Research paper

An empirical comparison of methods for analyzing over-dispersed zero-inflated count data from stratified cluster randomized trials

Sayem Borhan [a, b, c, d, *], Courtney Kennedy [d], George Ioannidis [d], Alexandra Papaioannou [d, e], Jonathan Adachi [e], Lehana Thabane [a, b, f]

[a] *Department of Health Research Methods, Evidence and Impact, McMaster University, Hamilton, ON, Canada*
[b] *Biostatistics Unit, Research Institute of St Joseph's Healthcare, Hamilton, ON, Canada*
[c] *Department of Family Medicine, McMaster University, Hamilton, ON, Canada*
[d] *GERAS Centre, Hamilton Health Sciences, Hamilton, ON, Canada*
[e] *Department of Medicine, McMaster University, Hamilton, ON, Canada*
[f] *Departments of Pediatrics and Anesthesia, McMaster University, Hamilton, ON, Canada*

A B S T R A C T

Background: The assessment of methods for analyzing over-dispersed zero inflated count outcome has received very little or no attention in stratified cluster randomized trials. In this study, we performed sensitivity analyses to empirically compare eight methods for analyzing zero inflated over-dispersed count outcome from the Vitamin D and Osteoporosis Study (ViDOS) – originally designed to assess the feasibility of a knowledge translation intervention in long-term care home setting.

Method: Forty long-term care (LTC) homes were stratified and then randomized into knowledge translation (KT) intervention (19 homes) and control (21 homes) groups. The homes/clusters were stratified by home size ($<250/> = 250$) and profit status (profit/non-profit). The outcome of this study was number of falls measured at 6-month post-intervention. The following methods were used to assess the effect of KT intervention on number of falls: i) standard Poisson and negative binomial regression; ii) mixed-effects method with Poisson and negative binomial distribution; iii) generalized estimating equation (GEE) with Poisson and negative binomial; iv) zero inflated Poisson and negative binomial — with the latter used as a primary approach. All these methods were compared with or without adjusting for stratification.

Results: A total of 5,478 older people from 40 LTC homes were included in this study. The mean ($=1$) of the number of falls was smaller than the variance ($=6$). Also 72% and 46% of the number of falls were zero in the control and intervention groups, respectively. The direction of the estimated incidence rate ratios (IRRs) was similar for all methods. The zero inflated negative binomial yielded the lowest IRRs and narrowest 95% confidence intervals when adjusted for stratification compared to GEE and mixed-effect methods. Further, the widths of the 95% confidence intervals were narrower when the methods adjusted for stratification compared to the same method not adjusted for stratification.

Conclusion: The overall conclusion from the GEE, mixed-effect and zero inflated methods were similar. However, these methods differ in terms of effect estimate and widths of the confidence interval.

Trial registration: ClinicalTrials.gov: NCT01398527. Registered: 19 July 2011.

1. Background

Randomized trials involving allocation of intact groups or clusters of subjects, instead of independent individuals, are commonly referred to as cluster randomized trials [1]. The rate of adopting cluster randomization trials is increasing [2]. Allocation units are diverse in such studies, and can include families or households, classrooms or schools [3], long-term care homes [4] or even entire communities [5].

Depending on the allocation of clusters, most cluster randomization trials can be classified as using one of three basic types of de-

---

signs: (a) completely randomized, (b) matched-pair, or (c) stratified. Completely randomized designs omit pre-stratification and matching on baseline prognostic factors. This design is most suited for trials enrolling fairly large numbers of clusters [6]. Random assignment of one of the two clusters in a stratum to each intervention group is termed a matched-pair design [6]. The stratified design extends the matched-pair design where more than two clusters are randomly allocated to intervention groups within strata. For example, Vitamin D and Osteoporosis Study (ViDOS) [4,7] conducted a pilot stratified cluster randomized trial – where long-term care (LTC) home were stratified by size and profit status, to assess the effect of a multifaceted knowledge translation (KT) intervention on prescribing vitamin D, calcium and osteoporosis medication in long-term care home.

