



Minerva Access is the Institutional Repository of The University of Melbourne

**Author/s:**

Yelland, LN;Sullivan, TR;Price, DJ;Lee, KJ

**Title:**

Sample size calculations for randomised trials including both independent and paired data

**Date:**

2017-04-15

**Citation:**

Yelland, L. N., Sullivan, T. R., Price, D. J. & Lee, K. J. (2017). Sample size calculations for randomised trials including both independent and paired data. *Statistics in Medicine*, 36 (8), pp.1227-1239. <https://doi.org/10.1002/sim.7201>.

**Persistent Link:**

<https://hdl.handle.net/11343/292317>

## Sample Size Calculations for Randomised Trials Including Both Independent and Paired Data

Lisa N Yelland<sup>a,b</sup>, Thomas R Sullivan<sup>a</sup>, David J Price<sup>c</sup>, Katherine J Lee<sup>d,e</sup>

<sup>a</sup> School of Public Health, The University of Adelaide, South Australia, Australia

<sup>b</sup> South Australian Health and Medical Research Institute, South Australia, Australia

<sup>c</sup> School of Mathematical Sciences, The University of Adelaide, South Australia, Australia

<sup>d</sup> Murdoch Children's Research Institute, Victoria, Australia

<sup>e</sup> Department of Paediatrics, University of Melbourne, Victoria, Australia

Short title: Sample Size Calculations for Trials with Independent and Paired Data

Keywords: sample size; clustered data; generalised estimating equations; design effect

Author for correspondence: Dr Lisa Yelland, School of Public Health, Mail Drop DX650 511, The University of Adelaide, SA 5005, Australia, Email: [lisa.yelland@adelaide.edu.au](mailto:lisa.yelland@adelaide.edu.au), Phone: +61 8 8313 3215, Fax: +61 8 8223 4075.

Author Manuscript

This is the author manuscript accepted for publication and has undergone full peer review but has not been through the copyediting, typesetting, pagination and proofreading process, which may lead to differences between this version and the [Version of Record](#). Please cite this article as doi: [10.1002/sim.7201](https://doi.org/10.1002/sim.7201)

## SUMMARY

Randomised trials including a mixture of independent and paired data arise in many areas of health research, yet methods for determining the sample size for such trials are lacking. We derive design effects algebraically assuming clustering due to paired data will be taken into account in the analysis using generalised estimating equations with either an independence or exchangeable working correlation structure. Continuous and binary outcomes are considered, along with three different methods of randomisation: cluster randomisation, individual randomisation and randomisation to opposite treatment groups. The design effect is shown to depend on the intraclass correlation coefficient, proportion of observations belonging to a pair, working correlation structure, type of outcome and method of randomisation. The derived design effects are validated through simulation and example calculations are presented to illustrate their use in sample size planning. These design effects will enable appropriate sample size calculations to be performed for future randomised trials including both independent and paired data.

Author Manuscript

## 1. INTRODUCTION

Mixtures of independent and paired data arise in many areas of health research. For example, in neonatology, outcomes may be collected on infants from single or twin births; in ophthalmology, one or both eyes may be affected by the condition of interest and included in the study; and in orthopedics, patients may require surgery on one or both knees. If any paired data are to be collected in a randomised trial, the clustering present due to the correlation between outcomes of members belonging to the same pair should be taken into account in the analysis to produce appropriate standard errors [1], and in the sample size calculations to avoid conducting an under- or over-powered trial [2]. Clustering is typically handled in sample size calculations by determining the sample size required assuming all observations are independent and then multiplying by an appropriate design effect (DEFF).

For randomised trials involving paired data only, methods for determining the sample size that account for the clustering are available and depend on the method of randomisation. If cluster randomisation is used, where members of the same pair are assigned to the same treatment group, the sample size calculated assuming independence is multiplied by a DEFF of  $1 + \rho$ , where  $\rho$  is the intracluster correlation coefficient (ICC) [3]. If one member of each pair is randomly assigned a treatment and the other member is automatically assigned the opposite treatment, the sample size can be calculated based on a paired t-test for continuous outcomes or McNemar's test for binary outcomes [4, 5].

Determining the sample size for randomised trials including both independent and paired data is more challenging. For cluster randomisation, the DEFF used to handle varying cluster sizes of  $1 + (\tilde{m} - 1)\rho$  could be applied, where  $\tilde{m} = \sum_i m_i^2 / \sum_i m_i$  and  $m_i$  is the size of the  $i$ th cluster [6]. Replacing  $\tilde{m}$  by the average cluster size to give an approximate DEFF [3], or calculating the DEFF using the coefficient of variation of cluster size [7], have also been suggested when individual cluster sizes are unknown. It is unclear whether these DEFFs hold when some of the data are independent and hence  $m_i = 1$  for some clusters. A more general equation for the DEFF that can accommodate varying cluster sizes and different methods of randomisation has been derived from a two-way analysis of variance for continuous outcomes [2]. However, this DEFF may not extend to settings involving alternative methods of analysis, binary outcomes or mixtures of independent and paired data. More flexible sample size approaches have been proposed assuming the clustering will be taken into account in the analysis using generalised estimating equations (GEEs) [8-11], but these involve complex equations that are not readily applicable to trials involving a mixture of independent and paired data.

The aim of this article is to derive DEFFs that can be used to determine the sample size required for two-arm, parallel group, randomised controlled trials that include both independent and paired data in both treatment groups. As this work was motivated by the need to account for mixtures of singletons and twins in sample size calculations for neonatal trials, we focus on this setting as an example, although the methods are applicable to other situations where combinations of independent and paired data arise. We consider both continuous and binary outcomes, as well as scenarios where members of the same pair are randomised to the same treatment group (cluster randomisation), independently of the

treatment assigned to the other member of the pair (individual randomisation), or to opposite treatment groups, since all three methods of randomisation are commonly used in practice [12-14]. DEFFs are derived assuming the marginal or population-average treatment effect is of interest and the analysis will be performed using GEEs [15], as this is the most common analysis approach used to account for twins in neonatal trials [12, 14] and GEEs have been shown to perform well in this setting [16]. Both the independence and exchangeable working correlation structures are considered, since the exchangeable structure is more realistic and often more efficient than the independence structure [17, 18] but makes stronger assumptions about the data [19] and may not be appropriate when the outcome of interest is related to the size of the cluster [20, 21].

The remainder of this article is arranged as follows. DEFFs are derived algebraically in Section 2 and their validity is assessed through simulation in Section 3. Example calculations are presented in Section 4 to illustrate how the derived DEFFs can be used in sample size calculations for future randomised trials involving both independent and paired data. We conclude with a discussion and recommendations in Section 5.

