Contemporary Clinical Trials Communications 15 (2019) 100405

Contents lists available at ScienceDirect

Contemporary Clinical Trials Communications

journal homepage: www.elsevier.com/locate/conctc

# Sensitivity of methods for analyzing continuous outcome from stratified cluster randomized trials – an empirical comparison study

Sayem Borhan[a,b,c,**], Rizwana Mallick[d], Mershen Pillay[e], Harsha Kathard[d], 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] University of Cape Town, Rondebosch, Cape Town, South Africa
[e] University of KwaZulu Natal, Durban, South Africa
[f] Department of Pediatrics and Anesthesia, McMaster University, Hamilton, ON, Canada

ARTICLE INFO

*Keywords:*
Stratification
Cluster randomized trial
Sensitivity analysis
Continuous

ABSTRACT

The assessment of the sensitivity of statistical methods has received little attention in cluster randomized trials (CRTs), especially for stratified CRT when the outcome of interest is continuous. We empirically examined the sensitivity of five methods for analyzing the continuous outcome from a stratified CRT - aimed to investigate the efficacy of the Classroom Communication Resource (CCR) compared to usual care to improve the peer attitude towards children who stutter among grade 7 students. Schools – the clusters, were divided into quintile based on their socio-political resources, and then stratified by quintile. The schools were then randomized to CCR and usual care groups in each stratum. The primary outcome was Stuttering Resource Outcomes Measure. Five methods, including the primary method, were used in this study to examine the effect of CCR. The individual-level methods were: (i) linear regression; (ii) mixed-effects method; (iii) GEE with exchangeable correlation structure (primary method of analysis). And the cluster-level methods were: (iv) cluster-level linear regression; and (v) meta-regression. These methods were also compared with or without adjustment for stratification. Ten schools were stratified by quintile, and then randomized to CCR (223 students) and usual care (231 students) groups. The direction of the estimated differences was same for all the methods except meta-regression. The widths of the 95% confidence intervals were narrower when adjusted for stratification. The overall conclusion from all the methods was similar but slightly differed in terms of effect estimate and widths of confidence intervals.

*Trialregistration:* Clinicaltrials.gov, NCT03111524. Registered on 9 March 2017.

## 1. Background

Randomization of intact groups, namely clusters, into intervention groups are known as cluster randomized trials (CRT) [1]. Over the years, the number of adopting CRTs is increasing [2]. Diverse types of clusters can be allocated in CRTs including: geographical areas [3]; health care districts [4]; and schools [5]. Like trials on individuals', most CRTs use one of the following three experimental design strategy such as: (a) completely randomized; (b) matched-pair; or (c) stratified. A completely randomized design is satisfactory with substantial number of clusters while stratified design is suitable for small number of clusters [6]. In stratified designs, clusters are randomly allocated to the intervention and control groups within each stratum. For example, Mallick et al. [5] conducted a school-based CRT to investigate the effect of the Classroom Communication Resource (CCR), vs Usual Care, to improve the peer attitude towards children who stutter (CWS). In this trial, schools were first divided into quintile (1–3: lower and 4–5: higher) and stratified as a high vs low school based on the socio-economic resources [5].

Due to the randomization of intact clusters, the outcome from the same cluster may be similar. The intra-cluster correlation coefficient (ICC) is used to measure the degree of similarity [1]. The variance of the estimated intervention effect is inflated due to this correlation and may produce spurious statistically significant results [1,7]. This

[*] Corresponding author. Biostatistics Unit, Research Institute of St Joseph's Healthcare, 3rd Floor, Martha Wing, Room H-325, 50 Charlton Avenue East, Hamilton, ON, L8N 4A6, Canada.

[**] Corresponding author. Department of Health Research Methods, Evidence and Impact, McMaster University, Hamilton, ON, Canada.

*E-mail addresses:* borhana@mcmaster.ca (S. Borhan), thabanl@mcmaster.ca (L. Thabane).

https://doi.org/10.1016/j.conctc.2019.100405
Received 6 March 2019; Received in revised form 20 June 2019; Accepted 3 July 2019
Available online 05 July 2019

inflation can be quantified by the design effect, given by $1 + (\bar{m} - 1)ICC$, where $\bar{m}$ is the average cluster size [1]. Thus, the statistical methods should take into account the potential correlation among the outcomes from the same cluster. Further, the methodologies need to be adjusted for stratification due to stratified design. Researchers have recommended several approaches to analyze the continuous data from completely randomized CRT, which can be extended to stratified designs [1]. The methodologies are broadly classified into two categories: individual- and cluster-level methods. Individual-level methods use the individual-level data such as mixed-model [8] or generalized estimating equation (GEE) [9]. Similarly, we can employ the meta-analytic approach (cluster-level method), which commonly used to combine the results from different studies [10]. This approach helps to aggregate the treatment effects over multiple stratum, like multicentre trials [11–13].