Random allocation of clusters may result in similarity among the outcomes from the same cluster, which is measured using an intra-cluster correlation coefficient (ICC) [1]. This correlation among the responses from the same cluster invalidates the application of statistical techniques which assume independence of observations. Thus, standard statistical methodology needs to be adjusted for this clustering effect, which can be quantified by the design effect, or variance inflation factor, given by $1 + (\bar{m} - 1) ICC$, where $\bar{m}$ is the average cluster size [1].

Donner and Klar [1] discussed about several approaches to analyze count data from cluster randomized trials including cluster-specific and population-average extension of Poisson regression. They also discussed we can easily extend these approaches for stratified cluster randomized trials. Similarly, Young et al. [8] compared the performance of cluster-specific and population-average extension of Poisson regression using data from a non-randomized study while Pacheco et al. [9] investigated the performance of methods for analyzing over-dispersed – variance is greater than the mean, count outcome from completely randomized CRT. Further, to account the count outcome with excess zeros we need to use the zero-inflated models. To the best of our knowledge, no study examined the methods for analyzing over dispersed and zero-inflated count data from stratified cluster randomized trials.

On the other hand, Thabane et al. [10] rightfully emphasized the importance of performing a sensitivity analysis, which help us to assess the robustness of the results. For cluster randomized trials we can perform sensitivity analyses with or without taking clustering into account. We can also compare the methods with or without considering the stratification. Borhan et al. [11] examined the sensitivity of methods for analyzing continuous outcome from stratified cluster randomized trials and found the overall conclusion from all the methods were similar.

In this study, we performed sensitivity analyses to empirically compare eight methods for analyzing zero inflated over-dispersed count outcome from the ViDOS study [4].

2. Methods

2.1. Motivating example: ViDOS study

We used the data from an LTC-based pilot stratified cluster randomized trial – details can be found elsewhere [4,7], for this study. A total of 5,478 older people from 40 LTC homes (19 Intervention and 21 Control) were randomized into two groups KT intervention and control groups. The LTC homes were stratified by size (<250 vs ≥ 250 beds) and profit status (profit vs non-profit). Seven LTC homes withdrew before the study began. The outcome, number of falls were measured at 6- and 12-month post-randomization. For this study, we used the number of falls measured at 12-month. The variance of the number of falls is greater than the mean number of falls (variance = 6 > mean = 1). Similarly, for each cluster the mean number of falls is smaller than the variance of the number of falls.

Thus, the number of falls was over-dispersed. Further, the number of falls was zero inflated as 72% and 46% of the number of falls were zero in the control and intervention groups, respectively.

2.2. Statistical analysis methods

Both cluster-specific (mixed-effect method) and population-average (generalized estimating equation) methods were used to analyze the number of falls from the ViDOS study. The mixed-effect zero-inflated negative binomial model was considered as the primary method since it can take into account both overdispersion and zero-inflation as well as clustering. The adjustment for stratification covariates – home size and profit status, were applicable for cluster- and individual-level methods, since these were cluster-level covariates. The results from the analyses were reported in terms of the incidence rate ratios (IRRs) along with 95% confidence intervals (CIs) and associated p-values. All statistical tests were two-sided at the significance level of 0.05. The p-value less than 0.001 were reported as <0.001 The reporting of the results follows the CONSORT (Consolidated Standards for Reporting Trials) guidelines for reporting cluster-randomized trials [12].

Data were analyzed using Intention-to-treat (ITT) principles and missing data analysis approach – where missing data were imputed using multiple imputation technique assuming missing data follows a missing at random (MAR) pattern. Overall, five datasets were generated, and pooled estimates were reported.

2.3. Standard Poisson/Negative binomial (NB) model

The standard Poisson and negative binomial model for count data is given by

$$\log(E(Y_{ijkl}) = \mu_{ijkl}) = \beta_0 + \beta_1 X_{ijkl} + \beta_2 S_{1ijkl} + \beta_3 S_{2ijkl} + e_{ijkl}$$

Where, Y_{ijkl} is the outcome, number of falls, of the $i-th$ subject of the $j-th$ cluster in the $k-th (k = 0,1)$ and $l-th (l = 0,1)$ stratum. X_{ijkl} is the intervention (0: Control; 1: KT Intervention). S_{1ijkl} (0:<250; 1> = 250) is the home size and S_{2ijkl} (0: Non-profit; 1: Profit) is the profit status of the cluster.