## 2. DESIGN EFFECTS

### 2.1 Notation and Setting

Let  $Y_{ij}$  be a continuous or binary outcome measured on the  $j$ th member of the  $i$ th cluster ( $i = 1, \dots, M; j = 1, \dots, n_i; \sum_{i=1}^M n_i = N$ ). Suppose an analysis will be performed using GEEs to estimate the effect of the randomised treatment group  $X_{ij}$  (1=intervention, 0=control) on the outcome based on the unadjusted mean model:

$$g(\mu_{ij}) = \beta_0 + \beta_1 X_{ij} = \mathbf{X}_{ij}^T \boldsymbol{\beta}, \quad (1)$$

where  $g$  is the link function,  $\mu_{ij} = E[Y_{ij}]$ ,  $\mathbf{X}_{ij} = (1, X_{ij})^T$ , and  $\boldsymbol{\beta} = (\beta_0, \beta_1)^T$  is a vector of model parameters. The GEE estimates of these model parameters ( $\hat{\boldsymbol{\beta}}$ ) are obtained by solving the estimating equations:  $\sum_{i=1}^M \mathbf{D}_i^T \mathbf{V}_i^{-1} (\mathbf{Y}_i - \boldsymbol{\mu}_i) = \mathbf{0}$ , where  $\mathbf{Y}_i = (Y_{i1}, \dots, Y_{in_i})^T$ ,  $\boldsymbol{\mu}_i = E[\mathbf{Y}_i]$ ,  $\mathbf{V}_i = \mathbf{A}_i^{1/2} \mathbf{R}_i \mathbf{A}_i^{1/2}$ ,  $\mathbf{A}_i = \text{diag}(\text{var}(Y_{i1}), \dots, \text{var}(Y_{in_i}))$ ,  $\mathbf{R}_i$  is a working correlation matrix for  $\mathbf{Y}_i$ , and  $\mathbf{D}_i = \partial \boldsymbol{\mu}_i / \partial \boldsymbol{\beta}$ . When  $g$  is the identity link,  $\mathbf{D}_i = \mathbf{X}_i = (\mathbf{X}_{i1}, \dots, \mathbf{X}_{in_i})^T$ . The covariance matrix of the GEE estimates of the model parameters is given by:

$$\text{cov}(\hat{\boldsymbol{\beta}}) = \left[ \sum_{i=1}^M \mathbf{D}_i^T \mathbf{V}_i^{-1} \mathbf{D}_i \right]^{-1} \left[ \sum_{i=1}^M \mathbf{D}_i^T \mathbf{V}_i^{-1} \text{cov}(\mathbf{Y}_i) \mathbf{V}_i^{-1} \mathbf{D}_i \right] \left[ \sum_{i=1}^M \mathbf{D}_i^T \mathbf{V}_i^{-1} \mathbf{D}_i \right]^{-1}, \quad (2)$$

where  $\text{cov}(\mathbf{Y}_i) = \mathbf{A}_i^{1/2} \mathbf{C}_i \mathbf{A}_i^{1/2}$  and  $\mathbf{C}_i$  is the true correlation matrix for  $\mathbf{Y}_i$ . Equation (2) simplifies to  $\text{cov}(\hat{\boldsymbol{\beta}}) = \left[ \sum_{i=1}^M \mathbf{D}_i^T \mathbf{V}_i^{-1} \mathbf{D}_i \right]^{-1}$  if the working correlation matrix is correct, i.e. if  $\mathbf{R}_i = \mathbf{C}_i$  [15, 22]. The quantities of primary interest are the marginal treatment effect estimate,  $\hat{\beta}_1$ , and its variance,  $\text{var}(\hat{\beta}_1)$ .

Let each cluster consist of a single member or a pair of members, such that  $n_i = 1$  or  $2$ , and let the number of single clusters and paired clusters be denoted by  $M_S$  and  $M_P$ , respectively. Without loss of generality, we can assume that the true correlation matrix is exchangeable, i.e.  $\mathbf{C}_i = 1$  when  $n_i = 1$  and  $\mathbf{C}_i = \begin{bmatrix} 1 & \rho \\ \rho & 1 \end{bmatrix}$  when  $n_i = 2$ . The working correlation matrix is  $\mathbf{R}_i = \mathbf{C}_i$  when  $n_i = 1$ , and  $\mathbf{R}_i = \mathbf{I}$  or  $\mathbf{R}_i = \mathbf{C}_i$  when  $n_i = 2$  and an independence or exchangeable working correlation structure is specified, respectively. Clusters of size 2 may be intervention only (i.e. both cluster members assigned to the intervention group), control only or a mixture of intervention and control, where the number of clusters of each type is denoted by  $M_{PI}$ ,  $M_{PC}$  and  $M_{PM}$ , respectively, and  $M_{PI} + M_{PC} + M_{PM} = M_P$ . All cluster types can occur under individual randomisation, while  $M_{PM} = 0$  for cluster randomisation and  $M_{PI} = M_{PC} = 0$  for randomisation to opposite groups. Finally, let  $\gamma_S = M_S/N$  and  $\gamma_P = 2M_P/N$  denote the proportion of cluster members or observations that belong to a cluster of size 1 and 2 respectively, with  $\gamma_{PI} = 2M_{PI}/N$ ,  $\gamma_{PC} = 2M_{PC}/N$  and  $\gamma_{PM} = 2M_{PM}/N$  defined similarly.

## 2.2 Methods for Deriving Design Effects

In the current setting, the DEFF of interest is the variance of the treatment effect estimate accounting for clustering using GEEs, divided by the variance of the treatment effect estimate assuming all observations are independent. This DEFF can be used to account for clustering in sample size calculations by determining the sample size required assuming independence and then multiplying by the DEFF [2]. For continuous outcomes, DEFFs were derived algebraically based on a linear model ( $g = \text{identity link}$ ), where the treatment effect is expressed as a difference in means. For binary

outcomes, DEFFs were derived algebraically based on a logistic ( $g = \text{logit}$  link) or log binomial ( $g = \text{log}$  link) model, where the treatment effect is expressed as an odds ratio or a relative risk, respectively. Numerators for DEFFs were obtained using equation (2). Denominators were derived based on a linear regression model for continuous outcomes, and a logistic or log binomial regression model for binary outcomes, where all observations were assumed to be independent.

DEFFs were derived under the assumption that the total sample size  $N$  will be balanced between treatment groups. This assumption is generally made in standard sample size calculations, unless the groups are intended to be unbalanced by design (e.g. using 2:1 randomisation). For individual randomisation, an additional assumption was made that 25%, 25% and 50% of clusters of size 2 will be intervention only, control only, and a mixture of intervention and control, respectively (i.e.  $\gamma_{PI} = \gamma_{PC} = 0.5M_p/N$ ,  $\gamma_{PM} = M_p/N$ ). While this distribution is expected, it cannot be guaranteed using standard approaches for randomising cluster members. For cluster randomisation, an additional assumption was made that the sample size will be balanced between treatment groups conditional on cluster size. Such balance is expected when the sample size is balanced between treatment groups overall, but can only be guaranteed if the randomisation is stratified by cluster size. Note that this last assumption holds automatically for individual randomisation and randomisation to opposite groups under the other assumptions stated above. All assumptions were made to ensure that the resulting DEFFs only involved parameters that could reasonably be estimated before a trial commences.

### 2.3 Design Effects for Continuous Outcomes

Suppose  $Y_{ij}$  is a continuous outcome with constant variance  $\sigma^2$ . The general formula for the DEFF is given by:

$$\text{DEFF} = 1 + \rho(\gamma_{PI} + \gamma_{PC} - \gamma_{PM}) \quad (3)$$

for an independence working correlation structure and:

$$\text{DEFF} = \frac{1 - \rho^2}{1 - \rho^2 \gamma_S - \rho(\gamma_{PI} + \gamma_{PC} - \gamma_{PM})} \quad (4)$$

for an exchangeable working correlation structure (see Supporting Information for the derivation). These DEFFs only depend on the ICC and the proportion of observations that belong to a cluster of size 1, an intervention only cluster of size 2, a control only cluster of size 2 and a mixed cluster of size 2, all of which may be reasonably specified at the trial planning stage.

Based on the method of randomisation to be used for pairs (cluster, individual or opposite groups), equations (3) and (4) can be simplified to produce the DEFFs given in Table I. These now only depend on the ICC and the proportion of observations that belong to a cluster of size 1 or size 2. Note that the

DEFFs for the independence working correlation structure can be rewritten as weighted averages of the DEFFs for clusters of size 1 (DEFF=1) and size 2 (DEFF=1+ $\rho$ , 1 and 1- $\rho$  for cluster randomisation, individual randomisation and randomisation to opposite groups, respectively) with weights given by the proportions of observations that belong to clusters of size 1 and 2, respectively. Assuming outcomes of members of the same pair are positively correlated, DEFF>1 for cluster randomisation, indicating that a larger sample size is required for a trial involving both independent and paired data compared with a trial involving independent data only. In contrast, DEFF $\leq$ 1 for individual randomisation and randomisation to opposite groups, and hence a smaller sample size may suffice when paired data are included in a trial. Irrespective of the chosen method of randomisation, the DEFF is larger for the independence working correlation structure than the exchangeable working correlation structure and the difference increases with increasing ICC (Figure IA). This is not surprising, since the true correlation matrix was assumed to be exchangeable and the independence working correlation structure provides a poorer approximation to the truth for larger ICCs. As expected, the DEFF moves further away from 1 as the ICC increases (Figure IA) or the proportion of observations that belong to a pair increases (Figure IB).

