUD, bank) to address the challenges presented above for continuous outcomes. Our specific objectives were to: 1) provide a database of correlation parameter estimates useful for designing multiple-period cluster randomised trials; and 2) describe the patterns in the correlation estimates in relation to key study design and trial characteristics.

Below we outline the methods used to address these objectives, including datasets collected to form the CLOUD bank and the three correlation structures applied to each dataset. We describe

the resulting repository and patterns observed in the correlation estimates, and provide an RShiny app[18] for readers to easily access and visualise the estimates from these analyses. Finally, we demonstrate how these estimates can be used to inform sample size calculations.

Methods

Datasets: Identifying and selecting studies for the CLOUD bank

Datasets were sourced in two ways: co-authors of this manuscript had existing permission for use of a dataset, or recommended an open access data repository. Online repositories (Dryad, London School of Hygiene and Tropical Medicine data compass, PLoS Medicine and UK Dataservice) were searched for 'cluster randomised' keywords (and other spelling variations).

Datasets were accepted for inclusion in the databank if they met all of the following criteria:

1) Participants were clustered within some higher-level units in the dataset;
2) The dataset had a continuous outcome recorded for each cluster in at least two time points, i.e. two or more periods, or a minimum of baseline and one other time point; and
3) If the dataset was from a randomised trial, the cluster was the level of randomisation.

Given that the CLOUD bank's purpose is to provide a repository of within-cluster correlation parameters corresponding to a range of real-world scenarios, studies that investigated any type of treatment, population or had other study designs consistent with the restrictions above were considered. Observational studies (e.g. pilot studies or routinely collected datasets without the randomisation of clusters to intervention sequences) were also included.

This study was approved by the Monash University Human Research Ethics committee (Ethics approval). Analysis results were accessed in one of three ways: 1) the co-author with access

permissions performed the analysis and shared the required results, 2) the dataset was downloaded from the freely available online repository and analysed by the authors, and 3) data custodians of the potential datasets were sent a request for access and, once approved, the dataset was supplied to the authors who performed the analysis.

Models

We fit three linear mixed effects regression models for each continuous outcome in the CLOUD bank, one for each of the following correlation structures: 1) exchangeable[7]; 2) block-exchangeable[8,9]; and 3) discrete-time decay within-cluster correlation structures[10]. Further details and syntax to fit these models, including extensions for cohort designs, are available in Supplementary 1. Although they are presented as three distinct models, they can each be considered as a specific case of the class of models described in Kasza et al.[10] and Li et al.[19]

Each model includes fixed intervention effects (i.e. where the receipt of the intervention is either 'yes' or 'no'), categorical period effects, and a random effect for cluster. Note that we do not allow for partial intervention effects or for the intervention effect to change over time.

The "exchangeable" correlation structure assumes the correlation between observations within a cluster are the same regardless of the time between their measurement. The correlation is thus described with a single intra-cluster correlation coefficient. The "block-exchangeable" correlation structure (known also as two-period decay and the Hooper/Girling model[10]) extends the exchangeable correlation structure by supposing that the correlation between observations within the same cluster in the same period (WP-ICC, the within-period correlation coefficient), and the correlation between observations in the same cluster but different periods (between-period ICC) differ. We can calculate the cluster autocorrelation coefficient (CAC), given by the ratio of the between-period ICC and the WP-ICC[8,9]. This is the correlation between two

population means from the same cluster at different time points – where a CAC of zero implies independence between observations in different periods and a CAC of 1 implies no decay in correlation as duration between periods increases. These models included an additional random effect for each cluster-period.

The discrete-time decay correlation structure implies that the between-period ICC depends on the length of time between the periods of the observations being compared, with a decay parameter quantifying the relative reduction in correlation with each successive period of time between observations. We refer to this reduction in correlation between adjacent periods as the decaying CAC (Supplementary 1).

Some datasets included separate participants in each period of the study (referred to as "cross-sectional" sampling schemes), while in other studies, participants were measured in multiple periods ("cohort" sampling schemes). The cohort studies can be further divided into "closed" and "open" cohort sampling schemes, where closed cohorts do not allow new participants to enter the study after it has begun, while open cohorts allow participants to provide observations in variable numbers of study periods[20]. For studies with open or closed cohort sampling schemes, each correlation structure is extended to account for the additional within-participant correlation by including a random effect for participant in the underlying model (Supplementary 1).

Fitting the models

Models were fitted using the SAS/STAT® (version 9.4) MIXED procedure. Table S2 presents example code. The key components to each analysis are: the outcome, treatment (if applicable), participant ID (if applicable), period and cluster. We did not adjust for other covariates, allow for treatment effect heterogeneity or treatment-by-period interactions.

