# Choosing appropriate analysis methods for cluster randomised cross-over trials with a binary outcome: Online appendix

Katy E. Morgan*,[1], Andrew B. Forbes[2], Ruth H. Keogh[1], Vipul Jairath[3] and Brennan C. Kahan[4]

## 1 Weights used with a linear regression of cluster-level summaries

In the Methods section of our paper we introduced a cluster-level summary linear regression that used various weights. In addition to the weights defined in the main paper, we also looked at weights that relax the assumption of the event rates being equal:

$$\left(P_A(1-P_A)\frac{1+(n_{Ai}-1)\hat{\rho}_c}{n_{Ai}}+P_B(1-P_B)\frac{1+(n_{Bi}-1)\hat{\rho}_c}{n_{Bi}}-2\hat{\eta}\sqrt{P_A(1-P_A)P_B(1-P_B)}\right)^{-1}, \tag{A1}$$

where $n_{Xi}$ is the number of subjects in the period of cluster $i$ in which patients receive treatment $X$, with $X$ running over treatments $A$ and $B$, $P_X$ is the proportion of events in all cluster-periods in which treatment $X$ is given, and $\hat{\rho}_c$ and $\hat{\eta}$ are estimates of the ICCs (see Appendix Section 2). The weights considered so far (aside from that based on the binomial variance inverse) use the variance for a difference in proportions in Giraudeau *et al.* in Ref. [1]. Using instead the variance in Donner, Klar and Zou in Ref. [2], which uses a cross-term of $(P_A(1-P_A)+P_B(1-P_B))/2$ instead of $\sigma_A\sigma_B=\sqrt{P_A(1-P_A)P_B(1-P_B)}$ as in Giraudeau *et al.* [1], we get two more weightings:

$$\left(\frac{P_A(1-P_A)}{n_{Ai}}+\frac{P_B(1-P_B)}{n_{Bi}}\right)^{-1}, \tag{A2}$$

which assumes $\rho_c=\eta$, and:

$$\left(P_A(1-P_A)\frac{1+(n_{Ai}-1)\hat{\rho}_c}{n_{Ai}}+P_B(1-P_B)\frac{1+(n_{Bi}-1)\hat{\rho}_c}{n_{Bi}}-\hat{\eta}\left(P_A(1-P_A)+P_B(1-P_B)\right)\right)^{-1}, \tag{A3}$$

which relaxes this assumption.

*Email: katy.morgan@lshtm.ac.uk
[1]Department of Medical Statistics, London School of Hygiene and Tropical Medicine, London, UK
[2]School of Public Health and Preventive Medicine, Monash University, Melbourne, Australia
[3]Department of Medicine, London Health Sciences Network & Western University, London, ON, Canada
[4]Pragmatic Clinical Trials Unit, Queen Mary University of London, London, UK

## 2 ICC estimators for use in weighted linear regressions

Some of the weights that we use with the cluster-level summary regression involve estimates of $\rho_c$ and the correlation between two observations in different periods within the same cluster, $\eta$. We used the ANOVA estimator for $\rho_c$ given in Reference [2]; for other examples of the ANOVA estimator with one level of clustering see for example Refs. [3–5]. The ANOVA estimator from Ref. [2] is defined as:

$$\hat{\rho}_c = \frac{\text{MSB} - \text{MSW}}{\text{MSB} + (n_0 - 1)\text{MSW}}\,, \tag{A4}$$

where:

$$\text{MSB} = \sum_{X=A}^{B}\sum_{i=1}^{C}\frac{n_{Xi}(P_{Xi} - P_X)^2}{2C-2}\,,\ \text{MSW} = \sum_{X=A}^{B}\sum_{i=1}^{C}\frac{n_{Xi}P_{Xi}(1-P_{Xi})}{N-2C}\,, \tag{A5}$$

and $n_0 = \left(N - \sum_{X=A}^{B}\sum_{i=1}^{C}\frac{n_{Xi}^2}{n_X}\right)\Big/(2C-2)$, where $N$ is the total number of individuals in the trial and other parameters are defined as in Section 1.

For the correlation $\eta$ we used the pairwise estimator that is given in Donner *et al.* in Ref. [2]:

$$\hat{\eta} = \frac{\sum_{i=1}^{C}(Y_{Ai} - n_{Ai}P)(Y_{Bi} - n_{Bi}P)}{\sqrt{\left(\sum_{i=1}^{C}n_{Ai}(Y_{Ai} - 2Y_{Ai}P + n_{Ai}P^2)\right)\left(\sum_{i=1}^{C}n_{Bi}(Y_{Bi} - 2Y_{Bi}P + n_{Bi}P^2)\right)}}\,. \tag{A6}$$

Note that this estimator is called $\rho_{12}$ in Ref. [2].

## 3 Stata code for methods of analysis

Here we provide the Stata [6] code for the analysis methods we used in the factorial simulation study. For the linear regression on cluster level summaries, the proportions of events in each cluster-period first need to be calculated. Given a data set of the following format and with these variable names:

| `cluster` | `period` | `treatment` | `proportion` | `weights` |
|---|---|---|---|---|
| 1 | 1 | 1 | $p_{11}$ | $w_{11}$ |
| 1 | 2 | 2 | $p_{12}$ | $w_{12}$ |
| 2 | 1 | 2 | $p_{21}$ | $w_{21}$ |
| 2 | 2 | 1 | $p_{22}$ | $w_{22}$ |
| ⋮ | ⋮ | ⋮ | ⋮ | ⋮ |

the Stata command is:

```
regress proportion i.treatment i.cluster i.period
```

Weights, which for example could be based on the size of the cluster, can be added by using the following:

```
regress proportion i.treatment i.cluster i.period [aw=weights]
```

For the hierarchical model with random effects for cluster and cluster-period, a data set with individual level data is needed. Given the following format and variable names:

| cluster | period | cluster_period | treatment | patient | outcome |
|---|---|---|---|---|---|
| 1 | 1 | 1 | 1 | 1 | $Y_{111}$ |
| 1 | 1 | 1 | 1 | 2 | $Y_{112}$ |
| ⋮ | ⋮ | ⋮ | ⋮ | ⋮ | ⋮ |
| 1 | 2 | 2 | 2 | 1 | $Y_{121}$ |
| 1 | 2 | 2 | 2 | 2 | $Y_{122}$ |
| ⋮ | ⋮ | ⋮ | ⋮ | ⋮ | ⋮ |
| 2 | 1 | 3 | 2 | 1 | $Y_{211}$ |
| 2 | 1 | 3 | 2 | 2 | $Y_{212}$ |
| ⋮ | ⋮ | ⋮ | ⋮ | ⋮ | ⋮ |
| 2 | 2 | 4 | 1 | 1 | $Y_{221}$ |
| 2 | 2 | 4 | 1 | 2 | $Y_{222}$ |
| ⋮ | ⋮ | ⋮ | ⋮ | ⋮ | ⋮ |

the Stata comand is:

```
meqrlogit outcome i.period i.treatment || cluster:  || cluster_period:
```

The command `melogit`, which uses a different estimation method, could also be used. Note that in this data set, the variable `cluster_period` contains a different number for each cluster-period.

## 4 Discussion of parameter choices

Here, we present a fuller discussion of the parameters we chose to use in our simulation studies.

**ICCs:** With a binary outcome, care has to be taken over whether ICCs are defined on the logistic or linear scale. This is because for a certain level of correlation the ICCs on these two scales will not be equal, and the difference between them will depend on the event rate of the outcome — see for example Ref. [7]. One way to move between the two scales is to use simulation. A random normal distribution was used to generate 100,000 log odds based on one particular logistic ICC, from which probabilities were then extracted and the corresponding linear ICC calculated. Using this method, a logistic ICC of 0.062 corresponds to a linear ICC of 0.03 for an event rate of 15%. Similarly, an ICC of 0.023 on the logistic scale corresponds to 0.01 on the linear scale for the same event rate. Values of 0.03 and 0.01 on the linear scale were chosen based on work by Eldridge *et al.* on patterns of ICCs in primary care research [8], in which a median ICC of 0.01 and an upper inter-quartile range (IQR) of 0.032 were found. The same ICCs were used for an event rate of 45% for consistency.

**Number of clusters:** In a sample of previously conducted CRXO trials (including at least one pilot study) the median number of clusters was found to be 6 (IQR: 5 to 29) [9–18]. Eldridge *et al.* found the median number of clusters to be 34 in the published cluster randomised trials in primary care in their review [19]. In the factorial set of simulations we therefore chose to look at 6 and 30 clusters.

**Coefficients of variation:** 0.65 is the coefficient of variation found in general practice list size by Eldridge, Ashby and Kerry [20]. 0.65/100 was chosen so that variation between periods within a cluster would be much smaller than variation between clusters — this is because we anticipate that variation within a cluster is likely to be due solely to random fluctuations in many settings.

## 5 Results for 45% event rate

Figures 1 and 2 show the Type 1 error and power for scenarios with an event rate of 45%. The unweighted linear regression gives appropriate Type I errors close to 5% across all scenarios. However, the power is again seen to decrease substantially for ICC combinations with non-zero $\rho_p$ in Figure 2. In the most extreme case, power falls to 23% for 6 clusters and an average of 400 patients per cluster-period. Even for 30 clusters and an average of 75 patients per cluster-period the power falls to around 70% for a $\rho_c$ of 0.062 and a $\rho_p$ of 0.023. Although the size-weighted regression and hierarchical model do not suffer from this loss of power to the same extent, the inflation of the Type I error rates for these methods are generally a bit greater than for an event rate of 15%.

As discussed in the main paper, the failure rates for the Random-random model are higher for a event rate of 45%.