In addition, it is vital to assess the robustness of the results obtained from the randomized controlled trials [14]. The sensitivity analysis helps us to assess the robustness of the results [14]. For CRTs, we can perform several sensitivity analyses. First, we can conduct sensitivity analyses with or without considering the clustering. Secondly, results are compared using different correlation structures [14]. For stratified designs, we can also assess robustness by comparing the methods with or without adjusted for stratification. The GEE with exchangeable correlation structure was used as the primary method of analysis in the Mallick et al. [5] study.

In this study, we empirically examined the sensitivity of methods for analyzing continuous outcome from the stratified CRT using the data from the Mallick et al. [5] study, which in turn demonstrated the robustness of the results obtained using the primary GEE method.

## 2. Methods

### 2.1. Overview of the mallick et al. study

The details about the Mallick et al. study can be found elsewhere [5,15]. In brief, this was a cluster randomized trial aimed at examining the effect of Classroom Communication Resource (CCR) on peer attitude towards Children Who Stutter (CWS) in South African schools in the Western Cape. Schools were the unit of randomization and the participants of this trial were the grade 7 students. The selected schools were first stratified to high or low quintile groups and then randomized to CCR or usual care groups. The grade 7 teachers in the intervention group received training on CCR and administered the intervention (including a social story, role-play and facilitated discussion) while participants in the control group received usual curriculum. The participants were assessed 6-month post intervention. The primary outcome was Stuttering Resource Outcomes Measure (SROM) completed at baseline and 6-month post intervention. The study flow chart is presented in Fig. 1.

### 2.2. Statistical methods

Both individual-level and cluster-level methods were used to analyze the data from the Mallick et al. [5] study. The cluster-and individual-level methods can be adjusted for cluster-level covariates, while individual-level methods can be adjusted for individual-level covariates. The adjustment for stratification covariate, quintile, was applicable for cluster- and individual-level methods, since this was a cluster-level covariate. The results from the analyses were reported in terms of difference (Intervention - Control) along with 95% confidence interval (CI) and associated p-value. 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 [16].

Data were analyzed using both intention-to-treat (ITT) and per-protocol principles. 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. All analyses were performed using statistical software R [17].

#### 2.2.1. Individual-level methods

#### 2.2.1.1. Linear regression

The linear regression can be expressed as

$$Y_{ijkl} = \beta_0 + \beta_1 X_{ijkl} + e_{ijkl}$$

Where $Y_{ijkl}$ is the outcome of the $l$-th subject in the $k$-th cluster, $j$-th intervention group and $i$-th stratum. $X_{ijkl}$ represents the intervention assignment ($X_{ijkl} = 1$ for the treatment group; $X_{ijkl} = 0$ for the control), and $e_{ijkl}$ is the random error assumed to follow a normal distribution with mean 0 and variance $\sigma_e^2$. The intercept ($\beta_0$) represents the mean outcome for the control group in all clusters, while the slope ($\beta_1$) represents the effect of the treatment on the mean outcome.

The linear regression model assumes that data from the participants are independent. This model was implemented using R package lm().

#### 2.2.1.2. Mixed-effects regression model

The mixed-effects regression model is given by

$$Y_{ijkl} = \beta_0 + \beta_1 X_{ijkl} + \beta_2 S_{ijkl} + C_{ijk} + e_{ijkl}$$

In this model, $\beta_1$ and $\beta_2$ represents the treatment and stratum effect, respectively, which are fixed. Random cluster effect is represented by $C_{ijk}$, which follows a normal distribution with mean 0 and variance $\sigma_b^2$. 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. We fitted this model using lme4() package in R with restricted maximum likelihood (REML) method [18,19].

#### 2.2.1.3. Generalized estimating equation (GEE)

The generalized estimating equation (GEE) [9] has the advantage of taking into account the correlation of the outcomes through specification of working correlation structure. The estimated treatment effect from the GEE model reflects the both within- and between – cluster relationship [20]. The sandwich covariance estimator yields a robust estimate of treatment effect in the case when the correlation structure is misspecified [21]. Also, small number of clusters leads to an underestimate of variance [22].