Comparison of correlation parameter estimates

The WP-ICC and CAC estimates were extracted from each of the three models. We compared the median and interquartile range (IQR) of the correlation parameter estimates between trial characteristics to identify patterns that may predict parameter values.

The selection of trial characteristics was informed by the hypotheses of Campbell et al.[13] and aspects of trial design considered during sample size calculation[2]. They include: cluster-period size (the average number of participants per cluster-period), period-length (months), outcome type (i.e. clinical outcome vs process outcome, where process outcomes examine elements of the process that ultimately lead to participant outcomes, e.g. physician adherence to a protocol[21]), and cluster type (e.g. primary vs secondary care). Campbell et al. hypothesised that increasing cluster-period size is expected to decrease the WP-ICC. We extended this hypothesis to period-length and its potential impact on both the WP-ICC and CAC. The type of cluster is hypothesised to have an influence over the correlation parameters (e.g. secondary care patient outcomes are likely to have a higher degree of similarity than those of general practitioner patients). The study country's income status (classified as high-income or middle/low income according to the World Bank indicator[22]). This was added to assist trialists in matching the correlation parameters of their planned study to those from previous studies.

For convenience, period-length and cluster-period size were classified as being above or below the characteristic's median among the datasets in the CLOUD bank.

Results

Here, we describe the CLOUD bank studies, the repository of correlation parameter estimates and the relationship between patterns of trial characteristics and the correlation parameters.

Characteristics of CLOUD bank studies

The CLOUD bank comprises 29 datasets and 44 continuous outcomes. Of these, 22 datasets were available for public access via online repositories, Dryad[23-28], LSHTM data compass[29,30], PLoS Medicine[31-39] and UK data service[40]. Seven datasets were included through permissions granted to co-authors[17,35,41-45]. Two datasets contained two sub trials, which were analysed and reported separately (Disinvestment[34], Cashbased[33]).

Details of each study are described in Table S3. Nineteen datasets were cluster randomised trials with a "parallel with baseline" design (i.e. had parallel randomisation of clusters with a baseline observation), 6 were from stepped-wedge trials (see Figure S1 for a schematic) and 4 were from observational studies. Eleven studies were conducted in low/middle income countries. Seven datasets used primary care clusters (e.g. general practice/primary care clinic), 6 secondary care clusters (e.g. hospital ward/intensive care unit), 13 community clusters (e.g. town/village) and 3 school clusters.

Twelve studies recruited participants with a cross-sectional sampling scheme, 16 with a closed cohort sampling scheme, and 1 with an open cohort sampling scheme. Most studies collected their observations at discrete timepoints, however, for nine of the studies the observations could have occurred at any time during the period. Models to account for continuous correlation decay have been developed[46], however, we do not consider them here. The CLOUD bank contains only one study recording a process outcome, eliminating the possibility of a comparison between clinical outcomes and process outcomes. The datasets in the bank had a median of 23 clusters (IQR: 16 – 41), a median of 24 participants per cluster-period (IQR: 15 – 60), a median of 3 periods (IQR: 2 – 6) and a median period-length of 2 months (IQR: 1 – 18 months). We break these characteristics down further by design in Table 1.

Repository of correlation parameters

This section includes the repository of parameter estimates associated with the exchangeable, block-exchangeable and discrete-time decay models (Table 2). For three outcomes from one dataset (Outreach[32]), the algorithm for fitting the discrete-time correlation decay structure did not converge. All other models converged in all datasets.

From the discrete-time decay model, the WP-ICCs ranged from 0.002 to 0.350, median 0.05 (IQR: 0.02 – 0.09), while the CACs ranged from -0.98 to 0.99 (note: negative CACs are possible in the discrete-time decay model but not in the block-exchangeable mixed model), median 0.73 (IQR: 0.19 – 0.91). Boxplots for the estimated WP-ICCs and CACs from the different models are presented in Figure 1. The ICC calculated from fitting the exchangeable correlation structure (median 0.03, IQR: 0.02 – 0.07), and the WP-ICCs from the block-exchangeable (median 0.05, IQR: 0.03 – 0.11) and discrete-time decay correlation structures were similar (Figure 1, Table 2). For most outcomes, the estimated within-period ICCs from the block-exchangeable and discrete-time correlation decay structures were more similar to each other, and generally higher, than the exchangeable correlation within-period ICC (Figure S2).

The discrete-time decay correlation structure resulted in a negative decaying CAC estimate for 8/41 outcomes, ranging from -0.98 to -0.12 (Figure 1). The median of the decaying CAC values provided by the discrete-time correlation decay model (median 0.73, IQR: 0.19 – 0.91) is higher than that of the block-exchangeable correlation structure (median 0.64, IQR: 0.18 – 0.89), however, the IQRs are similar.

[Insert Figure 1.]