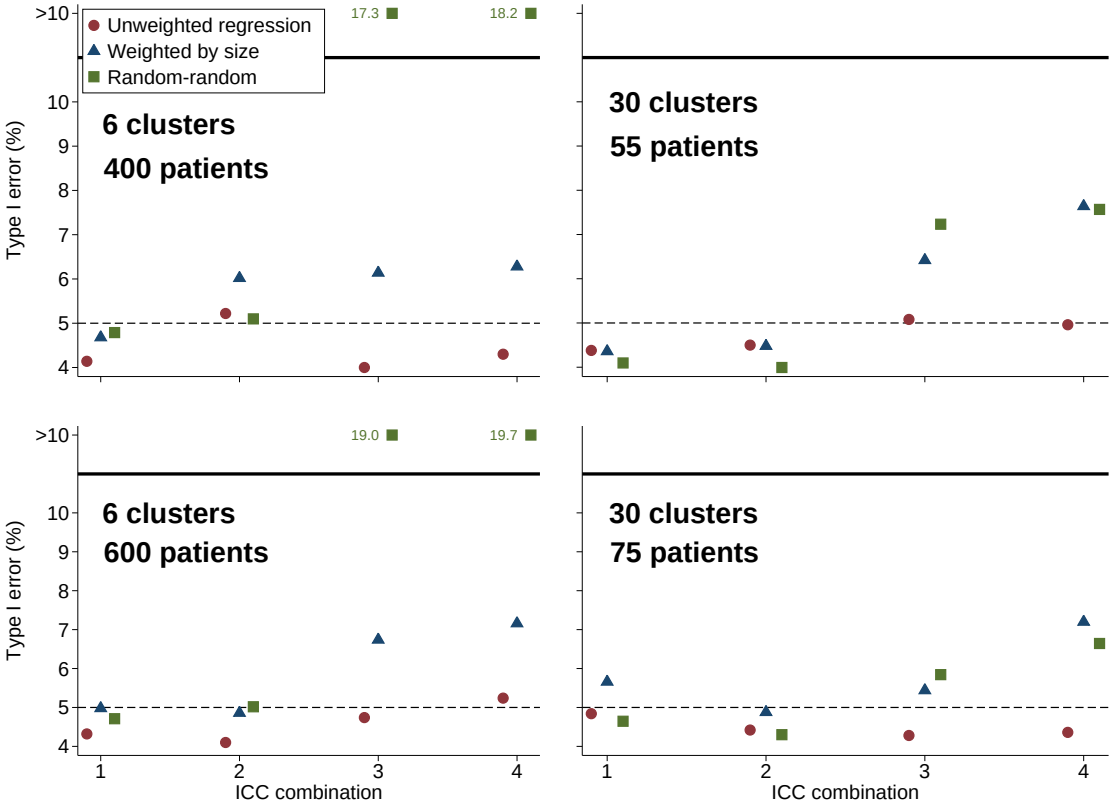


Figure 1: Type I errors across different ICC combinations: combination one is ($\rho_c$=0.023, $\rho_p$=0), combination two is (0.062, 0), combination three is (0.023, 0.01) and combination four is (0.062, 0.023). Graphs are labelled by number of clusters and average number of patients per cluster-period. Other simulation parameters are set to an event rate 45%, no treatment effect, and a fixed period effect OR of 0.92.

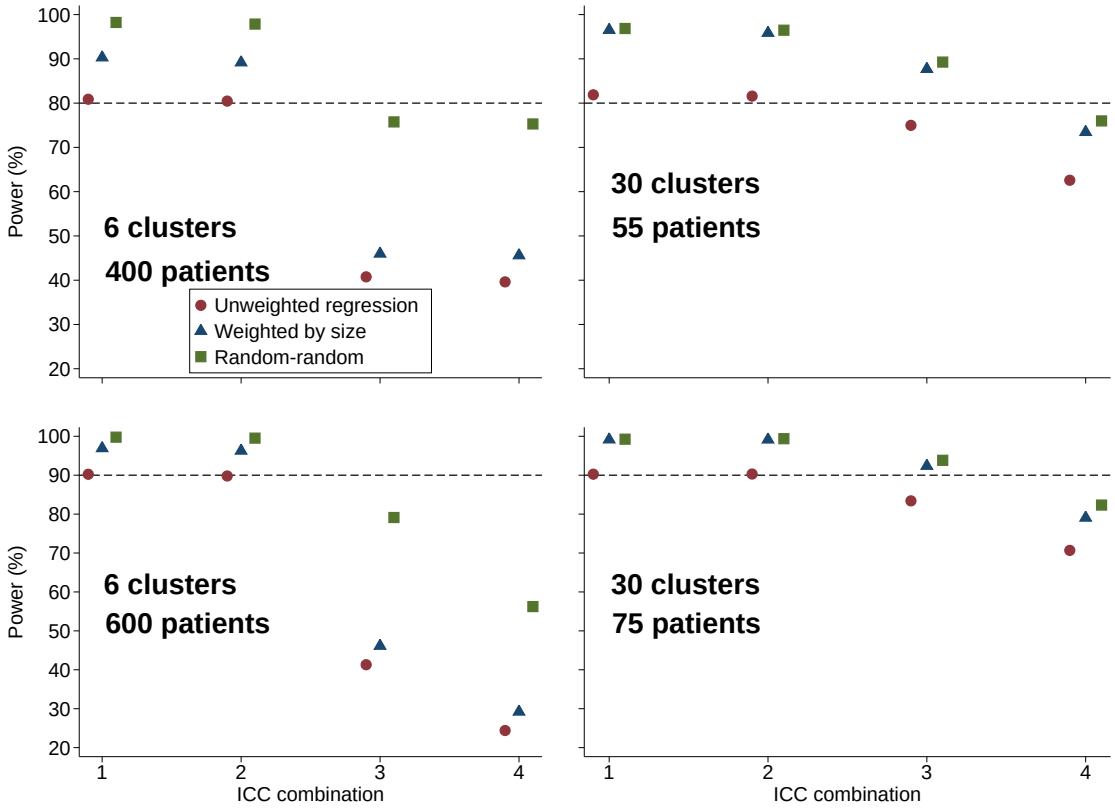


Figure 2: Power across different ICC combinations. Simulation parameters and ICC combinations are as in Figure 1 except that treatment OR is 0.75.

## 6 Tables of simulation results

Tabulated results for the initial simulations are given in the following tables: scenarios that increase the number of clusters while keeping other parameters constant are shown in Tables 1-3; scenarios that increase $\rho_p$, keeping other parameters constant, are shown in Tables 4-7. All initial simulations are for an event rate of 15%. Results from the full factorial simulation study are given in Tables 8-15. Results from the further simulations using the Random-random model are given in Table 16.

Mean treatment estimates for the hierarchical models and GEEs are quoted as odds ratios (ORs), *i.e.* the exponential of the mean of the individual treatment coefficients. The ratio of model based to empirical standard errors (SEs) is given on the logistic scale, *i.e.* the ratio of the square root of the mean of the treatment coefficient variances to the standard deviation (SD) of the treatment coefficients, where the treatment coefficients have not been exponentiated.

We provide the Type I error and power calculated using a t-distribution with $C-2$ degrees of freedom for the linear models. For the hierarchical models and GEEs we provide both the Type I error and power calculated based on a normal distribution (in the column "Type I error/Power with N"), which is the default used in Stata for these models, and a t-distribution with $C-2$ degrees of freedom (in the column "Type I error/Power with t"), to enable direct comparison with the results from the cluster-level summary regressions.

Due to the non-linearity of the logistic link, the treatment difference on the linear scale cannot be calculated directly from the scenario parameters used to generate data on the logistic scale. Averaging across the random effects for cluster and cluster-period from the data generating model on the logistic scale gives zero in expectation, since the random effects are generated from a normal distribution centred on zero. However, when averaging on the linear scale this will not be zero. We

therefore found the expected treatment differences for each scenario through simulation, defining the treatment effect as the difference in proportions when all patients received treatment compared with no patients receiving treatment. We generated data sets with 1 million clusters, and the relevant ICC combinations. The expected treatment differences were found to be:

- $-0.066$ for 15% event rate and ICC combinations ($\rho_c = 0.023, \rho_p = 0$) and $(0.023, 0.01)$,
- $-0.067$ for 15% event rate and ICC combinations $(0.062, 0)$ and $(0.062, 0.023)$,
- $-0.068$ for 45% event rate and ICC combinations $(0.023, 0)$ and $(0.023, 0.01)$,
- $-0.066$ for 45% event rate and ICC combinations $(0.062, 0)$ and $(0.062, 0.023)$.