For the primary GEE analysis, the exchangeable correlation structure, which based on the assumption that the individuals within the same cluster are equally correlated, was used. Also, this analysis was performed using sandwich method for standard error estimation. This analysis was performed using geepack () package in R.

#### 2.2.2. Cluster-level methods

#### 2.2.2.1. Cluster-level linear regression

This method consists of first estimating a summary measure by cluster such as mean, and then fitting a linear regression based on these summary measures [1].

#### 2.2.2.2. Meta-regression

This is a meta-analytic approach where cluster-level summary is used [10]. This can be extended to perform a stratified analysis on the mean difference in outcome between intervention and control arms within stratum. The overall treatment effect is estimated by a weighted average of individual mean differences across all strata. The principle of inverse-variance weighting is often used [10]. We implemented this method using the metacont() package in R.

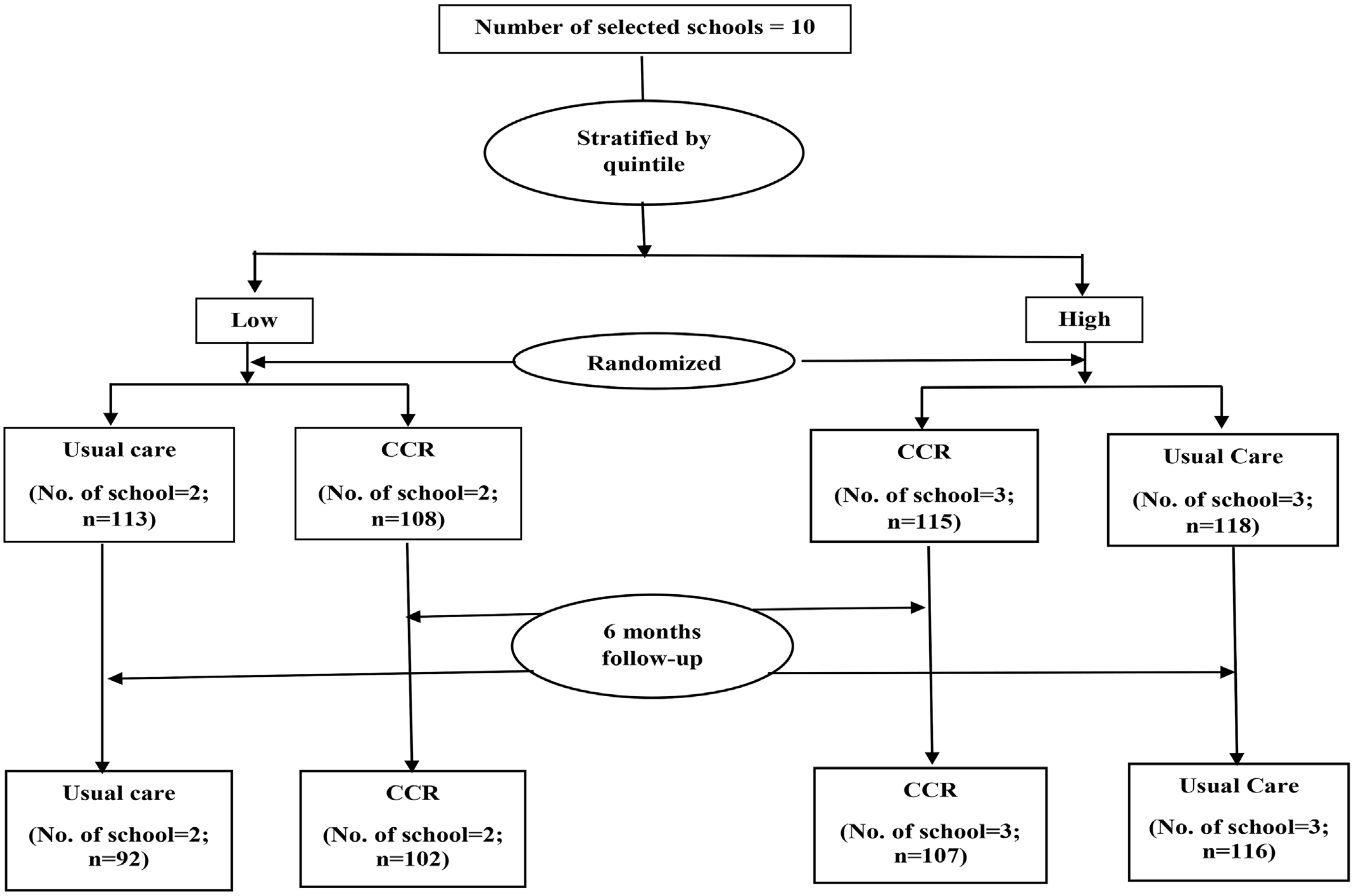
Number of selected schools = 10
Stratified by quintile
Low
High
Randomized
Usual care (No. of school=2; n=113)
CCR (No. of school=2; n=108)
CCR (No. of school=3; n=115)
Usual Care (No. of school=3; n=118)
6 months follow-up
Usual care (No. of school=2; n=92)
CCR (No. of school=2; n=102)
CCR (No. of school=3; n=107)
Usual Care (No. of school=3; n=116)