## 2.4 Design Effects for Binary Outcomes

Suppose  $Y_{ij}$  is a binary outcome with prevalence  $\pi_I$  in the intervention group and  $\pi_C$  in the control group. If the analysis will be performed using the logit link function, the general formulae for the DEFFs based on the independence and exchangeable working correlation structures are both complex equations involving  $\pi_I$ ,  $\pi_C$ , the ICC and the proportion of observations belonging to different types of clusters (see Supporting Information). However, by specifying the method of randomisation the DEFFs can be simplified somewhat to those presented in Table II. For cluster randomisation, the DEFFs for binary outcomes are identical to the continuous outcome case. For individual randomisation and randomisation to opposite groups, the DEFFs differ from those for continuous outcomes due to the dependence on the outcome prevalence in the intervention and control groups via the term

$$\frac{\sqrt{\pi_I \pi_C (1 - \pi_I)(1 - \pi_C)}}{\pi_I (1 - \pi_I) + \pi_C (1 - \pi_C)} \quad (5)$$

When  $\pi_I = \pi_C$  or  $\pi_I = 1 - \pi_C$ , equation (5) equals 0.5 and the DEFFs for binary and continuous outcomes coincide. Equation (5) is otherwise less than 0.5 and hence the DEFFs for binary outcomes are larger than those for continuous outcomes. To illustrate how much larger the DEFFs for binary outcomes may be compared with continuous outcomes for individual randomisation and randomisation to opposite groups, the effect of varying  $\pi_I$  to produce an odds ratio ranging from 0.1 to 5 when  $\pi_C = 0.1$  or 0.5 is shown in Figure II. Over this range, the DEFFs for binary outcomes are minimised

and coincide with the continuous case when  $\pi_I = \pi_C$  (odds ratio=1). The DEFFs remain relatively constant unless the odds ratio is quite small, suggesting that DEFFs for continuous and binary outcomes will be similar in most settings likely to be encountered in practice. As a result, the relationships between the DEFF and both the ICC and the proportion of observations that belong to a pair seen in Figure 1 for continuous outcomes also apply to binary outcomes.

If a binary outcome will be analysed using the log link function, the DEFFs are identical to those presented in Table II for the logit link, except that the term  $\pi_I(1-\pi_I)+\pi_C(1-\pi_C)$  is replaced by  $\pi_I(1-\pi_C)+\pi_C(1-\pi_I)$  for individual randomisation and randomisation to opposite groups (see Supporting Information).

### 3. SIMULATION STUDY

#### 3.1 Simulation Methods

A simulation study was conducted to validate the DEFFs presented in Section 2, and to test whether these are applicable in settings where the assumptions made in deriving the DEFFs are relaxed. Simulation parameters were chosen to represent scenarios likely to be encountered in the neonatal trial setting. Clusters (mothers) consisted of a pair of members (twins) with probability 0.015, 0.2 or 1 (i.e.  $\gamma_p = 0.03, 0.33$  or 1), or a single member (singleton) otherwise. These probabilities reflect twin birth rates typical of the general population of mothers [23], mothers who have a preterm birth [14], and mothers who have twins, respectively, all of which are commonly of interest. Although the last setting does not involve any independent data, it was included to assess whether the DEFFs derived above also apply to simpler trials involving paired data only. Outcomes of pairs were positively correlated with an ICC of 0.2 or 0.8, as we have observed ICCs of these magnitudes for twins in neonatal trials. For each simulation scenario, 10,000 datasets were generated for analysis and the seed was saved to allow the simulation results to be replicated. Cluster members (singleton and twin infants) were randomly assigned to the intervention or control group under a simple randomisation scheme such that half of all cluster members were assigned to each treatment group. For clusters of size 2, both cluster members were randomly assigned under individual randomisation, while only the first cluster member was randomly assigned under cluster randomisation and randomisation to opposite groups, as this determined the treatment group allocation for the second cluster member. The randomisation process ensured that the total sample size  $N$  was balanced between treatment groups for individual randomisation, but did not guarantee that any other assumptions relating to balance made in deriving the DEFFs were met (see Section 2.2).

Continuous outcomes were randomly generated from the linear mixed effects model  $Y_{ij} = \beta_0 + \beta_1 X_{1ij} + a_i + e_{ij}$ , where  $a_i$  is a random cluster effect drawn from an  $N(0, \sigma_a^2)$  distribution,  $e_{ij}$  is a random error drawn from an  $N(0, \sigma_e^2)$  distribution, and other quantities are as defined previously.

Model coefficients were set to  $\beta_0 = 0$  and  $\beta_1 = 0.3$ , while variances were chosen to give a total variance of  $\sigma_a^2 + \sigma_e^2 = 1$  and produce the desired ICC according to the equation  $ICC = \sigma_a^2 / (\sigma_a^2 + \sigma_e^2)$ . Binary outcomes were randomly generated from a Bernoulli distribution. The marginal prevalence of the outcome was set to 0.4 in the control group and 0.3 in the intervention group (odds ratio=0.64, relative risk=0.75), and the method of Qaqish [24] was used to produce the desired ICC. The total sample size was  $N = 200$  or  $500$  for continuous outcomes, and  $N = 400$  or  $1000$  for binary outcomes. These sample sizes provide around 50% and 90% power respectively to detect the true treatment effects when all observations are independent.

For each simulation scenario, the expected DEFF and power to detect the true treatment effect were determined using the appropriate DEFF equation presented in Section 2. The expected power was calculated based on a Wald test for independent data using the effective sample size, given by the actual sample size divided by the expected DEFF. Each simulated dataset was analysed using a standard regression model, as well as GEEs with an independence or exchangeable working correlation structure and empirical sandwich variance estimation, based on the unadjusted mean model in equation (1). The identity link function was specified for continuous outcomes and both the logit and log link functions were used for binary outcomes. Observed DEFFs were calculated for each simulated dataset as the estimated variance for the treatment effect obtained using GEEs, divided by the estimated variance obtained using standard regression. The percentage relative difference in observed and expected DEFFs was also calculated as  $100 \times (\text{DEFF}_{\text{obs}} - \text{DEFF}_{\text{exp}}) / \text{DEFF}_{\text{exp}}$ . DEFFs and percentage relative differences were summarised across simulated datasets using the median due to their skewed distributions in many scenarios. Observed power was calculated for each scenario as the percentage of simulated datasets where the treatment effect estimate was statistically significant ( $p < 0.05$ ) and the percentage relative difference in observed and expected power was calculated as  $100 \times (\text{Power}_{\text{obs}} - \text{Power}_{\text{exp}}) / \text{Power}_{\text{exp}}$ . All simulations were performed using SAS version 9.3 (Cary, NC, USA).