| Number of clusters | Average number of patients per cluster-period | Analysis method | | Mean treatment estimate | Ratio of model based to empirical SEs | Type I error with N | Type I error with t | Failure rate (out of 5000) |
|---|---|---|---|---|---|---|---|---|
| 6 | 200 | Linear summary models | | | | | | |
| | | Unweighted reg | | -0.0006 | 1.00 | - | 4.2 | 0 |
| | | Weighted by size | | -0.0010 | 0.92 | - | 5.7 | 0 |
| | | Weighted by size and ICC | | -0.0007 | 1.01 | - | 4.0 | 8 |
| | | Weighted by binomial var. | | -0.0009 | 0.94 | - | 5.0 | 0 |
| | | Weighted by Eq. (A1) | | -0.0007 | 1.01 | - | 4.0 | 8 |
| | | Weighted by Eq. (A2) | | -0.0010 | 0.92 | - | 5.7 | 0 |
| | | Weighted by Eq. (A3) | | -0.0007 | 1.01 | - | 4.0 | 8 |
| | | Hierarchical models | | | | | | |
| | | Random-random | | 0.992 | 0.81 | 16.0 | 5.6 | 4 |
| | | Fixed-random | | 0.991 | 0.64 | 22.9 | 9.4 | 24 |
| | | Random | | 0.990 | 0.57 | 26.8 | 12.2 | 1 |
| | | Fixed | | 0.990 | 0.57 | 26.9 | 12.2 | 0 |
| | | Generalised estimating equations | | | | | | |
| | | Exchangeable | robust | 0.991 | 0.77 | 21.9 | 10.9 | 13 |
| | | | conv. | 0.991 | 0.57 | 27.2 | 12.1 | 13 |
| | | Independent | robust | 0.992 | 0.77 | 21.6 | 10.8 | 0 |
| | | | conv. | 0.992 | 0.57 | 27.6 | 12.9 | 0 |
| | | Exch in period | robust | 0.993 | 1.25 | 4.5 | 1.1 | 7 |
| | | | conv. | 0.993 | 1.20 | 5.4 | 1.3 | 7 |
| 12 | 60 | Linear summary models | | | | | | |
| | | Unweighted reg | | -0.0006 | 1.00 | - | 4.4 | 0 |
| | | Weighted by size | | -0.0005 | 0.94 | - | 5.8 | 0 |
| | | Weighted by size and ICC | | -0.0005 | 1.02 | - | 4.6 | 51 |
| | | Weighted by binomial var. | | -0.0005 | 0.96 | - | 5.4 | 0 |
| | | Weighted by Eq. (A1) | | -0.0005 | 1.02 | - | 4.6 | 50 |
| | | Weighted by Eq. (A2) | | -0.0005 | 0.94 | - | 5.8 | 0 |
| | | Weighted by Eq. (A3) | | -0.0005 | 1.02 | - | 4.6 | 51 |
| | | Hierarchical models | | | | | | |
| | | Random-random | | 0.996 | 0.92 | 9.0 | 5.2 | 11 |
| | | Fixed-random | | 0.996 | 0.78 | 13.5 | 8.3 | 313 |
| | | Random | | 0.997 | 0.76 | 14.2 | 9.0 | 1 |
| | | Fixed | | 0.996 | 0.75 | 14.4 | 9.3 | 0 |
| | | Generalised estimating equations | | | | | | |
| | | Exchangeable | robust | 0.997 | 0.90 | 12.1 | 8.2 | 8 |
| | | | conv. | 0.997 | 0.75 | 14.0 | 9.0 | 8 |
| | | Independent | robust | 0.997 | 0.90 | 12.2 | 8.1 | 0 |
| | | | conv. | 0.997 | 0.76 | 13.8 | 9.0 | 0 |
| | | Exch in period | robust | 0.996 | 1.18 | 3.6 | 2.1 | 1 |
| | | | conv. | 0.996 | 1.17 | 4.0 | 2.3 | 1 |

conv.=conventional; var.=variance

Table 1: Mean treatment estimates, ratio of SEs, Type I error and number of failures for a scenario with no treatment effect, $\rho_c = 0.062$ and $\rho_p = 0.023$. The top half of the table is for 6 clusters, 200 patients per period per cluster, and the bottom half is for 12 clusters, 60 patients per cluster-period.

| Number of clusters | Average number of patients per cluster-period | Analysis method | | Mean treatment estimate | Ratio of model based to empirical SEs | Type I error with N | Type I error with t | Failure rate (out of 5000) |
|---|---|---|---|---|---|---|---|---|
| 20 | 34 | Linear summary models | | | | | | |
| | | Unweighted reg | | 0.0004 | 1.00 | - | 4.7 | 0 |
| | | Weighted by size | | 0.0002 | 0.95 | - | 5.9 | 0 |
| | | Weighted by size and ICC | | 0.0001 | 1.01 | - | 5.2 | 84 |
| | | Weighted by binomial var. | | 0.0002 | 0.97 | - | 5.5 | 0 |
| | | Weighted by Eq. (A1) | | 0.0002 | 1.01 | - | 5.2 | 80 |
| | | Weighted by Eq. (A2) | | 0.0002 | 0.95 | - | 5.9 | 0 |
| | | Weighted by Eq. (A3) | | 0.0001 | 1.01 | - | 5.2 | 84 |
| | | Hierarchical models | | | | | | |
| | | Random-random | | 1.002 | 0.96 | 6.7 | 5.1 | 9 |
| | | Random | | 1.002 | 0.82 | 10.4 | 8.3 | 4 |
| | | Fixed | | 1.002 | 0.82 | 10.7 | 8.4 | 0 |
| | | Generalised estimating equations | | | | | | |
| | | Exchangeable | robust | 1.001 | 0.93 | 8.8 | 7.2 | 2 |
| | | | conv. | 1.001 | 0.82 | 10.5 | 8.3 | 2 |
| | | Independent | robust | 1.001 | 0.93 | 8.9 | 7.2 | 0 |
| | | | conv. | 1.001 | 0.83 | 10.4 | 8.3 | 0 |
| | | Exch in period | robust | 1.001 | 1.14 | 3.4 | 2.3 | 0 |
| | | | conv. | 1.001 | 1.13 | 3.6 | 2.5 | 0 |
| 30 | 22 | Linear summary models | | | | | | |
| | | Unweighted reg | | -0.0002 | 1.00 | - | 4.5 | 0 |
| | | Weighted by size | | 0.0001 | 0.97 | - | 6.2 | 0 |
| | | Weighted by size and ICC | | 0.0001 | 1.00 | - | 5.4 | 60 |
| | | Weighted by binomial var. | | 0.0001 | 0.99 | - | 5.6 | 0 |
| | | Weighted by Eq. (A1) | | 0.0000 | 1.00 | - | 5.5 | 58 |
| | | Weighted by Eq. (A2) | | 0.0001 | 0.97 | - | 6.3 | 0 |
| | | Weighted by Eq. (A3) | | 0.0001 | 1.00 | - | 5.4 | 60 |
| | | Hierarchical models | | | | | | |
| | | Random-random | | 1.001 | 0.98 | 6.2 | 5.2 | 14 |
| | | Random | | 1.001 | 0.88 | 8.7 | 7.6 | 13 |
| | | Fixed | | 1.001 | 0.87 | 8.9 | 7.8 | 0 |
| | | Generalised estimating equations | | | | | | |
| | | Exchangeable | robust | 1.001 | 0.97 | 7.9 | 6.6 | 2 |
| | | | conv. | 1.001 | 0.88 | 8.8 | 7.5 | 2 |
| | | Independent | robust | 1.001 | 0.97 | 7.9 | 6.6 | 0 |
| | | | conv. | 1.001 | 0.89 | 8.7 | 7.3 | 0 |
| | | Exch in period | robust | 1.001 | 1.12 | 3.4 | 2.7 | 0 |
| | | | conv. | 1.001 | 1.12 | 3.6 | 3.0 | 0 |

conv.=conventional; var.=variance

Table 2: Mean treatment estimates, ratio of SEs, Type I error and number of failures for a scenario with no treatment effect, $\rho_c = 0.062$ and $\rho_p = 0.023$. The top half of the table is for 20 clusters, 34 patients per period per cluster, and the bottom half is for 30 clusters, 22 patients per cluster-period.

| Number of clusters | Average number of patients per cluster-period | Analysis method | | Mean treatment estimate | Ratio of model based to empirical SEs | Type I error with N | Type I error with t | Failure rate (out of 5000) |
|---|---|---|---|---|---|---|---|---|
| 50 | 14 | Linear summary models | | | | | | |
| | | Unweighted reg | | -0.0003 | 1.00 | - | 4.9 | 0 |
| | | Weighted by size | | -0.0002 | 0.96 | - | 5.9 | 0 |
| | | Weighted by size and ICC | | -0.0003 | 0.99 | - | 5.6 | 55 |
| | | Weighted by binomial var. | | -0.0003 | 1.00 | - | 4.7 | 0 |
| | | Weighted by Eq. (A1) | | -0.0003 | 0.99 | - | 5.6 | 54 |
| | | Weighted by Eq. (A2) | | -0.0002 | 0.96 | - | 5.9 | 0 |
| | | Weighted by Eq. (A3) | | -0.0003 | 0.99 | - | 5.6 | 55 |
| | | Hierarchical models | | | | | | |
| | | Random-random | | 0.999 | 0.98 | 5.9 | 5.4 | 14 |
| | | Random | | 0.998 | 0.91 | 7.6 | 7.1 | 5 |
| | | Fixed | | 0.999 | 0.89 | 8.3 | 7.4 | 0 |
| | | Generalised estimating equations | | | | | | |
| | | Exchangeable | robust | 0.998 | 0.97 | 6.6 | 6.1 | 0 |
| | | | conv. | 0.998 | 0.90 | 7.7 | 7.0 | 0 |
| | | Independent | robust | 0.998 | 0.97 | 6.7 | 6.1 | 0 |
| | | | conv. | 0.998 | 0.91 | 7.4 | 6.7 | 0 |
| | | Exch in period | robust | 0.999 | 1.08 | 3.8 | 3.2 | 0 |
| | | | conv. | 0.999 | 1.08 | 3.9 | 3.3 | 0 |
| 80 | 8 | Linear summary models | | | | | | |
| | | Unweighted reg | | 0.0002 | 1.00 | - | 4.9 | 0 |
| | | Weighted by size | | 0.0003 | 0.97 | - | 5.6 | 0 |
| | | Weighted by size and ICC | | 0.0003 | 0.98 | - | 5.4 | 31 |
| | | Weighted by binomial var. | | 0.0003 | 1.04 | - | 4.0 | 0 |
| | | Weighted by Eq. (A1) | | 0.0003 | 0.98 | - | 5.4 | 31 |
| | | Weighted by Eq. (A2) | | 0.0003 | 0.97 | - | 5.6 | 0 |
| | | Weighted by Eq. (A3) | | 0.0003 | 0.98 | - | 5.4 | 31 |
| | | Hierarchical models | | | | | | |
| | | Random-random | | 1.002 | 0.99 | 5.4 | 5.0 | 4 |
| | | Random | | 1.002 | 0.94 | 6.4 | 6.0 | 0 |
| | | Fixed | | 1.002 | 0.90 | 7.5 | 7.1 | 0 |
| | | Generalised estimating equations | | | | | | |
| | | Exchangeable | robust | 1.002 | 0.97 | 6.1 | 5.7 | 0 |
| | | | conv. | 1.002 | 0.94 | 6.5 | 6.0 | 0 |
| | | Independent | robust | 1.002 | 0.97 | 6.0 | 5.6 | 0 |
| | | | conv. | 1.002 | 0.94 | 6.2 | 5.6 | 0 |
| | | Exch in period | robust | 1.002 | 1.04 | 4.2 | 3.9 | 0 |
| | | | conv. | 1.002 | 1.04 | 4.1 | 3.8 | 0 |

conv.=conventional; var.=variance

Table 3: Mean treatment estimates, ratio of SEs, Type I error and number of failures for a scenario with no treatment effect, $\rho_c = 0.062$ and $\rho_p = 0.023$. The top half of the table is for 50 clusters, 14 patients per period per cluster, and the bottom half is for 80 clusters, 8 patients per cluster-period.