**Fig. 1.** Study flow chart of the Mallick et al. study.

## 3. Results

In total, the selected 10 schools were stratified into two groups: higher quintile (6 schools) and lower quintile (4 schools). The schools were then randomized into the intervention CCR group and control usual care group. The average cluster size was 45 (range: 30–54) and 46 (range: 18–68) in the CCR and usual care groups, respectively. Overall, 454 students (223 in the CCR group and 231 in the usual care group) participated in this study. The average age was 13 years for both groups.

We used the methods discussed above (see statistical methods section) to evaluate the effect of intervention. The results of the estimated intervention effect, using ITT, are provided in Fig. 2 with and without adjustment for stratification. Results from all the methods, for the outcome SROM, indicated that the intervention CCR had no statistically significant effect as all the p-values were greater than the nominal level of 0.05 (Fig. 2). The estimated mean differences (MDs) were negative for all the methods except meta regression approach when adjusted for stratification (MD = 0.01[-0.48, 0.50]) (Fig. 2). The p-values for all the methods were similar or lower when adjusted for stratification compared to the same method when not adjusted for stratification, while cluster-level linear regression yielded the lowest p-value (Fig. 2). The magnitude of the widths of the confidence intervals were narrower for cluster-level linear regression (1.06 (when adjusted for stratification); 1.36 (when not adjusted for stratification)) and meta regression (0.98 (when adjusted for stratification); 1.07 (when not adjusted for stratification)) compared to other methods. The widths of the confidence intervals were wider when the methods were not adjusted for stratification compared to the same method adjusted for stratification.

The estimated results of the intervention effect using per-protocol principle are provided in Fig. 3 with and without adjusted for stratification. Similar to ITT analyses, results from per-protocol analyses yielded that the intervention CCR had no statistically significant effect on the outcome SROM as all the p-values were greater than the nominal level of 0.05 for both with and without adjustment for stratification (Fig. 3). The p-values were lower for all the methods when adjusted for stratification (Fig. 3). Also, like ITT, the estimated mean difference was positive (MD = 0.08 [-0.99, 1.15]) for the meta regression method in case of per-protocol analysis. The magnitude of the effect size was higher in the per-protocol analyses compared to ITT analyses except GEE with exchangeable correlation structure (when not adjusted for stratification) (Fig. 3).

For both ITT and per-protocol approaches, the standard errors (SEs) were lower for methods when adjusted for stratification compared to the same method when not adjusted for stratification (results are not presented here).

## 4. Discussion

In this study, we had empirically investigated the sensitivity of several methods for analyzing continuous outcome from the stratified cluster randomized trial using data from the Mallick et al. [5] study. We used five methods in a frequentist framework to assess the effect of the intervention CCR on SROM compared to usual care. These methods can be differentiated by whether they account the clustering effect or adjust for stratification or both. The overall conclusion, based on intention-to-treat and per-protocol analyses, from all the methods was similar to the primary method (GEE with exchangeable correlation structure) i.e. there was no significant difference between the intervention groups – Classroom Communication Resources (CCR), and the control group –