### 3.2 Simulation Results

The simulation results for a continuous outcome and a total sample size of 500 are shown in Tables III and IV. As GEEs with an exchangeable working correlation structure occasionally failed to converge when the probability of a pair was only 0.015, results for this method were determined after excluding simulated datasets where convergence failed. This occurred for a maximum of 2% of simulated datasets. While no clear reasons were identified for the convergence problems that occurred, they were more common for individual randomisation as well as simulated datasets where few pairs were present and hence little information was available for estimating the working correlation parameter. For the independence working correlation structure, relative differences in the observed and expected DEFFs and power were small in all scenarios. Relative differences were also small for both the DEFF and power for the exchangeable working correlation structure in most scenarios. However, the observed DEFFs were noticeably larger than the expected DEFFs in scenarios where individual randomisation or

randomisation to opposite groups was performed, the probability of a pair was 0.015 or 0.2, and the ICC was 0.8. The largest difference was seen for randomisation to opposite groups when the probability of a pair was 0.015, where the median observed DEFF was 1.53 compared with the expected DEFF of 0.89. Additionally, the observed power was around 10% and 30% lower than expected for an ICC of 0.2 and 0.8, respectively, in scenarios where individual randomisation or randomisation to opposite groups was performed and the probability of a pair was 0.015. Further investigation of the problematic scenarios revealed that estimated treatment effects and their variances were highly variable for simulated datasets where the working correlation parameter was very close to 1, often resulting in huge observed DEFFs. The estimated correlation parameter was set to the maximum possible value of 0.9999 for around 5% and 50% of simulated datasets when the ICC was 0.2 and 0.8, respectively, and results were greatly improved when these datasets were excluded (data not shown). When the sample size was reduced from 500 to 200, convergence problems were slightly more common for the exchangeable working correlation structure, occurring in up to 3% of simulated datasets. Underestimation of power was again observed for the exchangeable working correlation structure with individual randomisation or randomisation to opposite groups, although underestimation could occur when the probability of a pair was 0.2 in addition to 0.015. Results for DEFFs with a sample size of 200 were similar to those obtained for a sample size of 500 (see Supporting Information).

The simulation results for a binary outcome with a logit link function and a total sample size of 1000 are shown in Tables V and VI. For the independence working correlation structure, the observed DEFFs and power were similar to the expected values in all scenarios. This was also true for the majority of scenarios for the exchangeable working correlation structure after excluding up to 2% of simulated datasets where the model failed to converge. The exception occurred when individual randomisation or randomisation to opposite groups was performed, the probability of a pair was 0.015 and the ICC was 0.8, where the observed DEFFs were somewhat larger than expected and the observed power was around 30% lower than expected. Consistent with the continuous outcome case, these problems could be attributed to highly variable estimated treatment effects and variances in datasets where the working correlation parameter was very close to 1. With a smaller sample size of 400, convergence problems increased for the exchangeable working correlation structure and affected up to 4% of simulated datasets. Observed DEFFs and power remained close to expected values for the independence working correlation structure. Unlike other settings, observed DEFFs for the exchangeable working correlation structure were substantially smaller than expected under individual randomisation or randomisation to opposite groups in scenarios where the probability of a pair was 0.015 and the ICC was 0.8. However, power was also around 60% lower than expected in these scenarios, consistent with other settings (see Supporting Information). These seemingly contradictory results, where both the observed DEFFs and power were too small, resulted from tiny estimated treatment effects and variances in many simulated datasets when the working correlation parameter was very close to 1. For example, in the worst case scenario where individual randomisation was performed and the median observed DEFF was 0.00 compared with the expected DEFF of 0.95, the estimated correlation parameter was 0.9999 for 63% of simulated datasets and the median estimated treatment effect and variance were both 0.00 in these datasets. Simulation results were very similar when the log link function was used instead of the logit link (see Supporting Information).

#### 4. EXAMPLE CALCULATIONS

To illustrate how the DEFFs presented in Section 2 can be used to determine the sample size for future trials involving a mixture of independent and paired data, we consider two hypothetical examples. First, suppose a trial is planned to assess the effect of a nutritional supplement in the general population of pregnant women on cognitive development in the infant. Women with a singleton or twin pregnancy will be randomised to receive the intervention or placebo (cluster randomisation) and cognitive development will be assessed using the Bayley Scales of Infant and Toddler Development (mean=100, standard deviation=15) [25]. It is expected that the ICC will be 0.7, 1.5% of mothers will have twins and hence approximately 3% of infants will be from a twin birth (i.e.  $\gamma_p = 0.03$ ). If the outcomes of all infants included in the trial were independent, 444 infants would be required (222 per group) to provide 80% power to detect a clinically important 4 point improvement in the mean cognitive score, based on an independent samples t-test with two-sided  $\alpha = 0.05$ . However, the inclusion of twins violates the assumption of independence and hence this sample size should be multiplied by the DEFF for cluster randomisation from Table I. If the analysis will be performed using GEEs with an independence working correlation structure, the DEFF is 1.02 and the required total sample size increases to 454 infants to account for the twins. In comparison, the DEFF for GEEs with an exchangeable working correlation structure is 1.01 and the total sample size increases to 450 infants to account for the twins. These relatively small changes in the sample size reflect the low twin birth rate in the target population. With 1.5% of mothers delivering twins, it is expected that consent will be needed from 448 or 444 pregnant women in order to include 454 or 450 infants in the trial, respectively.

Now consider a trial designed to assess the effect of a new infant formula intended to reduce neonatal morbidity among infants born preterm. Mothers giving birth to a singleton or twins before 37 weeks' gestation will be approached for consent and infants will be randomised individually to receive the new formula or the standard formula. It is anticipated that 30% of infants born preterm will be from a twin birth (i.e.  $\gamma_p = 0.3$ ), the ICC for neonatal morbidity will be 0.5, and the new formula will reduce the risk of neonatal morbidity from 20% to 14% (odds ratio=0.65). If there were no twins included in the trial, a total of 1294 infants (647 per group) would be required to detect this reduction in the risk of neonatal morbidity between the treatment groups, based on a continuity-corrected chi-square test with two-sided  $\alpha = 0.05$  and 80% power. This is also the number of infants required after accounting for the twins using the DEFF for the logit link and individual randomisation from Table II, provided the analysis will be performed using GEEs with an independence working correlation structure, since the DEFF is 1.00 in this case. Alternatively, if the exchangeable working correlation structure will be used then the DEFF is 0.91 and the total number of infants required reduces by 116 to 1178 (589 per group).

#### 5. DISCUSSION

We have derived DEFFs that can be used to calculate the sample size for two-arm, parallel group, randomised controlled trials including both independent and paired data when the analysis will be

performed using GEEs. These DEFFs are relatively simple to calculate and depend on quantities that can usually be specified at the trial planning stage, namely the ICC and the proportion of observations that belong to a pair. We have shown through simulation that the derived DEFFs are valid across a range of scenarios. However, the DEFFs presented for the exchangeable working correlation structure should be used with caution when members of the same pair are to be randomised independently or to opposite treatment groups, especially when the probability of a pair is low or the ICC is high, as use of these equations may result in lower than expected power. A conservative approach when one of these methods of randomisation is chosen is to use the DEFF for the independence working correlation structure, since this will result in a larger sample size than the exchangeable working correlation structure and provide the desired power if an independence working correlation structure is specified in the analysis. If an exchangeable working correlation structure is used for analysis, careful attention should be paid to the working correlation parameter, as we found that working correlations close to 1 can lead to unusually large or small estimates for the treatment effect and its variance. Large correlations are uncommon for clustered data in general [26-28] but ICCs as high as 0.9 have been reported in twin studies [29, 30].