| $\rho_p$ | Analysis method | | Mean treatment estimate | Ratio of model based to empirical SEs | Type I error with N | Type I error with t | Failure rate (out of 5000) |
|---|---|---|---|---|---|---|---|
| 0.001 | Linear summary models | | | | | | |
| | Unweighted reg | | -0.0001 | 0.99 | - | 4.3 | 0 |
| | Weighted by size | | -0.0000 | 0.98 | - | 4.7 | 0 |
| | Weighted by size and ICC | | -0.0000 | 1.04 | - | 4.1 | 33 |
| | Weighted by binomial variance | | -0.0001 | 0.98 | - | 5.2 | 0 |
| | Weighted by Eq. (A1) | | -0.0000 | 1.04 | - | 4.1 | 33 |
| | Weighted by Eq. (A2) | | -0.0000 | 0.98 | - | 4.7 | 0 |
| | Weighted by Eq. (A3) | | -0.0000 | 1.04 | - | 4.1 | 33 |
| | Hierarchical models | | | | | | |
| | Random-random | | 0.999 | 0.99 | 5.7 | 0.8 | 6 |
| | Random-fixed | | 0.999 | 0.94 | 6.5 | 1.0 | 8 |
| | Fixed-fixed | | 0.999 | 0.94 | 6.6 | 0.9 | 0 |
| | Generalised estimating equations | | | | | | |
| | Exchangeable | robust | 1.000 | 0.81 | 19.0 | 9.8 | 10 |
| | | conventional | 1.000 | 0.94 | 6.2 | 1.0 | 10 |
| | Independent | robust | 1.000 | 0.82 | 19.1 | 9.2 | 0 |
| | | conventional | 1.000 | 0.93 | 6.5 | 1.2 | 0 |
| | Exch in period | robust | 1.000 | 1.96 | 0.8 | 0.3 | 14 |
| | | conventional | 1.000 | 1.89 | 1.2 | 0.2 | 14 |
| 0.005 | Linear summary models | | | | | | |
| | Unweighted reg | | -0.0001 | 1.00 | - | 4.7 | 0 |
| | Weighted by size | | 0.0000 | 0.97 | - | 5.8 | 0 |
| | Weighted by size and ICC | | -0.0000 | 1.04 | - | 4.8 | 24 |
| | Weighted by binomial variance | | 0.0001 | 0.98 | - | 5.7 | 0 |
| | Weighted by Eq. (A1) | | -0.0000 | 1.04 | - | 4.8 | 24 |
| | Weighted by Eq. (A2) | | 0.0000 | 0.97 | - | 5.8 | 0 |
| | Weighted by Eq. (A3) | | -0.0000 | 1.04 | - | 4.8 | 24 |
| | Hierarchical models | | | | | | |
| | Random-random | | 1.001 | 0.93 | 8.6 | 2.0 | 5 |
| | Random-fixed | | 1.001 | 0.84 | 10.5 | 2.6 | 3 |
| | Fixed-fixed | | 1.002 | 0.84 | 10.4 | 2.6 | 0 |
| | Generalised estimating equations | | | | | | |
| | Exchangeable | robust | 1.001 | 0.81 | 20.2 | 10.8 | 14 |
| | | conventional | 1.001 | 0.84 | 10.8 | 2.4 | 14 |
| | Independent | robust | 1.001 | 0.81 | 20.0 | 10.3 | 0 |
| | | conventional | 1.001 | 0.83 | 11.1 | 2.6 | 0 |
| | Exch in period | robust | 1.000 | 1.77 | 1.5 | 0.5 | 16 |
| | | conventional | 1.000 | 1.72 | 2.1 | 0.5 | 16 |

Table 4: Mean treatment estimates, ratio of SEs, Type I error and number of failures for a scenario with no treatment effect, 6 clusters, 200 patients per period per cluster, and $\rho_c = 0.062$. The top half of the table is for $\rho_p = 0.001$, and the bottom half is for $\rho_p = 0.005$.

| $\rho_p$ | Analysis method | | Mean treatment estimate | Ratio of model based to empirical SEs | Type I error with N | Type I error with t | Failure rate (out of 5000) |
|---|---|---|---|---|---|---|---|
| 0.01 | Linear summary models | | | | | | |
| | Unweighted reg | | -0.0000 | 1.02 | - | 4.5 | 0 |
| | Weighted by size | | -0.0001 | 0.94 | - | 5.6 | 0 |
| | Weighted by size and ICC | | 0.0000 | 1.03 | - | 4.5 | 20 |
| | Weighted by binomial variance | | -0.0001 | 0.96 | - | 5.5 | 0 |
| | Weighted by Eq. (A1) | | 0.0000 | 1.03 | - | 4.5 | 20 |
| | Weighted by Eq. (A2) | | -0.0001 | 0.94 | - | 5.6 | 0 |
| | Weighted by Eq. (A3) | | 0.0000 | 1.03 | - | 4.5 | 20 |
| | Hierarchical models | | | | | | |
| | Random-random | | 1.000 | 0.87 | 11.7 | 3.0 | 7 |
| | Random-fixed | | 0.999 | 0.73 | 16.0 | 4.7 | 2 |
| | Fixed-fixed | | 0.999 | 0.73 | 16.3 | 4.7 | 0 |
| | Generalised estimating equations | | | | | | |
| | Exchangeable | robust | 0.999 | 0.79 | 21.4 | 10.3 | 16 |
| | | conventional | 0.999 | 0.73 | 16.0 | 4.7 | 16 |
| | Independent | robust | 0.999 | 0.80 | 21.2 | 10.2 | 0 |
| | | conventional | 0.999 | 0.72 | 17.0 | 5.1 | 0 |
| | Exch in period | robust | 1.000 | 1.58 | 2.3 | 0.6 | 9 |
| | | conventional | 1.000 | 1.53 | 2.9 | 0.8 | 9 |
| 0.05 | Linear summary models | | | | | | |
| | Unweighted reg | | -0.0000 | 0.98 | - | 5.0 | 0 |
| | Weighted by size | | 0.0004 | 0.88 | - | 6.9 | 0 |
| | Weighted by size and ICC | | 0.0001 | 0.99 | - | 5.1 | 4 |
| | Weighted by binomial variance | | 0.0003 | 0.90 | - | 6.5 | 0 |
| | Weighted by Eq. (A1) | | 0.0001 | 0.99 | - | 5.1 | 4 |
| | Weighted by Eq. (A2) | | 0.0004 | 0.88 | - | 6.9 | 0 |
| | Weighted by Eq. (A3) | | 0.0001 | 0.99 | - | 5.1 | 4 |
| | Hierarchical models | | | | | | |
| | Random-random | | 1.001 | 0.75 | 20.0 | 9.0 | 9 |
| | Random-fixed | | 1.003 | 0.42 | 41.8 | 25.8 | 1 |
| | Fixed-fixed | | 1.003 | 0.42 | 42.1 | 25.9 | 0 |
| | Generalised estimating equations | | | | | | |
| | Exchangeable | robust | 1.004 | 0.74 | 23.3 | 12.0 | 31 |
| | | conventional | 1.004 | 0.42 | 41.9 | 25.7 | 31 |
| | Independent | robust | 1.004 | 0.74 | 23.1 | 12.1 | 0 |
| | | conventional | 1.004 | 0.41 | 42.1 | 26.3 | 0 |
| | Exch in period | robust | 1.003 | 0.93 | 10.3 | 3.3 | 7 |
| | | conventional | 1.003 | 0.89 | 12.0 | 4.1 | 7 |

Table 5: Mean treatment estimates, ratio of SEs, Type I error and number of failures for a scenario with no treatment effect, 6 clusters, 200 patients per period per cluster, and $\rho_c = 0.062$. The top half is for $\rho_p = 0.01$, and the bottom half is for $\rho_p = 0.05$.