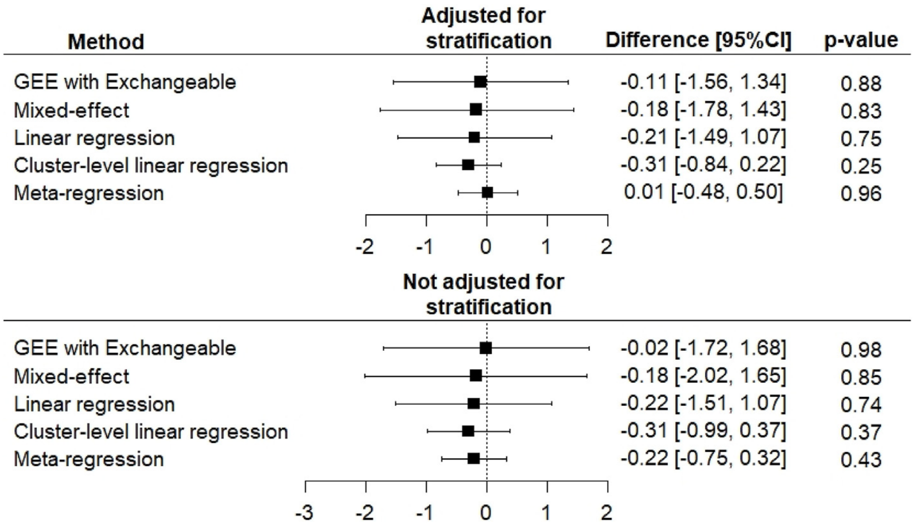

**Fig. 2.** Results of ITT analyses from different methods with and without adjustment for stratification.

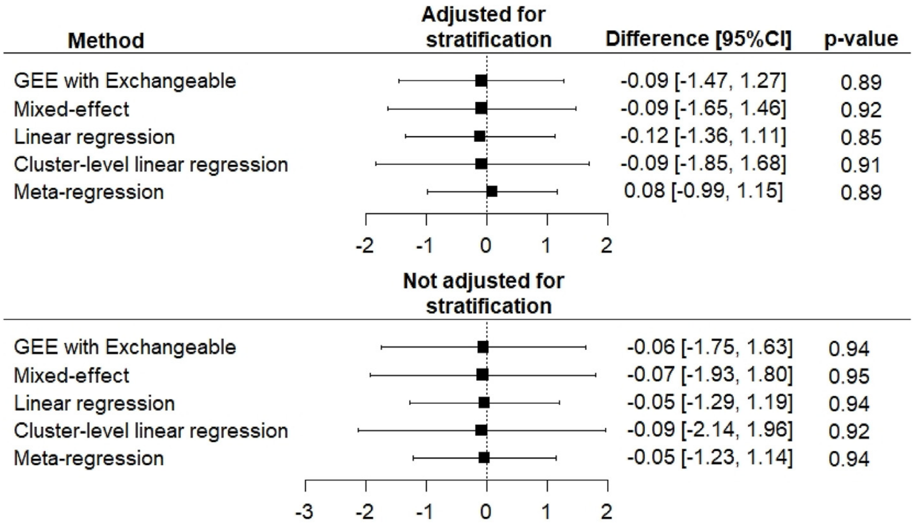

**Fig. 3.** Results of per-protocol analyses from different methods with and without adjustment for stratification.

usual care, in improving the peer attitude towards CWS.

The conclusion from the linear regression method was matched with other methods, but this method is not appropriate for analyzing data from CRT, as this method does not account the potential correlation among the outcomes from the same cluster. The meta-regression method yielded the narrowest 95% confidence intervals compared to other methods. There is very little variation among the summary measure, mean, of control and intervention groups in low and high stratum i.e. very low heterogeneity, which might lead to yield narrowest confidence interval for meta-regression since the width of the confidence interval decreases as the heterogeneity decreases [23]. Further, the direction of the estimated difference was opposite (positive) for this method compared to other methods when adjusted for stratification. The cluster-level linear regression yielded the widest confidence intervals for per-protocol approach, which are similar to the findings of Walter et al. [22]. However, for ITT approach, the cluster-level methods yielded narrower 95% confidence interval compared to individual-level methods. These results support the findings of Ukoumunne et al. [24] as the authors reported that the cluster-level method performed well, in case of binary data, when ICC is small.

The magnitudes of the estimated differences were similar among the methods with or without adjusted for stratification. However, the widths of the 95% confidence intervals were narrower for adjustment of stratification compared to without adjustment for stratification. These findings matched with the findings of Ma et al. [25] and Kahan et al. [26], where the authors compared several methods for analyzing binary data from stratified CRT and continuous data from stratified randomized controlled trial on individual, respectively. The p-values for all the methods were lower or similar when adjusted for stratification compared to the same method when not adjusted for stratification, which is in line with the findings of Kahan et al. [26].