Here, β_1 represents the treatment effect while β_2 and β_3 represents the two strata effect corresponding to home size (0:<250; 1: ≥250) and profit status (0: Non-profit; 1: Profit), respectively.

We considered two distributional assumptions for number of falls:

(a) Number of falls follows a Poisson distribution i.e. $Y_{ijkl} \sim Poi(\mu_{ijkl})$, with variance function $V(Y_{ijkl}) = \varphi v(\mu_{ijkl}) = \mu_{ijkl}$, where φ is assumed to be 1 i.e. mean and variance are equal.

(b) Number of falls follows a Negative Binomial (NB) distribution i.e. $Y_{ijkl} \sim NB(s, \mu_{ijkl})$, with variance function $V\left(Y_{ijkl}\right) = \varphi v\left(\mu_{ijkl}\right) = \varphi \left(\mu_{ijkl} + s\mu_{ijkl}^2\right)$, where φ is assumed to be 1 and s is the overdispersion parameter indicating that the NB distribution models overdispersion implicitly by its parameter s. The NB distribution is preferred when there is overdispersion in the data i.e. mean < variance.

The standard Poisson and negative binomial model were fitted using glm() and glm.nb() in R [13].

2.4. Mixed-effect model (Poisson/Negative binomial)

The mixed-effect model for count data is given by

$$\log(E(Y_{ijkl}) = \mu_{ijkl}) = \beta_0 + \beta_1 X_{ijkl} + \beta_2 S_{1ijkl} + \beta_3 S_{2ijkl} + C_{ijk} + e_{ijkl}$$

In this model, like the previous model, β_1 represents the treatment effect while β_2 and β_3 represents the two stratum effect corresponding

to home size (0: < 250; 1: ≥250) and profit status (0: Non-profit; 1: Profit), respectively, which are fixed. Random cluster effect is represented by C_{ijk}, which follows a normal distribution with mean 0 and variance σ_b^2. The intra-cluster correlation that measures the correlation among the outcomes within cluster is given by $\frac{\sigma_b^2}{\sigma_b^2+\sigma_e^2}$, assumed equal for all clusters. β_1 is the log of the Rate Ratio (RR) of the intervention X_{ijkl} (0 = Control, 1 = KT Intervention). We used glmer() and glmer.nb() in R to fit mixed-effect with Poisson and negative binomial, respectively.

2.5. Generalized estimating equation (GEE) (Poisson/Negative binomial)

The GEE model for count data is given by

$$\log(E(Y_{ijkl}) = \mu_{ijkl}) = \beta_0 + \beta_1 X_{ijkl} + \beta_2 S_{1ijkl} + \beta_3 S_{2ijkl}$$

Like before, β_1 represents the treatment effect while β_2 and β_3 represents the two stratum effect corresponding to home size (0: < 250; 1: ≥250) and profit status (0: Non-profit; 1: Profit), respectively. Similar to mixed-effect method we considered two distributional assumption for count data: Poisson and negative binomial. For GEE method we considered exchangeable working correlation structure. GEE with Poisson was fitted using geeglm() in R while GEE with negative binomial was fitted using PROC GENMOD in SAS [14]. GEE with negative binomial was the primary method of analysis.