| $\rho_p$ | Analysis method | | Mean treatment estimate | Ratio of model based to empirical SEs | Type I error with N | Type I error with t | Failure rate (out of 5000) |
|---|---|---|---|---|---|---|---|
| 0.001 | Linear summary models | | | | | | |
| | Unweighted reg | | 0.0003 | 0.99 | - | 4.5 | 0 |
| | Weighted by size | | 0.0004 | 0.98 | - | 5.2 | 0 |
| | Weighted by size and ICC | | 0.0003 | 0.98 | - | 5.3 | 208 |
| | Weighted by binomial variance | | 0.0003 | 0.99 | - | 4.7 | 0 |
| | Weighted by Eq. (A1) | | 0.0003 | 0.98 | - | 5.3 | 207 |
| | Weighted by Eq. (A2) | | 0.0004 | 0.98 | - | 5.2 | 0 |
| | Weighted by Eq. (A3) | | 0.0003 | 0.98 | - | 5.4 | 208 |
| | Hierarchical models | | | | | | |
| | Random-random | | 1.003 | 1.01 | 4.5 | 3.8 | 5 |
| | Random-fixed | | 1.003 | 0.97 | 5.3 | 4.5 | 6 |
| | Fixed-fixed | | 1.004 | 0.96 | 5.7 | 4.8 | 0 |
| | Generalised estimating equations | | | | | | |
| | Exchangeable | robust | 1.003 | 0.95 | 7.7 | 6.5 | 1 |
| | | conventional | 1.003 | 0.97 | 5.3 | 4.4 | 1 |
| | Independent | robust | 1.003 | 0.95 | 7.5 | 6.3 | 0 |
| | | conventional | 1.003 | 0.98 | 5.1 | 4.2 | 0 |
| | Exch in period | robust | 1.003 | 1.22 | 1.9 | 1.5 | 0 |
| | | conventional | 1.003 | 1.21 | 1.9 | 1.4 | 0 |
| 0.005 | Linear summary models | | | | | | |
| | Unweighted reg | | 0.0005 | 1.00 | - | 4.4 | 0 |
| | Weighted by size | | 0.0003 | 0.99 | - | 5.1 | 0 |
| | Weighted by size and ICC | | 0.0002 | 0.98 | - | 5.7 | 205 |
| | Weighted by binomial variance | | 0.0004 | 1.01 | - | 4.9 | 0 |
| | Weighted by Eq. (A1) | | 0.0002 | 0.98 | - | 5.7 | 205 |
| | Weighted by Eq. (A2) | | 0.0003 | 0.99 | - | 5.1 | 0 |
| | Weighted by Eq. (A3) | | 0.0002 | 0.98 | - | 5.7 | 205 |
| | Hierarchical models | | | | | | |
| | Random-random | | 1.003 | 1.02 | 4.7 | 4.0 | 5 |
| | Random-fixed | | 1.003 | 0.97 | 5.3 | 4.8 | 7 |
| | Fixed-fixed | | 1.003 | 0.95 | 5.8 | 5.0 | 0 |
| | Generalised estimating equations | | | | | | |
| | Exchangeable | robust | 1.003 | 0.96 | 7.3 | 6.2 | 0 |
| | | conventional | 1.003 | 0.97 | 5.4 | 4.6 | 0 |
| | Independent | robust | 1.003 | 0.96 | 7.3 | 6.1 | 0 |
| | | conventional | 1.003 | 0.98 | 5.3 | 4.5 | 0 |
| | Exch in period | robust | 1.002 | 1.21 | 2.1 | 1.6 | 0 |
| | | conventional | 1.002 | 1.21 | 2.1 | 1.6 | 0 |

Table 6: Mean treatment estimates, ratio of SEs, Type I error and number of failures for a scenario with no treatment effect, 30 clusters, 22 patients per period per cluster, and $\rho_c = 0.062$. The top half is for $\rho_p = 0.001$, and the bottom half is for $\rho_p = 0.005$.

| $\rho_p$ | Analysis method | | Mean treatment estimate | Ratio of model based to empirical SEs | Type I error with N | Type I error with t | Failure rate (out of 5000) |
|---|---|---|---|---|---|---|---|
| 0.01 | Linear summary models | | | | | | |
| | Unweighted reg | | -0.0003 | 1.01 | - | 4.3 | 0 |
| | Weighted by size | | -0.0000 | 1.00 | - | 5.3 | 0 |
| | Weighted by size and ICC | | 0.0001 | 1.01 | - | 5.2 | 143 |
| | Weighted by binomial variance | | -0.0001 | 1.02 | - | 5.1 | 0 |
| | Weighted by Eq. (A1) | | 0.0001 | 1.01 | - | 5.1 | 142 |
| | Weighted by Eq. (A2) | | -0.0000 | 1.00 | - | 5.3 | 0 |
| | Weighted by Eq. (A3) | | 0.0001 | 1.01 | - | 5.1 | 143 |
| | Hierarchical models | | | | | | |
| | Random-random | | 0.999 | 1.01 | 4.9 | 3.9 | 12 |
| | Random-fixed | | 1.000 | 0.95 | 6.2 | 5.0 | 8 |
| | Fixed-fixed | | 0.999 | 0.93 | 6.7 | 5.3 | 0 |
| | Generalised estimating equations | | | | | | |
| | Exchangeable | robust | 1.000 | 0.97 | 7.0 | 5.9 | 0 |
| | | conventional | 1.000 | 0.95 | 6.1 | 5.0 | 0 |
| | Independent | robust | 1.000 | 0.97 | 7.0 | 5.9 | 0 |
| | | conventional | 1.000 | 0.96 | 5.9 | 4.8 | 0 |
| | Exch in period | robust | 0.999 | 1.20 | 2.3 | 1.7 | 0 |
| | | conventional | 0.999 | 1.19 | 2.2 | 1.7 | 0 |
| 0.05 | Linear summary models | | | | | | |
| | Unweighted reg | | -0.0004 | 0.98 | - | 5.0 | 0 |
| | Weighted by size | | -0.0008 | 0.93 | - | 7.2 | 0 |
| | Weighted by size and ICC | | -0.0007 | 0.99 | - | 5.8 | 14 |
| | Weighted by binomial variance | | -0.0007 | 0.96 | - | 5.9 | 0 |
| | Weighted by Eq. (A1) | | -0.0007 | 0.99 | - | 5.8 | 14 |
| | Weighted by Eq. (A2) | | -0.0008 | 0.93 | - | 7.2 | 0 |
| | Weighted by Eq. (A3) | | -0.0007 | 0.99 | - | 5.8 | 14 |
| | Hierarchical models | | | | | | |
| | Random-random | | 0.994 | 0.93 | 8.0 | 6.9 | 13 |
| | Random-fixed | | 0.994 | 0.77 | 12.9 | 11.4 | 10 |
| | Fixed-fixed | | 0.993 | 0.76 | 13.3 | 11.8 | 0 |
| | Generalised estimating equations | | | | | | |
| | Exchangeable | robust | 0.993 | 0.94 | 8.2 | 7.3 | 4 |
| | | conventional | 0.993 | 0.77 | 13.0 | 11.4 | 4 |
| | Independent | robust | 0.994 | 0.94 | 8.3 | 7.2 | 0 |
| | | conventional | 0.994 | 0.78 | 12.8 | 11.4 | 0 |
| | Exch in period | robust | 0.994 | 0.99 | 5.9 | 4.9 | 0 |
| | | conventional | 0.994 | 0.99 | 6.1 | 4.9 | 0 |

Table 7: Mean treatment estimates, ratio of SEs, Type I error and number of failures for a scenario with no treatment effect, 30 clusters, 22 patients per period per cluster, and $\rho_c = 0.062$. The top half of the table is for $\rho_p = 0.01$, and the bottom half is for $\rho_p = 0.05$.

| $\rho_c$ | $\rho_p$ | Mean treatment estimate | Ratio of SEs | Type I error with N | Type I error with t | Failure rate (out of 5000) | Mean treatment estimate | Ratio of SEs | Power with N | Power with t | Failure rate (out of 5000) |
|---|---|---|---|---|---|---|---|---|---|---|---|
| Unweighted regression | | | | | | | | | | | |
| 0.023 | 0 | 0.0003 | 1.01 | - | 4.3 | 0 | -0.0655 | 1.00 | - | 82.3 | 0 |
| 0.062 | 0 | 0.0002 | 0.98 | - | 4.4 | 0 | -0.0677 | 0.99 | - | 79.4 | 0 |
| 0.023 | 0.01 | -0.0003 | 1.01 | - | 4.3 | 0 | -0.0663 | 1.01 | - | 71.7 | 0 |
| 0.062 | 0.023 | -0.0006 | 1.00 | - | 4.2 | 0 | -0.0674 | 1.00 | - | 54.5 | 0 |
| Size-weighted regression | | | | | | | | | | | |
| 0.023 | 0 | 0.0003 | 1.01 | - | 5.1 | 0 | -0.0653 | 0.99 | - | 90.7 | 0 |
| 0.062 | 0 | 0.0002 | 0.97 | - | 5.0 | 0 | -0.0676 | 0.95 | - | 87.3 | 0 |
| 0.023 | 0.01 | -0.0002 | 0.95 | - | 5.7 | 0 | -0.0661 | 0.95 | - | 80.0 | 0 |
| 0.062 | 0.023 | -0.0010 | 0.92 | - | 5.7 | 0 | -0.0674 | 0.91 | - | 61.8 | 0 |
| Random-random model | | | | | | | | | | | |
| 0.023 | 0 | 1.002 | 1.04 | 4.2 | 0.4 | 10 | 0.501 | 1.04 | 98.9 | 94.1 | 6 |
| 0.062 | 0 | 1.002 | 1.01 | 5.2 | 0.4 | 4 | 0.497 | 1.02 | 98.7 | 94.2 | 4 |
| 0.023 | 0.01 | 0.998 | 0.86 | 11.2 | 3.0 | 10 | 0.497 | 0.88 | 96.9 | 88.3 | 14 |
| 0.062 | 0.023 | 0.992 | 0.81 | 16.0 | 5.6 | 4 | 0.500 | 0.83 | 90.3 | 74.9 | 3 |

Table 8: Mean treatment estimates, ratio of SEs, Type I error or power, and number of failures for a 15% event rate, 6 clusters, 200 patients per period per cluster. Results on the left are for a scenario with no treatment effect, and those on the right are for a scenario with treatment OR 0.5.