Patterns of correlation coefficients for a range of trial characteristics

We compare the distributions of the WP-ICC for the three correlation structures, and compare the CAC and decaying CAC for the block-exchangeable and discrete-time decay correlation structures respectively, for outcomes belonging to datasets with different trial characteristics (Figures 2, 3, S5 and S6). The corresponding median and IQR for these estimated coefficients are reported in Table 2, separated into categories of each trial characteristic.

Within-period ICC, WP-ICC. Outcomes from larger studies (larger number of participants in each cluster-period) tended to have slightly lower median WP-ICCs compared to those from smaller studies (Figure 2, Table 2), while studies with longer periods had lower median WP-ICCs compared to those with shorter periods (Figure S5). Studies with primary care clusters had lower median WP-ICCs than studies with secondary care clusters (Figure 3), while studies with clusters of communities had higher median WP-ICCs than studies with clusters of schools. The median WP-ICCs from high income country studies were comparable to those from low/middle income countries (Figure S6).

Cluster autocorrelation, CAC. Estimated CACs were similar for block-exchangeable and discrete-time decay correlation structures for most characteristics. Estimated CACs were similar between short versus long period-lengths (Figure S5), but studies with fewer participants per cluster-period tended to have smaller CACs (Figure 2). Studies with primary care and secondary care clusters had comparable median CACs (Figure 3). Studies with community clusters had higher median CACs than studies with school clusters when estimated with the block-exchangeable model, though this was reversed when estimated with the discrete-time decay model (Figure 3). Studies from high income countries tended to have a higher median CAC (Figure S6).

[Insert Figure 2.]

[Insert Figure 3.]

RShiny app

All analysis results are provided in our online RShiny app[18]. Users can subset the table, or search directly, for particular trials of interest https://monash-biostat.shinyapps.io/CLOUDbank/.

Example of discrete-time decay model sample size calculation

In the absence of prior studies or available routinely collected data, trialists will want to select a likely range of correlation parameter estimates for a primary trial outcome that is fairly typical for studies with similar cluster-period sizes and period-lengths. We present an example of how the CLOUD bank repository of ICCs and CACs may be used, to estimate the required number of clusters for a stepped-wedge trial with seven sequences, eight 6-month periods and 20 different participants per cluster-period, aiming to detect a standardised effect size of 0.2 for a continuous outcome with 80% power and a two-sided significance level of 0.05. To choose a plausible WP-ICC and CAC for the sample size calculation, we assume a discrete-time decay model for the outcome, and filter the CLOUD bank by cluster-period size and period-length in the RShiny app. In the CLOUD bank repository, we find that studies in the same characteristic categories (above/below the median) for cluster-period size and period-length as our planned trial had a median (IQR) for the WP-ICC of 0.053 (0.031 – 0.113) and decaying CAC of 0.258 (0.00 – 0.689)

By entering the sample size parameters including plausible range of correlation coefficients into the RShiny app developed by Hemming et al.[2] (Supplementary 4), researchers are able to compare the required number of clusters and thus, total sample size, across a range of scenario, for example, at each combination of the 25th percentile, median, 75th percentile WP-ICC and decaying CAC values (Table 3). In general, we would recommend that researchers consider a range of ICCs in their sample size calculations to assess the sensitivity to their chosen value.

Discussion

This study provides the first repository of correlation parameter estimates to inform the design of longitudinal cluster randomised trials, where clusters are followed over multiple treatment periods, for the discrete-time decay correlation structure for continuous outcomes and expands upon the existing repositories for the exchangeable and block-exchangeable correlation structures.

We have reported and compared the correlation coefficients for three different correlation structures for 29 studies with a variety of study characteristics, to extend the 'rules of thumb' available when informing the correlation parameter estimates that can be explored during sample size calculations. We have shown that estimates of the within-period ICC vary little between the three correlation structures, and that, overall, the CAC for the discrete-time decay correlation structure is larger than the CAC from the block-exchangeable correlation structure – as expected given the discrete-time decay model measures the decay per period, whereas the block-exchangeable model measures the "average" decay over the entire period. When each correlation structure is considered separately, correlation parameter estimates may change for different trial design characteristics. We provide 'rules of thumb' for characteristics such as cluster-period size and period-length. As expected, WP-ICCs for larger cluster-period sizes and longer period-lengths were lower than for smaller cluster-period sizes and shorter period-lengths, although this difference was not substantial. Unsurprisingly, the WP-ICC was found to be smaller for studies with primary care clusters compared to studies with secondary care clusters, and for studies with school clusters compared to community clusters. While cluster type and other characteristics such as country income status have been summarised, further investigation with a larger number of studies is required to disentangle their effects from other characteristics explored here, as well as characteristics unavailable in this study.