2.6. Zero inflated models (Poisson/Negative binomial)

For zero inflated models the distribution of Y_{ijkl} is

$$Y_{ijkl} = \begin{cases} 0; & \text{with probability } \varphi_{ijkl} \\ Poisson \text{ or } NB\left(\mu_{ijkl}\right); & \text{with probability } \left(1 - \varphi_{ijkl}\right) \end{cases}$$

The mixed-effect zero inflated Poisson or negative binomial model is given by:

$$logit(\varphi_{ijkl}) = \beta_0 + \beta_1 X_{ijkl} + \beta_2 S_{1ijkl} + \beta_3 S_{2ijkl} + C_{ijk} + e_{ijkl}$$

$$\log(E(Y_{ijkl}) = \mu_{ijkl}) = \beta_0 + \beta_1 X_{ijkl} + \beta_2 S_{1ijkl} + \beta_3 S_{2ijkl} + C_{ijk} + e_{ijkl}$$

The zero inflated Poisson and negative binomial models were fitted using the R package GLMMadaptive.

3. Results

Overall 40 clusters were randomized into KT intervention (19 clusters) and control (21 clusters) groups. The clusters were stratified by two variables cluster size and profit status. The average cluster size in the KT group was 115 (minimum = 43, maximum = 294) while the average cluster size in the control group was 157 (minimum = 49, maximum = 375). At the end of the follow-up there were 2,209 participants in the intervention group and 3,382 participants in the control group. The average age of the participants in both groups were 84 years while approximately 70% were female.

We used the methods discussed above to assess the effect of KT intervention on number of falls with mixed-effect zero-inflated with negative binomial distribution as the primary method of analysis. The results of the ITT analyses with or without adjusted for stratification are given in Fig. 1. The direction of the effect estimate incidence rate ratios were similar for all the methods. The standard Poisson and negative binomial regression methods yielded statistically significant results as p-values lower than the nominal level of 0.05 while the other methods yielded non-significant results (Fig. 1). The estimated IRRs varies from 1.11 to 1.37 when adjusted for stratification and 1.03 to 1.49 when not adjusted for stratification. The effect estimates IRRs were slightly higher for mixed-effect methods compared to other methods. The magnitude of the widths of the 95% confidence intervals were higher for mixed-effect Poisson and negative binomial methods compared to other methods when adjusted or not adjusted for stratification (Fig. 1). The Akaike's Information Criteria (AIC) were slightly lower when the methods adjusted for stratification compared to without such adjustment. Further, the AIC values were lower for negative binomial models (8391.00 and 8333.24 for mixed-effect and zero-inflated negative binomial models respectively) compared to GEE models (10858.00 and 9093.10 for mixed-effect and zero-inflated Poisson models respectively).

The results of the missing data analysis were given in Fig. 2. Unlike ITT approach, standard Poisson and negative binomial did not yield statistically significant results (Fig. 2). Similar to ITT approach, direction of effect estimate for all the methods were similar. The estimated IRRs varies from 1.35 to 2.12, when adjusted for stratification and 1.41 to 1.96 when not adjusted for stratification. The magnitudes of the widths of the 95% confidence intervals were higher for all methods compared to ITT approach. Similar to ITT 95% confidence intervals were wider for mixed-methods, when not accounted for zero inflation, compared to other methods (Fig. 2).

For all methods, the estimated IRRs were very similar with or without adjusting for stratification for both ITT and missing data analysis approaches (Figs. 1-2). Further, it is noticeable, that the estimated IRRs were slightly higher, for all methods, in missing data analysis approach compared to ITT approach (Figs. 1-2). Also, for ITT approach, the 95% confidence intervals were slightly narrower when
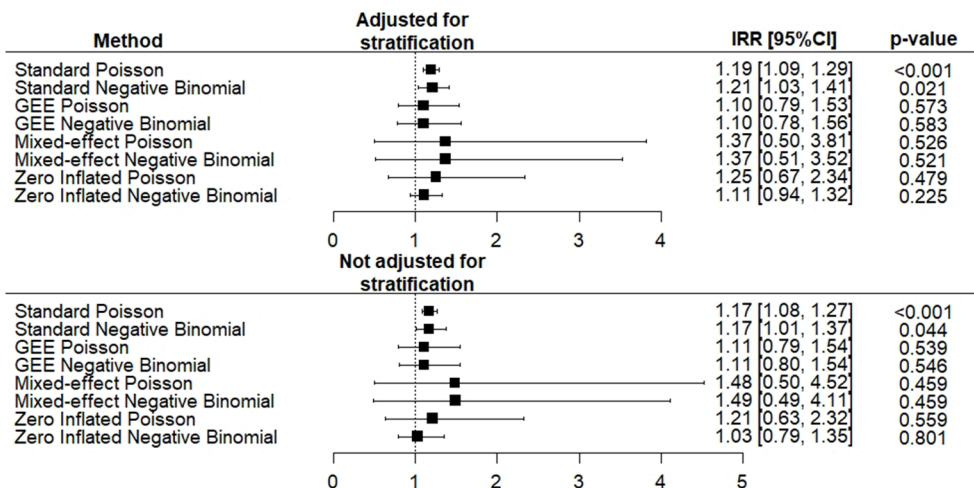