| $\rho_c$ | $\rho_p$ | Mean treatment estimate | Ratio of SEs | Type I error with N | Type I error with t | Failure rate (out of 5000) | Mean treatment estimate | Ratio of SEs | Power with N | Power with t | Failure rate (out of 5000) |
|---|---|---|---|---|---|---|---|---|---|---|---|
| Unweighted regression | | | | | | | | | | | |
| 0.023 | 0 | 0.0000 | 0.99 | - | 4.3 | 0 | -0.0655 | 1.00 | - | 92.6 | 0 |
| 0.062 | 0 | 0.0000 | 1.01 | - | 4.0 | 0 | -0.0672 | 1.01 | - | 89.2 | 0 |
| 0.023 | 0.01 | -0.0000 | 1.01 | - | 4.2 | 0 | -0.0656 | 1.00 | - | 82.5 | 0 |
| 0.062 | 0.023 | 0.0001 | 1.02 | - | 4.3 | 0 | -0.0673 | 1.02 | - | 63.1 | 0 |
| Size-weighted regression | | | | | | | | | | | |
| 0.023 | 0 | -0.0001 | 0.98 | - | 5.2 | 0 | -0.0655 | 0.97 | - | 97.8 | 0 |
| 0.062 | 0 | 0.0000 | 1.00 | - | 4.7 | 0 | -0.0673 | 0.94 | - | 94.3 | 0 |
| 0.023 | 0.01 | -0.0001 | 0.94 | - | 5.6 | 0 | -0.0657 | 0.93 | - | 88.7 | 0 |
| 0.062 | 0.023 | 0.0001 | 0.92 | - | 6.1 | 0 | -0.0673 | 0.92 | - | 69.3 | 0 |
| Random-random model | | | | | | | | | | | |
| 0.023 | 0 | 1.000 | 1.03 | 4.6 | 0.4 | 9 | 0.500 | 1.04 | 100.0 | 99.3 | 1 |
| 0.062 | 0 | 1.000 | 1.04 | 4.6 | 0.5 | 9 | 0.501 | 1.05 | 99.8 | 99.0 | 3 |
| 0.023 | 0.01 | 0.999 | 0.83 | 13.9 | 4.9 | 11 | 0.500 | 0.84 | 99.1 | 95.1 | 15 |
| 0.062 | 0.023 | 1.002 | 0.81 | 16.7 | 7.1 | 8 | 0.499 | 0.82 | 95.3 | 84.5 | 6 |

Table 9: Mean treatment estimates, ratio of SEs, Type I error or power, and number of failures for a 15% event rate, 6 clusters, 330 patients per period per cluster. Results on the left are for a scenario with no treatment effect, and those on the right are for a scenario with treatment OR 0.5.

| $\rho_c$ | $\rho_p$ | Mean treatment estimate | Ratio of SEs | Type I error with N | Type I error with t | Failure rate (out of 5000) | Mean treatment estimate | Ratio of SEs | Power with N | Power with t of 5000) | Failure rate (out |
|---|---|---|---|---|---|---|---|---|---|---|---|
| Unweighted regression | | | | | | | | | | | |
| 0.023 | 0 | 0.0002 | 1.00 | - | 4.3 | 0 | -0.0660 | 1.01 | - | 79.3 | 0 |
| 0.062 | 0 | -0.0001 | 0.99 | - | 4.5 | 0 | -0.0677 | 1.00 | - | 80.4 | 0 |
| 0.023 | 0.01 | -0.0001 | 1.00 | - | 4.6 | 0 | -0.0658 | 0.99 | - | 78.3 | 0 |
| 0.062 | 0.023 | -0.0002 | 1.00 | - | 4.5 | 0 | -0.0676 | 1.02 | - | 76.3 | 0 |
| Size-weighted regression | | | | | | | | | | | |
| 0.023 | 0 | 0.0003 | 1.00 | - | 5.2 | 0 | -0.0661 | 0.99 | - | 95.1 | 0 |
| 0.062 | 0 | -0.0000 | 1.00 | - | 4.7 | 0 | -0.0675 | 0.97 | - | 94.6 | 0 |
| 0.023 | 0.01 | -0.0000 | 0.98 | - | 5.3 | 0 | -0.0657 | 0.98 | - | 93.3 | 0 |
| 0.062 | 0.023 | 0.0001 | 0.97 | - | 6.2 | 0 | -0.0676 | 0.96 | - | 91.1 | 0 |
| Random-random model | | | | | | | | | | | |
| 0.023 | 0 | 1.003 | 1.03 | 4.1 | 3.2 | 10 | 0.495 | 1.02 | 95.9 | 94.8 | 11 |
| 0.062 | 0 | 1.000 | 1.03 | 4.2 | 3.3 | 8 | 0.497 | 1.02 | 95.2 | 94.2 | 6 |
| 0.023 | 0.01 | 1.000 | 0.99 | 5.5 | 4.3 | 11 | 0.497 | 0.98 | 94.6 | 93.8 | 14 |
| 0.062 | 0.023 | 1.001 | 0.98 | 6.2 | 5.2 | 14 | 0.497 | 0.97 | 92.9 | 91.4 | 7 |

Table 10: Mean treatment estimates, ratio of SEs, Type I error or power, and number of failures for a 15% event rate, 30 clusters, 22 patients per period per cluster. Results on the left are for a scenario with no treatment effect, and those on the right are for a scenario with treatment OR 0.5.

| $\rho_c$ | $\rho_p$ | Mean treatment estimate | Ratio of SEs | Type I error with N | Type I error with t | Failure rate (out of 5000) | Mean treatment estimate | Ratio of SEs | Power with N | Power with t | Failure rate (out of 5000) |
|---|---|---|---|---|---|---|---|---|---|---|---|
| Unweighted regression | | | | | | | | | | | |
| 0.023 | 0 | 0.0000 | 1.00 | - | 4.7 | 0 | -0.0655 | 1.02 | - | 89.9 | 0 |
| 0.062 | 0 | 0.0002 | 0.98 | - | 5.3 | 0 | -0.0681 | 1.00 | - | 90.8 | 0 |
| 0.023 | 0.01 | -0.0003 | 1.01 | - | 4.3 | 0 | -0.0660 | 0.99 | - | 88.6 | 0 |
| 0.062 | 0.023 | -0.0001 | 1.00 | - | 5.2 | 0 | -0.0675 | 1.01 | - | 87.1 | 0 |
| Size-weighted regression | | | | | | | | | | | |
| 0.023 | 0 | 0.0001 | 1.00 | - | 5.2 | 0 | -0.0658 | 1.00 | - | 99.0 | 0 |
| 0.062 | 0 | 0.0001 | 0.98 | - | 5.9 | 0 | -0.0678 | 0.99 | - | 99.0 | 0 |
| 0.023 | 0.01 | -0.0003 | 0.97 | - | 5.6 | 0 | -0.0659 | 0.96 | - | 98.2 | 0 |
| 0.062 | 0.023 | -0.0000 | 0.95 | - | 6.5 | 0 | -0.0673 | 0.97 | - | 96.6 | 0 |
| Random-random model | | | | | | | | | | | |
| 0.023 | 0 | 1.001 | 1.03 | 4.6 | 3.6 | 11 | 0.498 | 1.03 | 99.2 | 98.8 | 12 |
| 0.062 | 0 | 1.001 | 1.01 | 5.1 | 4.5 | 9 | 0.496 | 1.05 | 99.2 | 98.9 | 7 |
| 0.023 | 0.01 | 0.998 | 0.98 | 5.9 | 4.8 | 11 | 0.497 | 0.98 | 98.6 | 98.4 | 10 |
| 0.062 | 0.023 | 1.000 | 0.96 | 7.0 | 6.0 | 2 | 0.499 | 0.99 | 97.3 | 96.7 | 8 |

Table 11: Mean treatment estimates, ratio of SEs, Type I error or power, and number of failures for a 15% event rate, 30 clusters, 31 patients per period per cluster. Results on the left are for a scenario with no treatment effect, and those on the right are for a scenario with treatment OR 0.5.

| $\rho_c$ | $\rho_p$ | Mean treatment estimate | Ratio of SEs | Type I error with N | Type I error with t | Failure rate (out of 5000) | Mean treatment estimate | Ratio of SEs | Power with N | Power with t | Failure rate (out of 5000) |
|---|---|---|---|---|---|---|---|---|---|---|---|
| Unweighted regression | | | | | | | | | | | |
| 0.023 | 0 | 0.0000 | 1.03 | - | 4.1 | 0 | -0.0679 | 1.00 | - | 80.9 | 0 |
| 0.062 | 0 | 0.0004 | 1.00 | - | 5.2 | 0 | -0.0668 | 0.99 | - | 80.5 | 0 |
| 0.023 | 0.01 | 0.0002 | 1.00 | - | 4.0 | 0 | -0.0681 | 1.02 | - | 40.8 | 0 |
| 0.062 | 0.023 | 0.0008 | 1.02 | - | 4.3 | 0 | -0.0668 | 1.00 | - | 22.8 | 0 |
| Size-weighted regression | | | | | | | | | | | |
| 0.023 | 0 | 0.0002 | 1.00 | - | 4.7 | 0 | -0.0681 | 1.00 | - | 90.3 | 0 |
| 0.062 | 0 | 0.0003 | 0.98 | - | 6.0 | 0 | -0.0665 | 1.00 | - | 89.2 | 0 |
| 0.023 | 0.01 | -0.0000 | 0.90 | - | 6.1 | 0 | -0.0685 | 0.91 | - | 46.0 | 0 |
| 0.062 | 0.023 | 0.0007 | 0.90 | - | 6.3 | 0 | -0.0670 | 0.88 | - | 27.7 | 0 |
| Random-random model | | | | | | | | | | | |
| 0.023 | 0 | 1.001 | 1.04 | 4.8 | 0.5 | 72 | 0.750 | 1.04 | 98.2 | 91.7 | 50 |
| 0.062 | 0 | 1.001 | 1.02 | 5.1 | 0.5 | 116 | 0.749 | 1.04 | 97.9 | 91.9 | 89 |
| 0.023 | 0.01 | 1.000 | 0.80 | 17.3 | 6.5 | 126 | 0.749 | 0.80 | 75.8 | 56.1 | 79 |
| 0.062 | 0.023 | 1.003 | 0.81 | 18.2 | 8.0 | 211 | 0.748 | 0.80 | 54.7 | 52.2 | 148 |