Research in context

Similar to the observations of Campbell et al.[13], while larger clusters are expected to have smaller ICCs[12,47,48], we found only a slight tendency towards smaller WP-ICC estimates for studies with larger (compared to smaller) cluster-period sizes. Further, the secondary care outcomes in the CLOUD bank had higher WP-ICCs than primary care outcomes, supporting the observations made by Campbell et al.[13]. In the absence of available estimates, previous studies have recommended a CAC of between 0.8 – 1 for the block-exchangeable model[7,49]. However, similar to previous research[17] the block-exchangeable CAC estimates found here (median: 0.64, IQR: 0.19 – 0.89) indicate that CACs may be lower than the values that have previously been assumed - suggesting that more conservative values may be appropriate in sample size calculations during trial design, and highlights the necessity for trial results to include estimates of correlation parameters to inform future trial planning.

Surprisingly, we observed several (8/41) negative CAC values estimated by the discrete-time decay model (indicating between-period correlations oscillate between positive and negative values as distance between periods increases). However, such negative values are not regarded as plausible, and further empirical evidence is needed to ascertain if this counterintuitive correlation pattern is replicable or whether it is simply due to noise around a true small CAC.

Limitations

We have explored the relationships between trial characteristics and their correlation parameter estimates, however, our findings are based on a limited number of available datasets. We hope that additional databanks of correlation parameters will be made available which can help

strengthen future analyses of determinants of correlation estimates in longitudinal cluster randomised trials. Further, while we provide ranges of correlation parameter estimates from multiple studies, we have not provided confidence intervals around the correlation parameter estimates as there are, as yet, no agreed upon methods for constructing confidence intervals for ICC and CAC estimates in longitudinal cluster trials.

The analyses presented here assumed treatment effects that do not vary across clusters, periods, or time. Further, we have assumed models with fixed time effects that do not vary across cluster. When there are cluster-specific time trends, treatment effect heterogeneity or treatment-by-period interactions, failing to include these in the model may impact the correlation parameter estimates. However, the impact will likely depend on several factors, including the strength of the heterogeneity, the number of time periods in the study, and the number of clusters. Additionally, where studies had a cohort sampling scheme, the participant-level correlation was assumed to be constant over time, thus any participant-level decay would be attributed to the cluster. The number of clusters in most of the included studies was modest, with a median of 23 clusters. We did not consider small-sample adjustments as point estimates of correlation parameters have been shown to be unbiased with few clusters and we do not estimate uncertainty around their estimates.

We have explored the decay in correlation of clustered continuous outcomes within discrete periods of time. While the majority of the studies in the CLOUD bank are ‘survey-like’ in their measurement of outcomes, meaning all participants are observed at the same time point within a period, there are nine datasets where observations could occur at any point in time within a period interval (i.e. in continuous time). When this is the case, a correlation that decays continuously over time may be more appropriate[46,50]. Additionally, there may be study designs where it may be more appropriate to model time continuously or with a more complex seasonal form.

Correlation coefficients may be influenced by several other characteristics, one of which is whether or not a process outcome in considered[13]. The CLOUD bank contained one process outcome, and thus, no ICCs and CACs were compared for this characteristic. Binary and count outcomes available in the CLOUD bank were not considered here, as it has yet to be determined whether the correlation coefficients on the natural scale, obtained from fitting the discrete-time decay correlation structure above to binary outcomes, are appropriate. Some methods for ICCs for binary outcomes on the natural scale have recently been introduced in the context of marginal models/GEE in Li et al.[19,51]

Finally, we have not explored procedures for selecting which correlation structure is most appropriate. While trialists should pre-specify their primary analysis model, including the choice of within-cluster correlation structure, exploratory analyses may focus on finding the most appropriate within-cluster correlation structure. AICs and BICs are often used to guide model choice, but whether these are appropriate for the selection of within-cluster correlation structures is unknown. We note that for these datasets, AICs and BICs were similar for all three models, and present their values in the RShiny app. We are planning further research in this area.

Conclusions

The correlation coefficients provided by this study can be explored in the RShiny app by trialists when designing their trial. The design characteristics of the studies in the CLOUD bank may be compared with those of future trials to inform the selection of correlation coefficients for use in sample size and power calculations. The CLOUD bank will continue to seek new datasets for inclusion. Readers with access and appropriate permissions to such datasets are encouraged to get in touch with the first author.

Ethics approval

This study and analyses have been granted approval by Monash University Human Research Ethics Committee (Project ID 19998).

Data availability

All results are available through the RShiny app https://monash-biostat.shinyapps.io/CLOUDbank/.

Funding

EK is supported through an Australian Government Research Training Program (RTP) Scholarship administered by Monash University, Australia

The authors disclosed receipt of the following financial support for the research, authorship and publication of this article: This work was supported by the Monash University Network of Excellence grant. 2018-2019. NOE180009.