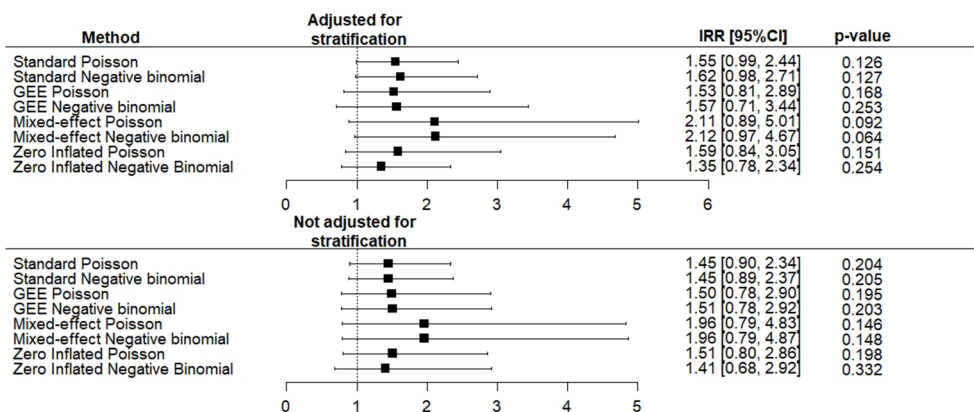**Fig. 1.** Results of ITT analysis using different methods with/without adjusted for stratification.

Fig. 2. Results of missing data analysis using different methods with/without adjusted for stratification.

adjusted for stratification (Fig. 1). The difference among the methods in terms of p-values were smaller for missing data analysis approach compared to ITT approach (Figs. 1 and 2).

4. Discussion

In this study, we empirically investigate the methods for analyzing overdispersed zero inflated count outcome from stratified cluster randomized trial using data from the ViDOS study – which was designed to investigate the effect of a KT intervention. We compared eight methods to assess the effect of KT intervention on number of falls. The direction of effect of estimate incidence rate ratios (IRRs) were similar for all methods for both adjusted and not adjusted for stratification. The conclusions from both ITT and missing data analyses indicated that, KT intervention had no effect on number of falls.

For ITT analyses, both standard Poisson and negative binomial methods yielded statistically significant results that the RRs of number of falls were slightly higher in the intervention group compared to control group. However, these two methods were not appropriate for analyzing count data from CRT as these methods do not take into account the degree of similarity among the outcomes from the same cluster.

In this study, we considered mixed-effect with zero-inflated negative binomial as the primary method of analysis to assess the effect of KT intervention on over dispersed number of falls. We performed sensitivity analyses to examine the robustness of the findings of the primary method. The overall conclusion from all the methods were similar. These findings match with the findings of the Borhan et al. [11] when they investigated the sensitivity of several methods for analyzing continuous outcome from the stratified CRT.

Overall, for all methods, the estimated IRRs and the corresponding widths of the 95% confidence intervals were slightly lower for ITT analyses compared to missing data analyses. GEE and mixed-effect with Poisson and negative binomial distributions, respectively, yielded approximately similar IRRs. The estimated IRRs and widths of the 95% confidence intervals were lower for zero inflated models compared to mixed-effect methods with Poisson and negative binomial distribution. The widths of the 95% confidence intervals were lower for GEE methods compared to mixed-effect methods for both ITT and missing data analyses. This is consistent with the findings of Pacheco et al. [9]. The authors reported that, GEE yielded the highest power and narrow CIs when the authors investigated the performance of methods for analyzing overdispersed count data from CRT. However, GEE underestimate the covariance among observations yielding downward biased standard errors when the number of clusters is small [15]. Also, we need to be cautious that, GEE method yields elevated type I error rates in small sample situations (<40 clusters) [9].