Table 12: Mean treatment estimates, ratio of SEs, Type I error or power, and number of failures for a 45% event rate, 6 clusters, 400 patients per period per cluster. Results on the left are for a scenario with no treatment effect, and those on the right are for a scenario with treatment OR 0.75.

| $\rho_c$ | $\rho_p$ | Mean treatment estimate | Ratio of SEs | Type I error with N | Type I error with t | Failure rate (out of 5000) | Mean treatment estimate | Ratio of SEs | Power with N | Power with t | Failure rate (out of 5000) |
|---|---|---|---|---|---|---|---|---|---|---|---|
| Unweighted regression | | | | | | | | | | | |
| 0.023 | 0 | 0.0001 | 0.99 | - | 4.3 | 0 | -0.0679 | 1.01 | - | 90.2 | 0 |
| 0.062 | 0 | -0.0001 | 1.00 | - | 4.1 | 0 | -0.0661 | 1.00 | - | 89.8 | 0 |
| 0.023 | 0.01 | 0.0002 | 1.00 | - | 4.7 | 0 | -0.0678 | 1.01 | - | 41.3 | 0 |
| 0.062 | 0.023 | -0.0007 | 1.00 | - | 5.2 | 0 | -0.0666 | 1.00 | - | 24.4 | 0 |
| Size-weighted regression | | | | | | | | | | | |
| 0.023 | 0 | 0.0002 | 0.99 | - | 5.0 | 0 | -0.0681 | 1.01 | - | 96.9 | 0 |
| 0.062 | 0 | -0.0002 | 1.00 | - | 4.9 | 0 | -0.0659 | 1.00 | - | 96.3 | 0 |
| 0.023 | 0.01 | 0.0000 | 0.90 | - | 6.7 | 0 | -0.0674 | 0.91 | - | 46.1 | 0 |
| 0.062 | 0.023 | -0.0003 | 0.90 | - | 7.2 | 0 | -0.0665 | 0.89 | - | 29.2 | 0 |
| Random-random model | | | | | | | | | | | |
| 0.023 | 0 | 1.001 | 1.03 | 4.7 | 0.4 | 160 | 0.750 | 1.05 | 99.8 | 97.9 | 122 |
| 0.062 | 0 | 1.000 | 1.02 | 5.0 | 0.4 | 237 | 0.749 | 1.04 | 99.5 | 97.9 | 168 |
| 0.023 | 0.01 | 1.001 | 0.78 | 19.0 | 9.0 | 359 | 0.749 | 0.78 | 79.1 | 59.4 | 290 |
| 0.062 | 0.023 | 0.997 | 0.80 | 19.7 | 8.8 | 467 | 0.751 | 0.79 | 56.2 | 36.2 | 434 |

Table 13: Mean treatment estimates, ratio of SEs, Type I error or power, and number of failures for a 45% event rate, 6 clusters, 600 patients per period per cluster. Results on the left are for a scenario with no treatment effect, and those on the right are for a scenario with treatment OR 0.75.

| $\rho_c$ | $\rho_p$ | Mean treatment estimate | Ratio of SEs | Type I error with N | Type I error with t | Failure rate (out of 5000) | Mean treatment estimate | Ratio of SEs | Power with N | Power with t | Failure rate (out of 5000) |
|---|---|---|---|---|---|---|---|---|---|---|---|
| Unweighted regression | | | | | | | | | | | |
| 0.023 | 0 | -0.0002 | 1.00 | - | 4.4 | 0 | -0.0681 | 1.01 | - | 81.9 | 0 |
| 0.062 | 0 | 0.0001 | 1.00 | - | 4.5 | 0 | -0.0658 | 1.00 | - | 81.6 | 0 |
| 0.023 | 0.01 | 0.0001 | 1.00 | - | 5.1 | 0 | -0.0678 | 1.00 | - | 75.0 | 0 |
| 0.062 | 0.023 | 0.0007 | 0.98 | - | 5.0 | 0 | -0.0655 | 1.01 | - | 62.6 | 0 |
| Size-weighted regression | | | | | | | | | | | |
| 0.023 | 0 | -0.0001 | 1.00 | - | 4.4 | 0 | -0.0684 | 0.99 | - | 96.5 | 0 |
| 0.062 | 0 | -0.0000 | 1.01 | - | 4.5 | 0 | -0.0658 | 1.00 | - | 95.9 | 0 |
| 0.023 | 0.01 | -0.0002 | 0.95 | - | 6.4 | 0 | -0.0679 | 0.94 | - | 87.7 | 0 |
| 0.062 | 0.023 | 0.0005 | 0.90 | - | 7.6 | 0 | -0.0654 | 0.92 | - | 73.5 | 0 |
| Random-random model | | | | | | | | | | | |
| 0.023 | 0 | 0.999 | 1.04 | 4.1 | 3.1 | 20 | 0.749 | 1.03 | 96.9 | 96.0 | 35 |
| 0.062 | 0 | 1.000 | 1.05 | 4.0 | 3.2 | 17 | 0.751 | 1.04 | 96.5 | 95.4 | 12 |
| 0.023 | 0.01 | 0.999 | 0.97 | 7.2 | 6.1 | 9 | 0.751 | 0.96 | 89.3 | 87.7 | 22 |
| 0.062 | 0.023 | 1.002 | 0.95 | 7.6 | 6.3 | 5 | 0.752 | 0.97 | 76.0 | 73.8 | 4 |

Table 14: Mean treatment estimates, ratio of SEs, Type I error or power, and number of failures for a 45% event rate, 30 clusters, 55 patients per period per cluster. Results on the left are for a scenario with no treatment effect, and those on the right are for a scenario with treatment OR 0.75.

| $\rho_c$ | $\rho_p$ | Mean treatment estimate | Ratio of SEs | Type I error with N | Type I error with t | Failure rate (out of 5000) | Mean treatment estimate | Ratio of SEs | Power with N | Power with t | Failure rate (out of 5000) |
|---|---|---|---|---|---|---|---|---|---|---|---|
| Unweighted regression | | | | | | | | | | | |
| 0.023 | 0 | -0.0004 | 0.99 | - | 4.8 | 0 | -0.0679 | 0.98 | - | 90.3 | 0 |
| 0.062 | 0 | 0.0002 | 1.00 | - | 4.4 | 0 | -0.0658 | 1.01 | - | 90.3 | 0 |
| 0.023 | 0.01 | -0.0007 | 1.02 | - | 4.3 | 0 | -0.0678 | 0.99 | - | 83.4 | 0 |
| 0.062 | 0.023 | 0.0008 | 1.01 | - | 4.4 | 0 | -0.0658 | 1.02 | - | 70.7 | 0 |
| Size-weighted regression | | | | | | | | | | | |
| 0.023 | 0 | 0.0000 | 0.98 | - | 5.7 | 0 | -0.0679 | 0.99 | - | 99.2 | 0 |
| 0.062 | 0 | 0.0001 | 1.02 | - | 4.9 | 0 | -0.0660 | 1.01 | - | 99.1 | 0 |
| 0.023 | 0.01 | -0.0005 | 0.97 | - | 5.4 | 0 | -0.0677 | 0.93 | - | 92.4 | 0 |
| 0.062 | 0.023 | 0.0009 | 0.91 | - | 7.2 | 0 | -0.0656 | 0.90 | - | 79.1 | 0 |
| Random-random model | | | | | | | | | | | |
| 0.023 | 0 | 1.000 | 1.02 | 4.6 | 4.0 | 28 | 0.750 | 1.03 | 99.3 | 99.0 | 22 |
| 0.062 | 0 | 1.000 | 1.06 | 4.3 | 3.5 | 24 | 0.750 | 1.05 | 99.4 | 99.2 | 29 |
| 0.023 | 0.01 | 0.998 | 0.99 | 5.8 | 5.0 | 20 | 0.751 | 0.96 | 93.8 | 92.8 | 11 |
| 0.062 | 0.023 | 1.003 | 0.97 | 6.6 | 5.7 | 3 | 0.751 | 0.97 | 82.3 | 80.0 | 3 |

Table 15: Mean treatment estimates, ratio of SEs, Type I error or power, and number of failures for a 45% event rate, 30 clusters, 75 patients per period per cluster. Results on the left are for a scenario with no treatment effect, and those on the right are for a scenario with treatment OR 0.75.