JT is jointly funded by the UK Medical Research Council (MRC) and the UK Department for International Development (DFID) under the MRC/DFID Concordat agreement and is part of the EDCTP2 programme supported by the European Union. Grant Ref: MR/R010161/1

Acknowledgements

We thank the researchers who provided datasets and analysis presented here, as well as those who contributed binary outcome results which have not been published here.

We thank the members of the International Network for Innovative Cluster Trial Designs for their contributions towards sourcing available datasets, performing analyses and providing correlation coefficients (https://clustertrials-research.netlify.app/).

Declaration of Conflicting Interests

The Authors declare that there is no conflict of interest

References

1. Eldridge S and Kerry S. *A practical guide to cluster randomised trials in health services research*. John Wiley & Sons, 2012.
2. Hemming K, Kasza J, Hooper R, et al. A tutorial on sample size calculation for multiple-period cluster randomized parallel, cross-over and stepped-wedge trials using the Shiny CRT Calculator. *Int J Epidemiol* 2020 2020/02/23. DOI: 10.1093/ije/dyz237.
3. Hemming K, Haines TP, Chilton PJ, et al. The stepped wedge cluster randomised trial: rationale, design, analysis, and reporting. *BMJ* 2015; 350: h391. 2015/02/11. DOI: 10.1136/bmj.h391.
4. 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. 2019/09/03. DOI: 10.1002/sim.8349.
5. Arnup SJ, Forbes AB, Kahan BC, et al. Appropriate statistical methods were infrequently used in cluster-randomized crossover trials. *J Clin Epidemiol* 2016; 74: 40-50. 2015/12/04. DOI: 10.1016/j.jclinepi.2015.11.013.
6. Teerenstra S, Eldridge S, Graff M, et al. A simple sample size formula for analysis of covariance in cluster randomized trials. *Stat Med* 2012; 31: 2169-2178. 2012/04/13. DOI: 10.1002/sim.5352.
7. Hussey MA and Hughes JP. Design and analysis of stepped wedge cluster randomized trials. *Contemp Clin Trials* 2007; 28: 182-191. 2006/07/11. DOI: 10.1016/j.cct.2006.05.007.
8. 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. 2016/06/29. DOI: 10.1002/sim.7028.
9. 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. 2016/01/11. DOI: 10.1002/sim.6850.
10. 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* 2017: 962280217734981. 2017/10/14. DOI: 10.1177/0962280217734981.
11. Feldman HA and McKinlay SM. Cohort versus cross-sectional design in large field trials: precision, sample size, and a unifying model. *Stat Med* 1994; 13: 61-78. 1994/01/15. DOI: 10.1002/sim.4780130108.
12. Adams G, Gulliford MC, Ukoumunne OC, et al. Patterns of intra-cluster correlation from primary care research to inform study design and analysis. *J Clin Epidemiol* 2004; 57: 785-794. 2004/10/16. DOI: 10.1016/j.jclinepi.2003.12.013.
13. Campbell MK, Fayers PM and Grimshaw JM. Determinants of the intracluster correlation coefficient in cluster randomized trials: the case of implementation research. *Clin Trials* 2005; 2: 99-107. 2005/11/11. DOI: 10.1191/1740774505cn071oa.
14. 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. 2014/04/24. DOI: 10.1186/1471-2288-14-54.
15. Pagel C, Prost A, Lewycka S, et al. Intracluster correlation coefficients and coefficients of variation for perinatal outcomes from five cluster-randomised controlled trials in low and middle-income countries: results and methodological implications. *Trials* 2011; 12: 151. 2011/06/16. DOI: 10.1186/1745-6215-12-151.
16. Thompson DM, Fernald DH and Mold JW. Intraclass correlation coefficients typical of cluster-randomized studies: estimates from the Robert Wood Johnson Prescription for Health projects. *Ann Fam Med* 2012; 10: 235-240. 2012/05/16. DOI: 10.1370/afm.1347.