We also compared the methods with or without adjusting for stratification. Zero inflated negative binomial yielded the lowest IRRs and narrowest 95% confidence intervals when adjusted for stratification among the valid methods. For ITT approach, the estimated IRRs and the widths of the 95% confidence intervals were almost similar or lower for both GEE methods. Similarly, for mixed-effect methods the estimated RRs and the magnitude of the widths of the 95% confidence intervals were slightly lower when we adjusted for stratification. These findings matched with the findings of Borhan et al. [11], Ma et al. [16] and Kahan et al. [17], where the authors compared several methods for analyzing continuous and binary data from stratified CRT and continuous data from stratified randomized controlled trial on individual, respectively. Similarly, for missing data approach, GEE yielded the similar results with or without adjusted for stratification. For all methods, the p-values were lower when adjusted for stratification compared to same method when not adjusted for stratification and matched with the findings of Kahan et al. [17].

The major strength of this study that, we empirically examined eight methods, including both cluster-specific and population-average methods, for analyzing count outcome from a stratified CRT - ViDOS study, under different scenarios including accounting for clustering and adjusting for stratification. We also compared the methods through ITT approach and imputing the missing data. In addition, we used appropriate method such as negative binomial to account for overdispersion and zero inflated models to account for excess zeros. Thus, this study will guide researchers about the sensitivity of these methods since there is no study, to the best of our knowledge, investigate the performance of these methods for analyzing count data from stratified CRT.

The major limitation of this study, that ViDOS study was a pilot trial designed to investigate the feasibility of the KT intervention. However, ViDOS was stratified by two cluster-level covariates cluster size and profit status, which is very rare in real life. It is possible that, we might have missed some falls data as it is difficult to measure the number of falls and varies between LTCs.

Data from 7 clusters were missing in the intervention group as 6 clusters declined to actively participate after randomization and 1 cluster withdrew after baseline measurement. Further study on missing data imputation techniques when the whole cluster is missing would be an important addition. Furthermore, a well-designed simulation study is warranted to examine the performance of these methods under different scenarios. It requires large number of clusters (>30) to get valid estimate using GEE and mixed-effect methods [18–21]. Researchers have suggested some corrections to address the requirement of large number of clusters [22–26] which can be extended to stratified CRT, especially when the outcome is count.

5. Conclusion

In this study, we empirically compared the eight methods for analyzing count outcome using the data from ViD*OS* study - a pilot stratified cluster randomized trial. The overall conclusion from all the methods were similar that the KT intervention had no effect on number of falls. The zero inflated negative binomial model yielded the lowest IRR and narrowest 95% confidence interval, when adjusted for stratification, compared to GEE and mixed-effect methods. A well-designed simulation study is warranted to assess the performance of these methods.

Acknowledgements

References

[1] A. Donner, N. Klar, Design and Analysis of Cluster Randomization Trials in Health Research, Arnold London, 2000.

[2] J. Bland, Cluster randomised trials in the medical literature: two bibliometric surveys, BMC Med. Res. Methodol. 4 (2004) 21–27.

[3] R. Mallick, H. Kathard, A.S.M. Borhan, M. Pillay, L. Thabane, A Cluster randomised trial of a classroom communication resource program to change peer attitudes towards children who stutter among grade 7 students, Trials 19 (2018) 664.

[4] Kennedy, et al., Successful knowledge translation intervention in long-term care: final results from the vitamin D and osteoporosis study (ViDOS) pilot cluster randomized controlled trial, Trials 16 (2015) 214.

[5] J. Kaczorowski, et al., Cardiovascular Health Awareness Program (CHAP): a community cluster-randomised trial among elderly Canadians, Prev. Med. 46 (6) (2008 Jun) 537–544.