The failure to adjust for clustering or centre in a multicentre trial results in inflated standard error and wider confidence interval [22,27]. Walters et al. [22] recommended to use cluster-level methods for number of cluster less than 15 per group as individual-level methods may not be reliable in this situation [28,29]. The estimates from the GEE and mixed-effect methods are connected through ICC [30] and in our case the estimates were similar due to the smaller ICC of 0.01.

We compared the results of five methods in several scenarios including: ITT and per-protocol analyses; with and without stratification; and account for potential correlation among the outcomes from the same cluster, which were pertaining to analyze continuous data from stratified CRTs. Moreover, we compared methods based on both individual-level and cluster-level summary data. Sensitivity analyses might help researchers to make informed decisions, since there is very limited guidance on which method is the best [14]. Furthermore, these analyses help to assess the sensitivity to conclusions to different scenarios such as, with or without clustering. However, we need to be cautious that, like binary data, the interpretation of the treatment effect using the marginal model and the mixed-effect model are may be different [31]. We only considered the multiple imputation technique to impute the missing data. Further investigation using other missing data imputation techniques are warranted.

Based on a simulation study on binary data it has been showed that, the statistical power of GEE is the highest compared to *t*-test, Wilcoxon rank sum test, permutation test, adjusted chi-square test and logistic random-effects model for the analysis of CRTs [32]. However, the estimated variance of from GEE is biased when the number of clusters is small for both binary and continuous data [33–35]. Researchers have reported the need for large number of clusters, 30–40 for mixed models and 40–50 for GEEs, in CRTs [1,36]. Also, some corrections have suggested - for mixed models corrections on degrees-of-freedom and for GEEs corrections to standard error estimations, for analyzing CRTs with small number of clusters [37–41]. Further studies are warranted to investigate how these corrections perform in the case of stratified cluster randomized trials.

## 5. Conclusion

We have empirically examined the sensitivity of five statistical methods for analyzing continuous outcome from stratified CRTs. The overall conclusions from all methods were similar i.e. no significant effect of the CCR intervention on improving the attitude of peers towards children who stutter. The adjustment for stratification yielded narrower standard errors and confidence intervals, thus it is important to adjust for stratification. Similarly, cluster-level methods yielded narrower confidence intervals compared to individual-level methods. However, further studies are warranted to assess the performance of these methods in wide ranging scenarios.

## Funding

There is no funding for this study.

## Conflicts of interest

All authors confirm that there are no known conflicts of interest associated with this study and there has been no financial support for this work that could have influenced its outcome.

## Acknowledgement

We wish to acknowledge the following for their contributions to the Mallick et al. study: (1) the University of Cape Town; (2) the SA National Research Fund (NRF); and (3) the Carnegie African Diaspora Fellowship Programme (CADFP).