17. 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. 2016/08/16. DOI: 10.1186/s13063-016-1532-9.
18. Chang W, Cheng J, Allaire J, et al. Shiny: web application framework for R. R package version 1.1. 0; 2018. 2019.
19. 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* 2020: 962280220932962. 2020/07/08. DOI: 10.1177/0962280220932962.
20. Kasza J, Hooper R, Copas A, et al. Sample size and power calculations for open cohort longitudinal cluster randomized trials. *Stat Med* 2020 2020/03/07. DOI: 10.1002/sim.8519.
21. Mason J, Wood J and Freemantle N. Designing evaluations of interventions to change professional practice. *J Health Serv Res Policy* 1999; 4: 106-111. 1999/07/01. DOI: 10.1177/135581969900400209.
22. Bank TW. World Bank Country and Lending Groups, https://datahelpdesk.worldbank.org/knowledgebase/articles/906519-world-bank-country-and-lending-groups (accessed 23 Sep 2020).
23. Ballard C, Corbett A, Orrell M, et al. Impact of person-centred care training and person-centred activities on quality of life, agitation, and antipsychotic use in people with dementia living in nursing homes: A cluster-randomised controlled trial. *PLoS Med* 2018; 15: e1002500. 2018/02/07. DOI: 10.1371/journal.pmed.1002500.
24. Blair C and Raver CC. Closing the achievement gap through modification of neurocognitive and neuroendocrine function: results from a cluster randomized controlled trial of an innovative approach to the education of children in kindergarten. *PLoS One* 2014; 9: e112393. 2014/11/13. DOI: 10.1371/journal.pone.0112393.
25. Llaurado E, Tarro L, Morina D, et al. EdAl-2 (Educacio en Alimentacio) programme: reproducibility of a cluster randomised, interventional, primary-school-based study to induce healthier lifestyle activities in children. *BMJ Open* 2014; 4: e005496. 2014/11/22. DOI: 10.1136/bmjopen-2014-005496.
26. Ogedegbe G, Plange-Rhule J, Gyamfi J, et al. Health insurance coverage with or without a nurse-led task shifting strategy for hypertension control: A pragmatic cluster randomized trial in Ghana. *Plos Medicine* 2018; 15. DOI: ARTN e1002561 10.1371/journal.pmed.1002561.
27. Rockers PC, Zanolini A, Banda B, et al. Two-year impact of community-based health screening and parenting groups on child development in Zambia: Follow-up to a cluster-randomized controlled trial. *Plos Medicine* 2018; 15. DOI: ARTN e1002555 10.1371/journal.pmed.1002555.
28. Tannenbaum C, Agnew R, Benedetti A, et al. Effectiveness of continence promotion for older women via community organisations: a cluster randomised trial. *Bmj Open* 2013; 3. DOI: ARTN e004135 10.1136/bmjopen-2013-004135.
29. Chibanda D, Weiss HA, Verhey R, et al. Effect of a Primary Care-Based Psychological Intervention on Symptoms of Common Mental Disorders in Zimbabwe: A Randomized Clinical Trial. *JAMA* 2016; 316: 2618-2626. 2016/12/28. DOI: 10.1001/jama.2016.19102.
30. Cresswell JA, Ganaba R, Sarrassat S, et al. The effect of the Alive & Thrive initiative on exclusive breastfeeding in rural Burkina Faso: a repeated cross-sectional cluster randomised controlled trial. *Lancet Glob Health* 2019; 7: e357-e365. 2019/02/21. DOI: 10.1016/S2214-109X(18)30494-7.
31. Andrew A, Attanasio O, Fitzsimons E, et al. Impacts 2 years after a scalable early childhood development intervention to increase psychosocial stimulation in the home: A follow-up of a cluster randomised controlled trial in Colombia. *PLoS Med* 2018; 15: e1002556. 2018/04/25. DOI: 10.1371/journal.pmed.1002556.
32. Fairall LR, Folb N, Timmerman V, et al. Educational Outreach with an Integrated Clinical Tool for Nurse-Led Non-communicable Chronic Disease Management in Primary