| $\rho_c$ | $\rho_p$ | Mean treatment estimate | Ratio of SEs | Type I error with N | Type I error with t | Failure rate (out of 5000) | Mean treatment estimate | Ratio of SEs | Power with N | Power with t | Failure rate (out of 5000) |
|---|---|---|---|---|---|---|---|---|---|---|---|
| 6 clusters | | | | | | | | | | | |
| 0.023 | 0 | 1.002 | 1.03 | 4.9 | 0.4 | 8 | 0.498 | 1.02 | 98.9 | 94.3 | 3 |
| 0.062 | 0 | 1.001 | 1.04 | 4.3 | 0.3 | 2 | 0.499 | 1.02 | 98.7 | 93.8 | 5 |
| 0.023 | 0.01 | 1.004 | 0.84 | 13.1 | 3.4 | 11 | 0.497 | 0.87 | 96.6 | 88.1 | 7 |
| 0.062 | 0.023 | 1.001 | 0.81 | 16.3 | 5.9 | 10 | 0.500 | 0.82 | 90.6 | 75.5 | 9 |
| 12 clusters | | | | | | | | | | | |
| 0.023 | 0 | 0.999 | 1.03 | 4.2 | 2.2 | 9 | 0.496 | 1.02 | 95.3 | 92.7 | 7 |
| 0.062 | 0 | 0.999 | 1.04 | 4.7 | 2.3 | 5 | 0.499 | 1.04 | 95.4 | 92.5 | 8 |
| 0.023 | 0.01 | 0.997 | 0.94 | 7.4 | 4.2 | 10 | 0.500 | 0.97 | 93.5 | 89.3 | 14 |
| 0.062 | 0.023 | 0.996 | 0.92 | 8.4 | 5.3 | 14 | 0.501 | 0.92 | 88.4 | 83.2 | 6 |
| 20 clusters | | | | | | | | | | | |
| 0.023 | 0 | 1.001 | 1.02 | 5.2 | 3.5 | 6 | 0.496 | 1.03 | 95.4 | 94.1 | 11 |
| 0.062 | 0 | 0.998 | 1.05 | 4.2 | 2.9 | 8 | 0.499 | 1.03 | 95.5 | 94.1 | 4 |
| 0.023 | 0.01 | 0.996 | 0.97 | 6.1 | 4.4 | 14 | 0.497 | 0.99 | 94.1 | 92.5 | 11 |
| 0.062 | 0.023 | 0.995 | 0.93 | 8.0 | 6.0 | 6 | 0.496 | 0.97 | 91.8 | 89.6 | 8 |
| 30 clusters | | | | | | | | | | | |
| 0.023 | 0 | 1.007 | 1.01 | 4.6 | 3.6 | 9 | 0.495 | 1.03 | 95.5 | 94.6 | 8 |
| 0.062 | 0 | 0.999 | 1.04 | 3.8 | 3.2 | 7 | 0.496 | 1.04 | 95.8 | 95.0 | 7 |
| 0.023 | 0.01 | 0.999 | 0.98 | 6.0 | 4.9 | 5 | 0.498 | 0.99 | 94.3 | 93.3 | 13 |
| 0.062 | 0.023 | 1.002 | 0.96 | 6.2 | 5.2 | 7 | 0.498 | 0.98 | 92.6 | 91.1 | 5 |
| 50 clusters | | | | | | | | | | | |
| 0.023 | 0 | 0.999 | 1.01 | 4.7 | 4.2 | 21 | 0.500 | 1.02 | 96.5 | 96.1 | 9 |
| 0.062 | 0 | 1.000 | 1.04 | 4.1 | 3.6 | 5 | 0.498 | 1.03 | 96.8 | 96.5 | 6 |
| 0.023 | 0.01 | 0.995 | 1.00 | 5.2 | 4.4 | 18 | 0.498 | 1.01 | 96.2 | 95.8 | 18 |
| 0.062 | 0.023 | 1.002 | 0.97 | 6.2 | 5.6 | 11 | 0.500 | 0.99 | 94.7 | 94.2 | 8 |
| 80 clusters | | | | | | | | | | | |
| 0.023 | 0 | 0.997 | 1.01 | 4.6 | 4.2 | 11 | 0.497 | 1.02 | 95.5 | 95.0 | 8 |
| 0.062 | 0 | 0.996 | 1.03 | 4.4 | 4.1 | 3 | 0.498 | 1.02 | 95.3 | 95.1 | 3 |
| 0.023 | 0.01 | 1.003 | 1.01 | 4.7 | 4.5 | 14 | 0.495 | 0.98 | 94.4 | 94.1 | 10 |
| 0.062 | 0.023 | 1.001 | 1.00 | 5.1 | 4.7 | 7 | 0.498 | 0.99 | 94.3 | 93.8 | 11 |
| 100 clusters | | | | | | | | | | | |
| 0.023 | 0 | 1.000 | 1.00 | 4.8 | 4.6 | 7 | 0.495 | 1.01 | 94.1 | 93.8 | 4 |
| 0.062 | 0 | 1.001 | 1.01 | 4.8 | 4.5 | 8 | 0.497 | 1.01 | 94.0 | 93.7 | 6 |
| 0.023 | 0.01 | 0.998 | 1.00 | 4.8 | 4.6 | 13 | 0.496 | 1.01 | 94.1 | 93.7 | 11 |
| 0.062 | 0.023 | 1.002 | 1.00 | 5.0 | 4.8 | 13 | 0.496 | 0.98 | 93.2 | 92.8 | 7 |

Table 16: Mean treatment estimates, ratio of SEs, Type I error or power, and number of failures for the Random-random model, for a 15% event rate and fixed period OR of 0.85. The average number of patients per cluster-period is 200 for 6 clusters, 60 for 12 clusters, 34 for 20 clusters, 22 for 30 clusters, 14 for 50 clusters, 8 for 80 clusters and 6 for 100 clusters. Results on the left are for a scenario with no treatment effect, and those on the right have a treatment OR of 0.5.

# References


[1] B. Giraudeau, P. Ravaud, and A. Donner. Sample size calculation for cluster randomized cross-over trials. *Statist. Med.*, 27:5578–5585, 2008.

[2] A. Donner, N. Klar, and G. Zou. Methods for the statistical analysis of binary data in split-cluster designs. *Biometrics*, 60:919–925, 2004.

[3] A. Donner and N. Klar. *Design and Analysis of Cluster Randomization Trials in Health Research.* Wiley, 2010.

[4] M. S. Ridout, C. G. Demétrio, and D. Firth. Estimating intraclass correlation for binary data. *Biometrics*, 55:137–48, 1999.

[5] S. Wu, C. M. Crespi, and W. K. Wong. Comparison of methods for estimating the intra-class correlation coefficient for binary responses in cancer prevention cluster randomized trials. *Contemporary Clinical Trials*, 33:869–880, 2012.

[6] StataCorp. Stata Statistical Software: Release 13., 2013. College Station, TX: StataCorp LP.

[7] S. M. Eldridge, O. C. Ukoumunne, and J. B. Carlin. The Intra-Cluster Correlation Coefficient in Cluster Randomized Trials: A Review of Definitions. *Int Statist Rev*, 77:378–394, 2009.

[8] G. Adams, M. C. Gulliford, O. C. Ukoumunne, S. Eldridge, S. Chinn, and M. J. Campbell. Patterns of intra-cluster correlation from primary care research to inform study design and analysis. *J Clin Epi*, 57:785–794, 2004.

[9] B. S. Dahler-Eriksen, T. Lauritzen, J. F. Lassen, E. D. Lund, and I. Brandslund. Near-Patient Test for C-Reactive Protein in General Practice: Assessment of Clinical, Organizational, and Economic Outcomes. *Clinical Chemistry*, 45 (4):478–485, 1999.

[10] J. Dearlove and S. Illingsworth. A controlled trial of parent initiated and conventional preschool health surveillance using personal child health records. *Arch Dis Child*, 80:507–510, 1999.

[11] A. Edwards, G. Elwyn, K. Hood, C. Atwell, M. Robling, H. Houston, P. Kinnersley, and I. Russell. Patient-based outcome results from a cluster randomized trial of shared decision making skill development and use of risk communication aids in general practice. *Family Practice*, 21 (4):347–354, 2004.

[12] J. R. Flanagan, B. N. Doebbeling, J. Dawson, and S. Beekmann. Randomized Study of Online Vaccine Reminders in Adult Primary Care. *Proc AMIA Symp.*, pages 755–9, 1999.

[13] M. A. Harris, J. E. Byles, J. Cockburn, and C. D'Este. A general practice-based recruitment strategy for colorectal screening. *Aust N Z J Public Health*, 24 (4):441–3, 2000.

[14] G. Hejblum, L. Chalumeau-Lemoine, V. Ioos, P.-Y. Boëlle, L. Salomon, T. Simon, J.-F. Vibert, and B. Guidet. Comparison of routine and on-demand prescription of chest radiographs in mechanically ventilated adults: a multicentre, cluster-randomised, two-period crossover study. *Lancet*, 347:1687–93, 2009.

[15] J.-J. Parienti and O. Kuss. Cluster-crossover design: A method for limiting clusters level effect in community-intervention studies. *Contemporary Clinical Trials*, 28:316–323, 2007.

[16] S. J. Connolly, F. Philippon, Y. Longtin, A. Casanova, D. H. Birnie, D. V. Exner, P. Dorian, R. Prakash, M. Alings, and A. D. Krahn. Randomized cluster crossover trials for reliable, efficient, comparative effectiveness testing: design of the prevention of arrythmia device infection trial (PADIT). *Canadian Journal of Cardiology*, 29:652–658, 2013.

[17] S. Johnson, A. Whitelaw, C. Glazebrook, C. Israel, R. Turner, I. R. White, T. Croudace, F. Davenport, and N. Marlow. Randomized Trial of a Parenting Intervention for Very Preterm Infants: Outcome at 2 Years. *The Journal of Pediatrics*, 155 (4):488–494, 2009.

[18] M. Garrouste-Orgeas, L. Soufir, A. Tabah, C. Schwebel, A. Vesin, C. Adrie, M. Thuong, and J. F. Timsit. A multifaceted program for improving quality of care in intensive care units: IATROREF study. *Crit Care Med*, 40 (2):468–476, 2012.

[19] S. M. Eldridge, D. Ashby, G. S. Feder, A. R. Rudnicka, and O. C. Ukoumunne. Lessons for cluster randomized trials in the twenty-first century: a systematic review of trials in primary care. *Clin Trials*, 1 (1):80–90, 2004.

[20] S. M. Eldridge, D. Ashby, and S. Kerry. Sample size for cluster randomized trials: effect of coefficient of variation of cluster size and analysis method. *Int. J. Epi.*, 35:1292–1300, 2006. doi:10.1093/ije/dyl129.