When calculating the required sample size for a trial, it may be tempting to choose the exchangeable working correlation structure over the independence structure in general due to the smaller DEFF. As seen in the example sample size calculations, this approach can lead to a substantial reduction in the required sample size compared with the independence structure, which translates into shorter recruitment periods and reduced trial costs. This finding is consistent with previous studies that have shown efficiency gains for the exchangeable compared to the independence working correlation structure in some settings [17, 18]. Despite these advantages, there are several reasons to prefer an independence working correlation structure for both sample size planning and analysis. First, convergence problems can occur for the exchangeable working correlation structure and an alternative method of analysis will then be required. If the sample size was determined assuming an exchangeable working correlation structure, this could result in an underpowered trial. Second, GEEs with an exchangeable working correlation structure do not necessarily estimate a marginal treatment parameter of interest when the cluster size is informative, that is, when the outcome of interest is related to the size of the cluster conditional on the covariates in the analysis model [20]. This issue has been investigated in the neonatal trial setting, where GEEs can be used to estimate marginal treatment effects with either a mother-level or infant-level interpretation and the latter interpretation is typically of primary interest. The independence working correlation structure is recommended to produce treatment effect estimates with this interpretation, while the exchangeable working correlation structure may not produce estimates of either marginal treatment effect [21].

Before the DEFF can be calculated, a choice must be made regarding the method of randomisation. Although cluster randomisation results in larger sample sizes than other methods of randomisation, it is often the more feasible approach. For example, in the neonatal setting cluster randomisation is necessary for interventions given to the mother and is usually more acceptable to parents for interventions given to the infant [12]. If individual randomisation is chosen, randomly permuted blocks should be used with caution as blocking affects the percentage of clusters of size 2 that are intervention only, control only, and a mixture of intervention and control, assumed to be 25%, 25% and 50%

respectively when deriving the DEFFs presented in Tables I and II. For example, in a trial consisting of paired data only, employing individual randomisation with randomly permuted blocks of size 2 will result in 100% of pairs being a mixture of intervention and control, equivalent to randomising members of the same pair to opposite treatment groups. With fewer pairs or larger block sizes, the percentage of pairs assigned to opposite treatment groups will decrease from 100% but remain greater than the assumed value of 50%. The more general DEFFs given in equations (3), (4) and the Supporting Information can be used to determine the sample size required under individual randomisation while accounting for different distributions of clusters of size 2 resulting from blocking. Alternatively, the DEFFs presented in Tables I and II can be used to calculate the sample size that will provide at least the desired power.

In order to calculate the sample size required for a new trial, an estimate of the ICC must be obtained for the outcome of interest. Unfortunately, a recent review of neonatal trials including twins found that none of the 26 included trials published ICCs [14]. Moderate to high ICCs have been reported for several neonatal outcomes in methodological papers [16, 31] and thus including twins in a trial is expected to have a non-negligible impact on the sample size unless the proportion of twins is very small. More published ICCs are needed to assist with sample size planning for future neonatal trials including twins. The best way to calculate ICCs for these trials, particularly whether calculations should involve data from twins only or all infants, is an important area for further research.

Several of our derived DEFFs can be shown algebraically to be consistent with previously published equations, providing further validation for their use. The DEFF reported by Gauderman and Barlow [4] for ophthalmology trials involving patients with one or both eyes included, cluster randomisation and a continuous or binary outcome agrees with our DEFF for the independence working correlation structure. The equations presented by Pan [10] for the GEE robust variance estimate assuming a binary outcome with a logit link, constant or varying treatment group within clusters, varying cluster size and an independence or exchangeable working correlation structure can be shown to produce DEFFs that match ours for cluster randomisation and randomisation to opposite treatment groups when the maximum cluster size is set to 2. The exact DEFF of  $1 + (\tilde{m} - 1)\rho$  for cluster randomisation with varying cluster size based on a summary statistics analysis with cluster size weights [6] reduces to our DEFF for a continuous or binary outcome with an independence working correlation structure when only clusters of size 1 and 2 are considered. The widely used approximation, where  $\tilde{m}$  is simply replaced by the average cluster size, is known to underestimate the DEFF [3] and hence we recommend using our DEFF instead to avoid underestimating the required sample size. Finally, the DEFF derived by Vierron and Giraudeau [2] based on a two-way analysis of variance for a continuous outcome reduces to our equation (3) for independent and paired data analysed using GEEs with an independence working correlation structure under the assumption that the total sample size is balanced between treatment groups. Our finding that  $\text{DEFF} > 1$  for cluster randomisation but  $\text{DEFF} \leq 1$  for individual randomisation and randomisation to opposite groups is also expected, since ignoring clustering is known to produce standard errors that are too small for between-cluster effects but too large for within-cluster effects [1]. Despite these consistencies with previous work, to our knowledge this is the first article to derive equations for DEFFs that are directly applicable to trials involving both independent

and paired data, and will allow the applied statistician to easily determine the sample size required for such trials. This is important, since recent systematic reviews of neonatal [14] and ophthalmic [13] trials indicate that clustering due to the inclusion of paired data is rarely taken into account in sample size calculations. Our simulation results suggest the DEFFs presented here can also be used in sample size planning for simpler trials involving paired data only, thus further extending their potential use.

Our study had several limitations. First, the maximum cluster size considered was 2. In most settings where independent and paired data arise, larger cluster sizes are not possible, such as trials measuring outcomes on eyes or knees of patients. In neonatal trials, clusters of size 3 (triplets) or more are sometimes eligible for inclusion, although these are usually so rare that they could reasonably be ignored at the sample size planning stage. Alternatively, the DEFFs presented in this article could be extended to allow for larger cluster sizes if the proportion of triplets was expected to be non-negligible and this is an interesting area for further research. Second, all DEFFs were derived assuming that the marginal or population-average treatment effect is of interest and hence clustering will be taken into account in the analysis using GEEs. Mixed effects models provide an alternative approach for handling clustered data and estimate conditional or cluster-specific treatment effects. The choice between these analysis approaches should be driven by the estimand of interest and is particularly important for the logistic model, where marginal and conditional treatment effects differ, but less important for the linear or log binomial model, where the two treatment effects coincide [32, 33]. Sample size methods based on mixed effects models have been proposed in other settings, including longitudinal studies [34] and trials where clusters of size 1 only occur in one treatment group [35, 36]. Developing sample size methods for trials involving independent and paired data in both treatment groups based on mixed effects models may be of limited value, since these models can produce biased treatment effect estimates and poor coverage rates in this setting [16]. GEEs are more popular for analysing trials involving independent and paired data in practice [12-14] and hence were the focus of this article. Finally, the scenarios considered in our simulation study involved a relatively large number of clusters (mothers), as would often be expected in trials including both independent and paired data due to the small cluster sizes of 1 or 2. However, few clusters may be required for trials where large treatment effects are anticipated. The DEFFs presented in this article may not be appropriate for trials with few clusters, since GEEs underestimate the variance of parameter estimates in this case and small sample corrections are recommended [37, 38]. Further research is needed to develop sample size methods that are applicable to trials involving independent and paired data with few clusters.

In conclusion, clustering can easily be taken into account in sample size calculations for randomised trials including both independent and paired data using the DEFFs presented in this article. The relevant DEFF should be applied after carefully considering whether GEEs with an independence or exchangeable working correlation structure are the most appropriate method of analysis for the trial in question.

## ACKNOWLEDGEMENTS

Lisa Yelland and Katherine Lee were supported by Australian National Health and Medical Research Council Fellowships [Early Career Fellowship ID 1052388 for LY, Career Development Fellowship ID

1053609 for KL]. This research was supported by a Women's and Children's Hospital Foundation Research Project Grant. We are grateful to an Associate Editor and two anonymous reviewers for providing helpful comments.