Care in South Africa: A Pragmatic Cluster Randomised Controlled Trial. *PLoS Med* 2016; 13: e1002178. 2016/11/23. DOI: 10.1371/journal.pmed.1002178.
33. Grijalva-Eternod CS, Jelle M, Haghparast-Bidgoli H, et al. A cash-based intervention and the risk of acute malnutrition in children aged 6-59 months living in internally displaced persons camps in Mogadishu, Somalia: A non-randomised cluster trial. *PLoS Med* 2018; 15: e1002684. 2018/10/30. DOI: 10.1371/journal.pmed.1002684.
34. Haines TP, Bowles KA, Mitchell D, et al. Impact of disinvestment from weekend allied health services across acute medical and surgical wards: 2 stepped-wedge cluster randomised controlled trials. *PLoS Med* 2017; 14: e1002412. 2017/11/01. DOI: 10.1371/journal.pmed.1002412.
35. Lombard C, Harrison C, Kozica S, et al. Preventing Weight Gain in Women in Rural Communities: A Cluster Randomised Controlled Trial. *PLoS Med* 2016; 13: e1001941. 2016/01/20. DOI: 10.1371/journal.pmed.1001941.
36. Menon P, Nguyen PH, Saha KK, et al. Impacts on Breastfeeding Practices of At-Scale Strategies That Combine Intensive Interpersonal Counseling, Mass Media, and Community Mobilization: Results of Cluster-Randomized Program Evaluations in Bangladesh and Viet Nam. *PLoS Med* 2016; 13: e1002159. 2016/10/26. DOI: 10.1371/journal.pmed.1002159.
37. Merom D, Mathieu E, Cerin E, et al. Social Dancing and Incidence of Falls in Older Adults: A Cluster Randomised Controlled Trial. *PLoS Med* 2016; 13: e1002112. 2016/08/31. DOI: 10.1371/journal.pmed.1002112.
38. Nickless A, Voysey M, Geddes J, et al. Mixed effects approach to the analysis of the stepped wedge cluster randomised trial-Investigating the confounding effect of time through simulation. *Plos One* 2018; 13. DOI: ARTN e0208876 10.1371/journal.pone.0208876.
39. Thankappan KR, Sathish T, Tapp RJ, et al. A peer-support lifestyle intervention for preventing type 2 diabetes in India: A cluster-randomized controlled trial of the Kerala Diabetes Prevention Program. *PLoS Med* 2018; 15: e1002575. 2018/06/07. DOI: 10.1371/journal.pmed.1002575.
40. Perra O. Cluster randomised trial of the Ready to Learn programme for children starting school 2010-2013. Colchester, Essex: UK Data Service, 2019.
41. Thiruganasambandamoorthy V, Kwong K, Wells GA, et al. Development of the Canadian Syncope Risk Score to predict serious adverse events after emergency department assessment of syncope. *CMAJ* 2016; 188: E289-E298. 2016/07/06. DOI: 10.1503/cmaj.151469.
42. Liddy C, Hogg W, Singh J, et al. A real-world stepped wedge cluster randomized trial of practice facilitation to improve cardiovascular care. *Implement Sci* 2015; 10: 150. 2015/10/30. DOI: 10.1186/s13012-015-0341-y.
43. Stow PJ, Hart GK, Higlett T, et al. Development and implementation of a high-quality clinical database: the Australian and New Zealand intensive care society adult patient database. *J Crit Care* 2006; 21: 133-141. DOI: 10.1016/j.jcrc.2005.11.010.
44. Tirlea L, Truby H and Haines TP. Pragmatic, Randomized Controlled Trials of the Girls on the Go! Program to Improve Self-Esteem in Girls. *Am J Health Promot* 2016; 30: 231-241. DOI: 10.1177/0890117116639572.
45. Turner J, Kelly B, Clarke D, et al. A tiered multidisciplinary approach to the psychosocial care of adult cancer patients integrated into routine care: the PROMPT study (a cluster-randomised controlled trial). *Support Care Cancer* 2017; 25: 17-26. 2016/08/18. DOI: 10.1007/s00520-016-3382-0.
46. 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. 2019/01/22. DOI: 10.1002/sim.8089.
47. Donner A. An empirical study of cluster randomization. *Int J Epidemiol* 1982; 11: 283-286. 1982/09/01. DOI: 10.1093/ije/11.3.283.
48. Gulliford MC, Ukoumunne OC and Chinn S. Components of variance and intraclass correlations for the design of community-based surveys and intervention studies: data from

the Health Survey for England 1994. *Am J Epidemiol* 1999; 149: 876-883. 1999/04/30. DOI: 10.1093/oxfordjournals.aje.a009904.
49. Hooper R and Bourke L. Cluster randomised trials with repeated cross sections: alternatives to parallel group designs. *BMJ* 2015; 350: h2925. 2015/06/10. DOI: 10.1136/bmj.h2925.
50. Hooper R and Copas A. Stepped wedge trials with continuous recruitment require new ways of thinking. *J Clin Epidemiol* 2019; 116: 161-166. 2019/07/06. DOI: 10.1016/j.jclinepi.2019.05.037.
51. Li F, Forbes AB, Turner EL, et al. Power and sample size requirements for GEE analyses of cluster randomized crossover trials. *Stat Med* 2019; 38: 636-649. 2018/10/10. DOI: 10.1002/sim.7995.

Table 1. Summary of CLOUD bank trial characteristics by type of design – median (IQR) reported

Design characteristics	Parallel with baseline (n = 19)	Stepped-wedge (n = 6)	Observational (n = 4)
No. Periods	2 (2-3)	7.5 (5-9)	6.5 (5-10.5)
Period-length (months) [a]	12 (6-18)	2 (1-6)	2.5 (1.5-7.5)
No. Clusters	30 (20-41)	11.5 (6-12)	74 (14-278)
Cluster-period size	23.7 (12.6-36.5)	43 (20-133)	89.5 (34-146)

[a] average period-length is reported for studies where period-lengths differ.