[6] N. Klar, A. Donner, The merits of matching in community intervention trials: a cautionary tale, Stat. Med. 16 (1997) 1753–1764.

[7] C.C. Kennedy, G. Ioannidis, L.M. Giangregorio, J.D. Adachi, L. Thabane, S.N. Morin, et al., An interdisciplinary knowledge translation intervention in long-term care: study protocol for the vitamin D and osteoporosis study (ViDOS) pilot cluster randomized controlled trial, Implement. Sci. 7 (2012) 48.

[8] M. Young, J. Preisser, B. Qaqish, M. Wolfson, Comparison of subject-specific and population averaged models for count data from cluster-unit intervention trials, Stat. Methods Med. Res. 16 (2007) 167–184.

[9] Pacheco, et al., Performance of analytical methods for overdispersed counts in cluster randomized trials: sample size, degree of clustering and imbalance, Stat. Med. 28 (2009) 2989–3011.

[10] Thabane, et al., A tutorial on sensitivity analyses in clinical trials: the what, why, when and how, BMC Med. Res. Methodol. 13 (1) (2013) 92 2013.

[11] S. Borhan, R. Mallick, M. Pillay, H. Kathard, L. Thabane, Sensitivity of methods for analyzing continuous outcome from stratified cluster randomized trials – an empirical comparison study, Contemp. Clin. Trials Commun. 15 (2019) 100405.

[12] M. Campbell, D. Elbourne, D. Altman, CONSORT group, CONSORT statement: extension to cluster randomised trials, BMJ 328 (7441) (2004) 702–708.

[13] R Core Team, R: A Language and Environment for Statistical Computing, R Foundation for Statistical Computing, Vienna, Austria, 2019 URL. https://www.R-project.org/.

[14] SAS/STAT_ User's Guide, SAS Institute, Inc., 2019.

[15] L.A. Mancl, T.A. DeRouen, A covariance estimator for GEE with improved small-sample properties, Biometrics 57 (1) (2001) 126–134.

[16] Ma, et al., Comparison of Bayesian and classical methods in the analysis of cluster randomized controlled trials with a binary outcome: the Community Hypertension Assessment Trial (CHAT), BMC Med. Res. Methodol. 9 (2009) 37.

[17] B. Kahan, T. Morris, Reporting and analysis of trials using stratified randomisation in leading medical journals: review and reanalysis, BMJ 345 (2012) e5840.

[18] R. Prentice, Correlated binary regression with covariates specific to each binary observation, Biometrics 44 (4) (1988) 1033–1048.

[19] L. Mancl, T. DeRouen, A covariance estimator for GEE with improved small-sample properties, Biometrics 57 (1) (2001) 126–134.

[20] Leyrat, et al., Cluster randomized trials with a small number of clusters: which analyses should be used?, Int. J. Epidemiol. 47 (1) (2018) 321–331.

[21] N. Ivers, M. Taljaard, S. Dixon, et al., Impact of CONSORT extension for cluster randomized trials on quality of reporting and study methodology: review of random sample of 300 trials, 2000–8, BMJ 343 (2011) d5886.

[22] M. Kenward, J. Roger, Small sample inference for fixed effects from restricted maximum likelihood, Biometrics 53 (1997) 983–997.

[23] M. Fay, B. Graubard, Small-sample adjustments for Wald-type tests using sandwich estimators, Biometrics 57 (2001) 1198–1206.

[24] L. Mancl, T. DeRouen, A covariance estimator for GEE with improved small-sample properties, Biometrics 57 (2001) 126–134.

[25] P. Li, D. Redden, Comparing denominator degrees of freedom approximations for the generalized linear mixed model in analysing binary outcome in small sample cluster-randomized trials, BMC Med. Res. Methodol. 15 (2015) 38.

[26] P. Li, D. Redden, Small sample performance of bias-corrected sandwich estimators for cluster-randomized trials with binary outcomes, Stat. Med. 34 (2015) 281–296.