## REFERENCES

1. Cannon MJ, Warner L, Taddei JA, Kleinbaum DG. What can go wrong when you assume that correlated data are independent: an illustration from the evaluation of a childhood health intervention in Brazil. *Statistics in Medicine* 2001; **20**: 1461-1467.
2. Vierron E, Giraudeau B. Design effect in multicenter studies: gain or loss of power? *BMC Medical Research Methodology* 2009; **9**: 39, DOI: 10.1186/1471-2288-9-39.
3. Donner A, Klar N. *Design and Analysis of Cluster Randomization Trials in Health Research*. Arnold: London, 2000.
4. Gauderman WJ, Barlow WE. Sample size calculations for ophthalmologic studies. *Archives of Ophthalmology* 1992; **110**: 690-692.
5. Borkhoff CM, Johnston PR, Stephens D, Atenafu E. The special case of the 2 x 2 table: asymptotic unconditional McNemar test can be used to estimate sample size even for analysis based on GEE. *Journal of Clinical Epidemiology* 2015; **68**: 733-739, DOI: 10.1016/j.jclinepi.2014.09.025.
6. Kerry SM, Bland JM. Unequal cluster sizes for trials in English and Welsh general practice: implications for sample size calculations. *Statistics in Medicine* 2001; **20**: 377-390.
7. Eldridge SM, Ashby D, Kerry S. Sample size for cluster randomized trials: effect of coefficient of variation of cluster size and analysis method. *International Journal of Epidemiology* 2006; **35**: 1292-1300, DOI: 10.1093/ije/dy1129.
8. Liu GH, Liang KY. Sample size calculations for studies with correlated observations. *Biometrics* 1997; **53**: 937-947, DOI: 10.2307/2533554.
9. Shih WJ. Sample size and power calculations for periodontal and other studies with clustered samples using the method of generalized estimating equations. *Biometrical Journal* 1997; **8**: 899-908.
10. Pan W. Sample size and power calculations with correlated binary data. *Controlled Clinical Trials* 2001; **22**: 211-227.
11. Li ZG, McKeague IW. Power and Sample Size Calculations for Generalized Estimating Equations Via Local Asymptotics. *Statistica Sinica* 2013; **23**: 231-250, DOI: 10.5705/ss.2011.081.
12. Hibbs AM, Black D, Palermo L, Cnaan A, Luan XQ, Truog WE, Walsh MC, Ballard RA. Accounting for multiple births in neonatal and perinatal trials: systematic review and case study. *Journal of Pediatrics* 2010; **156**: 202-208, DOI: 10.1016/j.jpeds.2009.08.049.
13. Lee CF, Cheng ACO, Fong DYT. Eyes or Subjects: Are Ophthalmic Randomized Controlled Trials Properly Designed and Analyzed? *Ophthalmology* 2012; **119**: 869-872, DOI: 10.1016/j.ophtha.2011.09.025.

14. Yelland LN, Sullivan TR, Makrides M. Accounting for multiple births in randomised trials: a systematic review. *Archives of Disease in Childhood-Fetal and Neonatal Edition* 2015; **100**: F116-120, DOI: 10.1136/archdischild-2014-306239.
15. Liang KY, Zeger SL. Longitudinal data analysis using generalized linear models. *Biometrika* 1986; **73**: 13-22.
16. Yelland LN, Salter AB, Ryan P, Makrides M. Analysis of binary outcomes from randomised trials including multiple births: when should clustering be taken into account? *Paediatric and Perinatal Epidemiology* 2011; **25**: 283-297, DOI: 10.1111/j.1365-3016.2011.01196.x.
17. Fitzmaurice GM. A caveat concerning independence estimating equations with multivariate binary data. *Biometrics* 1995; **51**: 309-317.
18. Mancl LA, Leroux BG. Efficiency of regression estimates for clustered data. *Biometrics* 1996; **52**: 500-511.
19. Pepe MS, Anderson GL. A cautionary note on inference for marginal regression-models with longitudinal data and general correlated response data. *Communications in Statistics-Simulation and Computation* 1994; **23**: 939-951.
20. Seaman S, Pavlou M, Copas A. Review of methods for handling confounding by cluster and informative cluster size in clustered data. *Statistics in Medicine* 2014; **33**: 5371-5387, DOI: 10.1002/sim.6277.
21. Yelland LN, Sullivan TR, Pavlou M, Seaman SR. Analysis of randomised trials including multiple births when birth size is informative. *Paediatric and Perinatal Epidemiology* 2015; **29**: 567-575, DOI: 10.1111/ppe.12228.
22. Dunlop DD. Regression for longitudinal data - a bridge from least-squares regression. *American Statistician* 1994; **48**: 299-303, DOI: 10.2307/2684838.
23. Blondel B, Macfarlane A, Gissler M, Breart G, Zeitlin J, Peristat Study Group. Preterm birth and multiple pregnancy in European countries participating in the PERISTAT project. *BJOG: an International Journal of Obstetrics and Gynaecology* 2006; **113**: 528-535, DOI: 10.1111/j.1471-0528.2006.00923.x.
24. Qaqish BF. A family of multivariate binary distributions for simulating correlated binary variables with specified marginal means and correlations. *Biometrika* 2003; **90**: 455-463.
25. Bayley N. *Bayley Scales of Infant and Toddler Development, Third Edition*. Pearson Education Inc.: San Antonio, TX, 2006.
26. Adams G, Gulliford MC, Ukoumunne OC, Eldridge S, Chinn S, Campbell MJ. Patterns of intra-cluster correlation from primary care research to inform study design and analysis. *Journal of Clinical Epidemiology* 2004; **57**: 785-794.
27. Taljaard M, Donner A, Villar J, Wojdyla D, Velazco A, Bataglia V, Faundes A, Langer A, Narvaez A, Valladares E, Carroli G, Zavaleta N, Shah A, Campodonico L, Romero M, Reynoso S, de Padua KS, Giordano D, Kublickas M, Acosta A. Intracluster correlation coefficients from the 2005 WHO Global Survey on Maternal and Perinatal Health: implications for implementation research. *Paediatric and Perinatal Epidemiology* 2008; **22**: 117-125, DOI: 10.1111/j.1365-3016.2007.00901.x.
28. Elley CR, Kerse N, Chondros P, Robinson E. Intraclass correlation coefficients from three cluster randomised controlled trials in primary and residential health care. *Australian and New Zealand Journal of Public Health* 2005; **29**: 461-467.

29. Baird J, Osmond C, MacGregor A, Snieder H, Hales CN, Phillips DI. Testing the fetal origins hypothesis in twins: the Birmingham twin study. *Diabetologia* 2001; **44**: 33-39, DOI: 10.1007/s001250051577.
30. Hur YM, Luciano M, Martin NG, Boomsma DI, Iacono WG, McGue M, Shin JS, Jun JK, Ooki S, van Beijsterveldt CE, Han JY. A comparison of twin birthweight data from Australia, the Netherlands, the United States, Japan, and South Korea: are genetic and environmental variations in birthweight similar in Caucasians and East Asians? *Twin Research and Human Genetics* 2005; **8**: 638-648, DOI: 10.1375/183242705774860169.
31. Gates S, Brocklehurst P. How should randomised trials including multiple pregnancies be analysed? *BJOG: an International Journal of Obstetrics and Gynaecology* 2004; **111**: 213-219, DOI: 10.1111/j.1471-0528.2004.00059.x.
32. Neuhaus JM, Kalbfleisch JD, Hauck WW. A comparison of cluster-specific and population-averaged approaches for analyzing correlated binary data. *International Statistical Review* 1991; **59**: 25-35.
33. Ritz J, Spiegelman D. Equivalence of conditional and marginal regression models for clustered and longitudinal data. *Statistical Methods in Medical Research* 2004; **13**: 309-323.
34. Dang QY, Mazumdar S, Houck PR. Sample size and power calculations based on generalized linear mixed models with correlated binary outcomes. *Computer Methods and Programs in Biomedicine* 2008; **91**: 122-127, DOI: 10.1016/j.cmpb.2008.03.001.
35. Moerbeek M, Wong WK. Sample size formulae for trials comparing group and individual treatments in a multilevel model. *Statistics in Medicine* 2008; **27**: 2850-2864, DOI: 10.1002/sim.3115.
36. Roberts C, Roberts SA. Design and analysis of clinical trials with clustering effects due to treatment. *Clinical Trials* 2005; **2**: 152-162, DOI: 10.1191/1740774505cn076oa.
37. Lu B, Preisser JS, Qaqish BF, Suchindran C, Bangdiwala S, Wolfson M. A comparison of two bias-corrected covariance estimators for generalized estimating equations. *Biometrics* 2007; **63**: 935-941, DOI: 10.1111/j.1541-0420.2007.00764.x.
38. Mancl LA, DeRouen TA. A covariance estimator for GEE with improved small-sample properties. *Biometrics* 2001; **57**: 126-134.