Table 2. Median (IQR) for the WP-ICC of each model, the CAC coefficient for the block-exchangeable model and the decaying CAC for the discrete time decay correlation structure in different trial characteristics

Characteristic	Comparison	No. outcomes n = 44	WP-ICC			CAC	
			Ex	BE	DTD [b]	BE	DTD [b]
Cluster-period size [a]	Small (<24)	22	0.03 (0.01-0.07)	0.06 (0.04-0.12)	0.05 (0.03-0.09)	0.32 (0.08-0.68)	0.40 (0.00-0.75)
	Large (>=24)	22	0.03 (0.02-0.07)	0.04 (0.02-0.09)	0.05 (0.02-0.10)	0.85 (0.61-0.91)	0.90 (0.25-0.96)
Period-length (months)	Short (<12)	17	0.05 (0.01-0.10)	0.09 (0.05-0.16)	0.08 (0.04-0.11)	0.49 (0.19-0.91)	0.80 (0.25-0.96)
	Long (>=12)	27	0.03 (0.02-0.06)	0.04 (0.02-0.08)	0.05 (0.02-0.08)	0.68 (0.15-0.88)	0.67 (-0.03-0.90)
Cluster type	Primary care	13	0.03 (0.02-0.04)	0.04 (0.02-0.05)	0.04 (0.02-0.07)	0.84 (0.55-0.88)	0.88 (0.78-0.91)
	Secondary care	7	0.05 (0.01-0.08)	0.08 (0.05-0.16)	0.05 (0.03-0.08)	0.86 (0.08-0.95)	0.95 (0.25-0.97)
	Community	19	0.04 (0.02-0.07)	0.07 (0.04-0.11)	0.07 (0.04-0.11)	0.42 (0.14-0.75)	0.27 (-0.12-0.80)
	School	7	0.01 (0.00-0.10)	0.02 (0.01-0.11)	0.02 (0.02-0.11)	0.22 (0.12-0.80)	0.69 (-0.16-0.80)
Country [c]	High-income countries	24	0.04 (0.01-0.09)	0.06 (0.02-0.11)	0.05 (0.02-0.10)	0.82 (0.32-0.91)	0.84 (0.34-0.96)
	Low/Middle income countries	20	0.03 (0.02-0.06)	0.05 (0.03-0.08)	0.05 (0.04-0.09)	0.46 (0.15-0.73)	0.27 (0.07-0.78)

[a] Cluster-period size is defined as the average number of participants within a cluster in each period over the entire trial.
[b] The number of outcomes varies for the DTD correlation structure as three outcomes did not converge.
[c] Datasets have been classified based on the country of their outcomes according to the World Bank indicator of income per capita (i.e. high-income countries USD$12,536 or more low/middle income countries USD$12,535 or less).
WP-ICC: within-period intra-cluster correlation, CAC: cluster autocorrelation coefficient, Ex: Exchangeable, BE: Block-exchangeable, DTD: Discrete-time correlation decay

Table 3. Total number of clusters (total sample size) required to detect an effect size of 0.2 with 80% power and a two-sided significance level of 0.05 with an eight-period (seven sequence) stepped-wedge design with 20 participants per cluster-period, assuming the decaying within-cluster correlation structure and varying choices of WP-ICC and CAC from the CLOUD repository.

	Decaying CAC		
WP-ICC	**0.00**	**0.258**	**0.689**
0.031	14 (2240)	21 (3360)	21 (3360)
0.053	21 (3360)	21 (3360)	28 (4480)
0.113	28 (4480)	35 (5600)	35 (5600)

WP-ICC: within-period intra-cluster correlation, CAC: cluster autocorrelation coefficient

Figure legends

Figure 1. WP-ICC and CAC coefficients for the exchangeable, block-exchangeable and discrete-time correlation decay models (left). The CAC coefficients for the block-exchangeable model, and Decaying CAC for the discrete-time correlation decay model (right). The exchangeable and block-exchangeable plots include 44 outcomes. Discrete-time decay models converged for 41 outcomes.

Figure 2. Patterns of correlation coefficients for the large and small cluster-period sizes (WP-ICC: top row; CAC: bottom row; for the exchangeable, block-exchangeable and discrete-time decay correlation structures (from left to right)). The exchangeable and block-exchangeable plots include 44 outcomes. Discrete-time decay models converged for 41 outcomes. Of the 44 outcomes, 22 had 24 or more participants per cluster; 22 had fewer. Of the 22 outcomes with 24 or more participants per cluster, discrete-time decay models converged for 19 outcomes.

Figure 3. Patterns of correlation coefficients for four categories of cluster types: primary care, secondary care, community and school. (WP-ICC: top row; CAC: bottom row; for the exchangeable, block-exchangeable and discrete-time decay correlation structures (from left to right)). The exchangeable and block-exchangeable plots include 44 outcomes. Discrete-time decay models converged for 41 outcomes. Of the 44 outcomes, 13 outcomes included primary care clusters; 7 included secondary care clusters; 19 outcomes included community clusters; and 5 outcomes included school clusters. Of the 13 primary care outcomes, discrete-time decay models converged for 10 outcomes.