HHS Public Access

Author manuscript

Stata J. Author manuscript; available in PMC 2022 March 23.

Published in final edited form as:

Stata J. 2020 June ; 20(2): 363–381. doi:10.1177/1536867x20931001.

xtgeebcv: A command for bias-corrected sandwich variance estimation for GEE analyses of cluster randomized trials

John A. Gallis,

Department of Biostatistics and Bioinformatics, Duke University, Duke Global Health Institute, Durham, NC

Fan Li,

Department of Biostatistics, Yale School of Public Health, New Haven, CT

Elizabeth L. Turner

Department of Biostatistics and Bioinformatics, Duke University, Duke Global Health Institute, Durham, NC

Abstract

Cluster randomized trials, where clusters (for example, schools or clinics) are randomized to comparison arms but measurements are taken on individuals, are commonly used to evaluate interventions in public health, education, and the social sciences. Analysis is often conducted on individual-level outcomes, and such analysis methods must consider that outcomes for members of the same cluster tend to be more similar than outcomes for members of other clusters. A popular individual-level analysis technique is generalized estimating equations (GEE). However, it is common to randomize a small number of clusters (for example, 30 or fewer), and in this case, the GEE standard errors obtained from the sandwich variance estimator will be biased, leading to inflated type I errors. Some bias-corrected standard errors have been proposed and studied to account for this finite-sample bias, but none has yet been implemented in Stata. In this article, we describe several popular bias corrections to the robust sandwich variance. We then introduce our newly created command, xtgeebcv, which will allow Stata users to easily apply finite-sample corrections to standard errors obtained from GEE models. We then provide examples to demonstrate the use of xtgeebcv. Finally, we discuss suggestions about which finite-sample corrections to use in which situations and consider areas of future research that may improve xtgeebcv.

john.gallis@duke.edu .

[7]Programs and supplemental materials

To install a snapshot of the corresponding software files as they existed at the time of publication of this article, type

```
. net sj 20-2
. net install st0599   (to install program files, if available)
. net get st0599       (to install ancillary files, if available)
```

Keywords

st0599; xtgeebcv; cluster randomized trials; bias-corrected variances; sandwich variance; generalized estimating equations; finite-sample correction

1 Introduction

The cluster randomized trial (CRT) is a study design used in many fields of research. In a CRT, randomization to intervention arms is carried out at the cluster level (for example, schools or clinics) and outcomes are assessed for each member of each cluster. The cluster randomization design is typically chosen when there is a high chance of treatment spillover across study arms, when the intervention is group based, or when individual randomization is not feasible (Turner et al. 2017a). For example, a recent trial in Ghana is evaluating an intervention designed to assist mothers with children that are under two years old to become more resilient and more effectively manage daily stress (Baumgartner 2018). The trial adopts a cluster randomized design because the intervention is designed to be delivered to groups of women. As another example, in the Thinking Healthy Program Peer-Delivered Plus study, the researchers recruited depressed women in their third trimester of pregnancy from 40 villages in Pakistan, with each village then being randomized to receive either the intervention or enhanced usual care (Sikander et al. 2015; Turner et al. 2016). Because this was a public health intervention delivered by community health workers, the risk of contamination (that is, the intervention being transmitted to women in the control group) would be too high if individual women were randomized, given that many of the women within each village live relatively close to one another.

Randomizing clusters instead of individuals poses unique challenges to the data analyses because the outcomes for members of the same cluster tend to be more similar than those for members of different clusters. The intraclass correlation coefficient (ICC) is a quantity that measures the degree of similarity for within-cluster observations and plays a central role in the design and analysis of CRTs (Murray 1998). Appropriate statistical methods used for trial analyses should properly reflect the within-cluster correlation and mainly include two classes of regression models: the cluster-specific (conditional) model and the population-averaged (marginal) model (Fitzmaurice, Laird, and Ware 2011). Although each modeling strategy has its own advantages, an important distinction between them is the difference in interpretation of the regression parameters (Preisser et al. 2003). A conditional model, such as the generalized linear mixed model, induces the within-cluster correlation through the latent random effects. Thus, the interpretation of the treatment effect is the average change in outcomes from control to intervention, conditional on the unobserved random effect. By contrast, marginal models separately specify a mean structure and a “working” correlation structure, and the interpretation of the corresponding treatment effect is the average change in outcomes due to intervention among the population defined by all participating clusters. Because CRTs are often conducted to evaluate public health intervention and inform policy decision, the marginal model carries a straightforward population-averaged interpretation and may be preferred (Li, Turner, and Preisser 2018). Furthermore, the estimation and inference of marginal models are often conducted through generalized estimating equations

(GEE) (Liang and Zeger 1986), a multivariate extension of the quasilikelihood inference (Wedderburn 1974).