Author

Table I: Design effects for continuous outcomes by method of randomisation and choice of working correlation structure

Randomisation Method for Pairs	Independence Working Correlation	Exchangeable Working Correlation
Cluster	$1 + \rho\gamma_p$	$\frac{1 + \rho}{1 + \rho\gamma_s}$
Individual	1	$\frac{1 - \rho^2}{1 - \rho^2\gamma_s}$
Opposite Groups	$1 - \rho\gamma_p$	$\frac{1 - \rho}{1 - \rho\gamma_s}$

Author Manuscript

Table II: Design effects for binary outcomes with a logit link by method of randomisation and choice of working correlation structure

Randomisation Method for Pairs	Independence Working Correlation	Exchangeable Working Correlation
Cluster	$1 + \rho\gamma_P$	$\frac{1 + \rho}{1 + \rho\gamma_S}$
Individual	$1 + \rho\gamma_P \left( \frac{1}{2} - \frac{\sqrt{\pi_I \pi_C (1 - \pi_I)(1 - \pi_C)}}{\pi_I (1 - \pi_I) + \pi_C (1 - \pi_C)} \right)$	$\frac{(1 - \rho^2) \left[ 1 - \rho^2 \gamma_S - \rho\gamma_P \left( \frac{1}{2} + \frac{\sqrt{\pi_I \pi_C (1 - \pi_I)(1 - \pi_C)}}{\pi_I (1 - \pi_I) + \pi_C (1 - \pi_C)} \right) \right]}{(1 - \rho^2 \gamma_S)(1 - \rho^2 \gamma_S - \rho\gamma_P)}$
Opposite Groups	$1 - \rho\gamma_P \left( \frac{2\sqrt{\pi_I \pi_C (1 - \pi_I)(1 - \pi_C)}}{\pi_I (1 - \pi_I) + \pi_C (1 - \pi_C)} \right)$	$\frac{(1 - \rho) \left[ 1 - \rho^2 \gamma_S - \rho\gamma_P \left( \frac{2\sqrt{\pi_I \pi_C (1 - \pi_I)(1 - \pi_C)}}{\pi_I (1 - \pi_I) + \pi_C (1 - \pi_C)} \right) \right]}{(1 - \rho\gamma_S)(1 - \rho^2 \gamma_S - \rho\gamma_P)}$

Author Manuscript

Table III: Observed and expected design effects for a continuous outcome with a sample size of 500

Randomisation Method for Pairs	Probability of a Pair	ICC	GEE Independence			GEE Exchangeable**		
			Observed DEFF*	Expected DEFF	Relative Difference (%)*	Observed DEFF*	Expected DEFF	Relative Difference (%)*
Cluster	0.015	0.2	1.00	1.01	-0.11	1.00	1.00	-0.13
		0.8	1.02	1.02	-0.33	1.01	1.01	-0.17
	0.2	0.2	1.06	1.07	-0.23	1.06	1.06	-0.27
		0.8	1.26	1.27	-0.39	1.17	1.17	-0.23
	1	0.2	1.19	1.20	-0.42	1.19	1.20	-0.42
		0.8	1.80	1.80	-0.09	1.80	1.80	-0.09
Individual	0.015	0.2	1.00	1.00	0.00	1.00	1.00	0.12
		0.8	1.00	1.00	-0.01	1.00	0.95	5.51
	0.2	0.2	1.00	1.00	-0.11	0.98	0.99	-0.28
		0.8	1.00	1.00	-0.17	0.67	0.63	6.89
	1	0.2	1.00	1.00	-0.07	0.96	0.96	-0.35
		0.8	1.00	1.00	-0.26	0.36	0.36	-0.33
Opposite Groups	0.015	0.2	1.00	0.99	0.13	1.00	0.99	0.84
		0.8	0.98	0.98	0.30	1.53	0.89	71.64
	0.2	0.2	0.93	0.93	0.10	0.93	0.92	0.23
		0.8	0.74	0.73	0.38	0.46	0.43	6.32
	1	0.2	0.80	0.80	0.08	0.80	0.80	0.08
		0.8	0.20	0.20	0.08	0.20	0.20	0.08

\* Median value across 10,000 simulated datasets.

\*\* Results calculated after excluding up to 2% of simulated datasets where model failed to converge.

Author Manuscript

Table IV: Observed and expected power for a continuous outcome with a sample size of 500

Randomisation Method for Pairs	Probability of a Pair	ICC	GEE Independence			GEE Exchangeable*		
			Observed Power	Expected Power	Relative Difference (%)	Observed Power	Expected Power	Relative Difference (%)
Cluster	0.015	0.2	91.78	91.69	0.10	90.75	91.71	-1.05
		0.8	91.14	91.23	-0.10	91.31	91.50	-0.20
	0.2	0.2	89.85	90.11	-0.28	90.23	90.31	-0.09
		0.8	84.69	84.62	0.08	87.56	87.20	0.42
	1	0.2	86.62	86.47	0.17	86.62	86.47	0.17
		0.8	71.12	70.54	0.82	71.12	70.54	0.82
Individual	0.015	0.2	92.05	91.84	0.23	85.80	91.87	-6.60
		0.8	91.87	91.84	0.04	67.80	93.07	-27.15
	0.2	0.2	91.76	91.84	-0.08	92.03	92.18	-0.16
		0.8	91.86	91.84	0.03	97.15	98.85	-1.72
	1	0.2	91.92	91.84	0.09	92.74	92.83	-0.10
		0.8	91.78	91.84	-0.06	99.95	99.99	-0.04
Opposite Groups	0.015	0.2	92.38	91.99	0.43	83.18	92.02	-9.61
		0.8	92.38	92.43	-0.05	62.70	94.37	-33.56
	0.2	0.2	93.10	93.47	-0.40	93.26	93.71	-0.48
		0.8	97.34	97.48	-0.15	99.84	99.92	-0.08
	1	0.2	96.54	96.33	0.22	96.54	96.33	0.22
		0.8	100.00	100.00	0.00	100.00	100.00	0.00

\* Results calculated after excluding up to 2% of simulated datasets where model failed to converge.

Author

Table V: Observed and expected design effects for a binary outcome with a logit link and a sample size of 1000

Randomisation Method for Pairs	Probability of a Pair	ICC	GEE Independence			GEE Exchangeable**		
			Observed DEFF*	Expected DEFF	Relative Difference (%)*	Observed DEFF*	Expected DEFF	Relative Difference (%)*
Cluster	0.015	0.2	1.01	1.01	-0.04	1.00	1.00	-0.03
		0.8	1.02	1.02	-0.08	1.01	1.01	-0.05
	0.2	0.2	1.07	1.07	-0.10	1.06	1.06	-0.08
		0.8	1.26	1.27	-0.13	1.17	1.17	-0.06
	1	0.2	1.20	1.20	-0.15	1.20	1.20	-0.15
		0.8	1.80	1.80	-0.01	1.80	1.80	-0.01
Individual	0.015	0.2	1.00	1.00	-0.02	1.00	1.00	-0.05
		0.8	1.00	1.00	0.00	0.96	0.95	0.63
	0.2	0.2	1.00	1.00	-0.03	0.98	0.99	-0.15
		0.8	1.00	1.00	-0.10	0.63	0.63	0.92
	1	0.2	1.00	1.00	-0.14	0.96	0.96	-0.22
		0.8	1.00	1.00	-0.24	0.36	0.36	-0.57
Opposite Groups	0.015	0.2	0.99	0.99	0.03	1.00	0.99	0.34
		0.8	0.98	0.98	0.06	0.96	0.89	6.94
	0.2	0.2	0.93	0.93	0.04	0.92	0.92	0.10
		0.8	0.73	0.73	0.14	0.43	0.43	0.76
	1	0.2	0.80	0.80	-0.01	0.80	0.80	-0.01
		0.8	0.20	0.20	0.00	0.20	0.20	0.00

\* Median value across 10,000 simulated datasets.