## 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] A. Kroeger, E.V. Avila, L. Morison, Insecticide impregnated curtains to control domestic transmission of cutaneous leishmaniasis in Venezuela: cluster randomized trial, Br. Med. J. 325 (7368) (2002) 810–813.
[4] M. Jordhoy, P. Fayers, T. Saltnes, M. Ahlner-Elmqvist, M. Jannert, S. Kaasa, A palliative-care intervention and death at home: a cluster randomized trial, Lancet 356 (9233) (2000) 888–893.
[5] 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.
[6] N. Klar, A. Donner, The merits of matching in community intervention trials: a cautionary tale, Stat. Med. 16 (1997) 1753–1764.
[7] D. Murray, S. Varnell, J. Blitstein, Design and analysis of group-randomized trials: a review of recent methodological developments, Am. J. Public Health 94 (2004) 423–432.
[8] D. Hedeker, R. Gibbons, B. Flay, Random-effects regression models for clustered data with an example from smoking prevention research, J. Consult. Clin. Psychol. 62 (1994) 757–765.
[9] L. Zeger, K.-Y. Liang, P. Albert, Models for longitudinal data: a generalized estimating equation approach, Biometrics 44 (1988) 1049–1060.
[10] A. Whitehead, Meta-analysis of Controlled Clinical Trials, first ed., John Wiley and Sons, Chichester, 2002.
[11] A. Gould, Multi-centre trial analysis revisited, Stat. Med. 17 (15–16) (1998) 1779–1797 discussion 1799-800.
[12] A. Agresti, J. Hartzel, Strategies for comparing treatments on a binary response with multi-centre data, Stat. Med. 19 (8) (2000) 1115–1139.
[13] J. Fleiss, Analysis of data from multiclinic trials, Contr. Clin. Trials 7 (4) (1986) 267–275.
[14] 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.
[15] R. Mallick, H. Kathard, L. Thabane, M. Pillay, The Classroom Communication Resource (CCR) intervention to change peer's attitudes towards children who stutter (CWS): study protocol for a randomised controlled trial, Trials 19 (2018) 43.
[16] M. Campbell, D. Elbourne, D. Altman, CONSORT group, CONSORT statement: extension to cluster randomised trials, BMJ 328 (7441) (2004) 702–708.
[17] R Core Team, R: A Language and Environment for Statistical Computing, R Foundation for Statistical Computing, Vienna, Austria, 2018 URL https://www.R-project.org/.
[18] H. Brown, R. Kempton, The application of REML in clinical trials, Stat. Med. 13 (16) (1994) 1601–1617.
[19] R. McLean, W. Sanders, Approximating the degrees of freedom for SE's in mixed linear models, Proceedings of the Statistical Computing Section of the American Statistical Association. New Orleans, Louisiana, 1988.
[20] J. Twisk, Applied Longitudinal Data Analysis for Epidemiology: a Practical Guide, Cambridge University Press, 2003.
[21] K.-Y. Liang, L. Zeger, Longitudinal data analysis using generalized linear models, Biometrika 73 (1986) 13–22.
[22] S. Walters, C. Morrell, P. Slade, Analysing data from a cluster randomized trial (cCRT) in primary care: a case study, J. Appl. Stat. 38 (10) (2011) 2253–2269.
[23] Cochrane handbook for systematic reviews of interventions version 5.1.0 [updated March 2011], in: J.P.T. Higgins, S. Green (Eds.), The Cochrane Collaboration, 2011 Available from:www.handbook.cochrane.org.
[24] O. Ukoumunne, J. Carlin, M. Gulliford, A simulation study of odds ratio estimation for binary outcomes from cluster randomized trials, Stat. Med. 26 (18) (2007) 3415–3428.
[25] 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.
[26] B. Kahan, T. Morris, Reporting and analysis of trials using stratified randomisation in leading medical journals: review and reanalysis, BMJ 345 (2012) e5840.
[27] R. Chu, L. Thabane, J. Ma, A. Holbrook, E. Pullenayegum, P. Devereaux, Comparing methods to estimate treatment effects on a continuous outcome in multicentre randomized controlled trials: a simulation study, BMC Med. Res. Methodol. 11 (2011) 21.
[28] R. Hayes, L. Moulton, Cluster Randomised Trials, Chapman and Hall/CRC, Boca Raton, FL, 2009.
[29] A. Petrie, C. Sabin, Medical Statistics at a Glance, second ed., Blackwell, Oxford, 2005.
[30] M. Campbell, A. Donner, N. Klar, Developments in cluster randomized trials and statistics in medicine, Stat. Med. 26 (1) (2007) 2–19.
[31] P. FitzGerald, M. Knuiman, Use of conditional and marginal odds-ratios for analysing familial aggregation of binary data, Genet. Epidemiol. 18 (3) (2000) 193–202.
[32] P. Austin, A comparison of the statistical power of different methods for the analysis of cluster randomization trials with binary outcomes, Stat. Med. 26 (19) (2007) 3550–3565.
[33] R. Prentice, Correlated binary regression with covariates specific to each binary observation, Biometrics 44 (4) (1988) 1033–1048.
[34] L. Mancl, T. DeRouen, A covariance estimator for GEE with improved small-sample properties, Biometrics 57 (1) (2001) 126–134.
[35] 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.
[36] 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.
[37] M. Kenward, J. Roger, Small sample inference for fixed effects from restricted maximum likelihood, Biometrics 53 (1997) 983–997.
[38] M. Fay, B. Graubard, Small-sample adjustments for Wald-type tests using sandwich estimators, Biometrics 57 (2001) 1198–1206.
[39] L. Mancl, T. DeRouen, A covariance estimator for GEE with improved small-sample properties, Biometrics 57 (2001) 126–134.
[40] 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.
[41] 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.