\*\* Results calculated after excluding up to 2% of simulated datasets where model failed to converge.

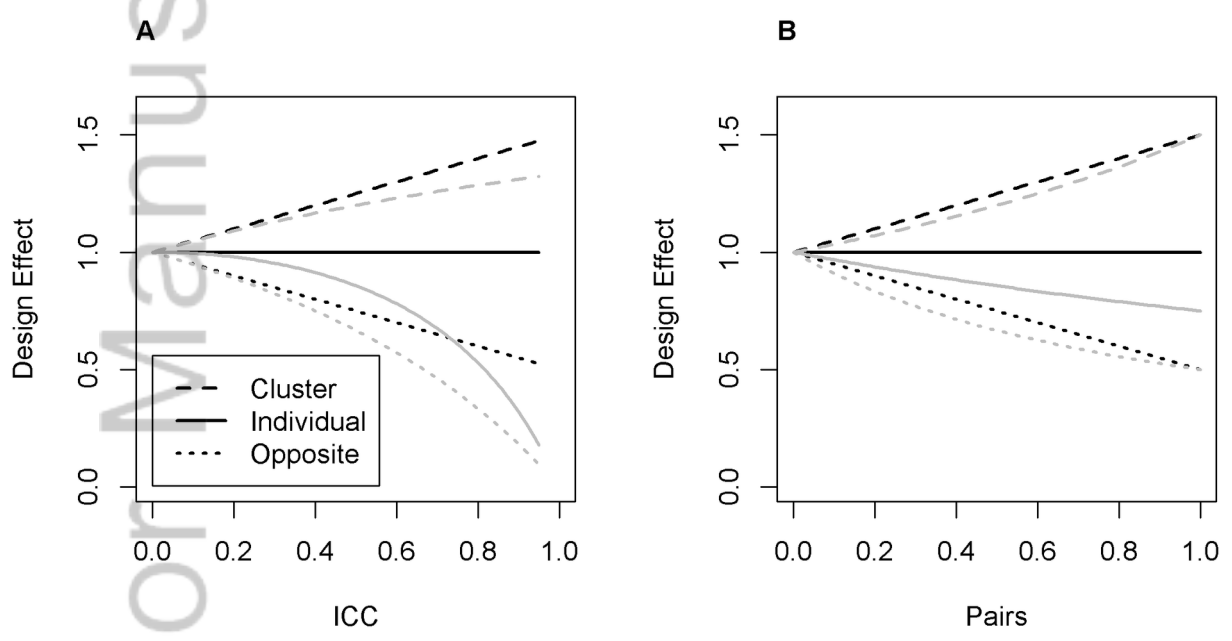
Author Manuscript

Table VI: Observed and expected power for a binary outcome with a logit link and a sample size of 1000

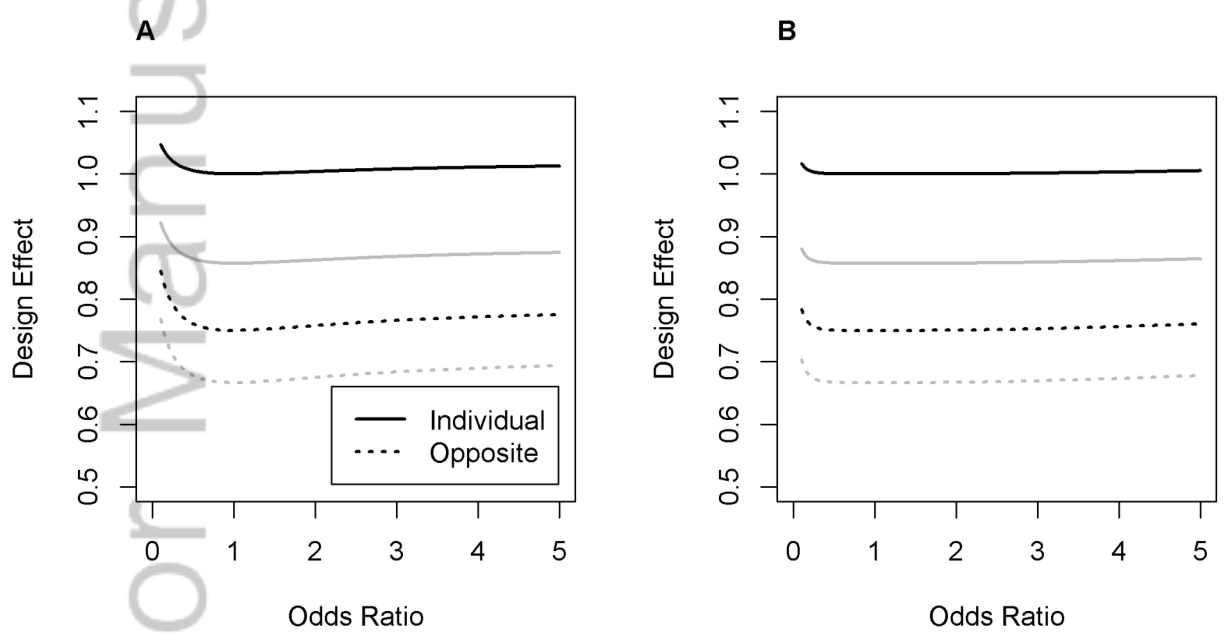
Randomisation Method for Pairs	Probability of a Pair	ICC	GEE Independence			GEE Exchangeable*		
			Observed Power	Expected Power	Relative Difference (%)	Observed Power	Expected Power	Relative Difference (%)
Cluster	0.015	0.2	91.24	91.18	0.07	91.30	91.20	0.11
		0.8	90.93	90.71	0.24	91.31	90.98	0.36
	0.2	0.2	89.47	89.55	-0.09	89.80	89.77	0.04
		0.8	83.96	83.96	0.00	86.85	86.58	0.31
	1	0.2	85.51	85.85	-0.39	85.51	85.85	-0.39
		0.8	69.41	69.83	-0.61	69.41	69.83	-0.61
Individual	0.015	0.2	91.04	91.33	-0.32	89.47	91.36	-2.07
		0.8	91.11	91.33	-0.24	64.87	92.61	-29.96
	0.2	0.2	91.62	91.33	0.32	91.86	91.69	0.19
		0.8	91.23	91.33	-0.10	99.62	98.70	0.93
	1	0.2	91.76	91.33	0.47	92.56	92.35	0.22
		0.8	91.72	91.31	0.45	100.00	99.98	0.02
Opposite Groups	0.015	0.2	91.43	91.49	-0.06	90.09	91.52	-1.56
		0.8	91.46	91.94	-0.53	69.01	93.96	-26.56
	0.2	0.2	92.83	93.02	-0.20	92.99	93.27	-0.30
		0.8	97.19	97.22	-0.03	99.99	99.90	0.09
	1	0.2	96.19	96.00	0.20	96.19	96.00	0.20
		0.8	100.00	100.00	0.00	100.00	100.00	0.00

\* Results calculated after excluding up to 2% of simulated datasets where model failed to converge.

Author Manuscript



SIM\_7201\_F1.tiff



SIM\_7201\_F2.tiff