In addition to straightforward interpretation of estimated model parameters, GEE maintains a robustness property in that the treatment-effects estimates are consistent even if the working correlation model deviates from the true correlation model. In this case, the sandwich variance estimator (Liang and Zeger 1986) remains consistent to the true variance. However, the approximate unbiasedness of the sandwich variance holds only when there are many clusters (a rule of thumb is ≥ 30, although this rule is sometimes given as ≥ 40 or even ≥ 50), whereas a frequent practical limitation of CRTs is that few clusters are available, because of resource constraints. In fact, a recent review by Fiero et al. (2016) found that, of the 86 studies included, about 50% randomized 24 or fewer clusters. In CRTs related to cancer published between 2002 and 2006, Murray et al. (2008) found similar results, with about 50% randomizing 24 or fewer clusters. Additionally, in their review of 300 CRTs published between 2000 and 2008, Ivers et al. (2011) found that, of the 285 studies reporting the number of clusters randomized, at least 50% randomized 21 or fewer clusters. Often, randomizing such few clusters is done because every cluster included in the study adds strain to limited financial and human resources. For example, in a study examining an intervention targeted at early childhood development among HIV-exposed children in Cameroon, only 10 total clusters were randomized because of resource and practical limitations (Baumgartner 2017).

When fewer than 30 to 40 clusters are randomized, the GEE sandwich variance estimator tends to be biased toward zero, leading to inflated type I error rates when testing for the intervention effect (Hayes and Moulton 2009). Proper analyses of CRTs should account for such finite-sample bias in variance estimation and adopt the bias-corrected variance estimator (Turner et al. 2017b). Several proposals for correcting such finite-sample bias have appeared in the statistical literature; see, for example, Mancl and DeRouen (2001); Kauermann and Carroll (2001); Fay and Graubard (2001) among others. These proposals have existed for over 15 years, but to our knowledge none has yet been implemented in Stata. Introducing the bias-corrected variance estimators to Stata has significant practical implications because Stata is a popular software tool for CRT analysts. The availability of this routine will help promote better statistical practice by allowing future analysts to report appropriate *p*-values and confidence intervals.

The remainder of this article is organized into four sections. In section 2, we introduce the theory of bias-corrected sandwich variance estimators for GEE analyses of CRTs. In section 3, we present our newly created command, xtgeebcv, which computes parameter estimates and bias-corrected variance in GEE models. In section 4, we present two examples of its use. We conclude in section 5 with recommendations to xtgeebcv users and ideas for future additions to the functionality of the program.

2 Statistical methods

2.1 GEE

We consider a parallel-arm CRT consisting of n clusters allocated into two intervention arms and note that the methods are generalizable to CRTs with more than two intervention arms. The outcome of each participant is typically measured at the end of the study and represented by Y_{ij} ($i = 1, \ldots, n, j = 1, \ldots, m_i$), where m_i is the number of individuals in cluster i. We denote the $p \times 1$ design vector by $\boldsymbol{X}_{ij}$, which includes 1 (intercept), the cluster-level binary indicator for treatment assignment, and possibly additional $p - 2$ baseline covariates. Note that, for CRTs with more than two arms, one could include additional dummy variables in the design vector $\boldsymbol{X}_{ij}$, and the following discussions remain unchanged. The marginal model parameterizes the marginal mean through a generalized linear model, $E(Y_{ij} \mid \boldsymbol{X}_{ij}) = \mu_{ij} = g^{-1}(\boldsymbol{X}_{ij}'\boldsymbol{\beta})$, where g is the link function and $\boldsymbol{\beta}$ is the p-vector of coefficients. The intervention effect is the component of $\boldsymbol{\beta}$ that corresponds to the treatment indicator. To characterize the similarity between individual responses within each cluster, we often employ the exchangeable working correlation so that $\text{corr}(Y_{ij}, Y_{ij'}) = \alpha$ for $j \neq j'$. The parameter α is interpreted as the ICC, a quantity that is vitally important for both the design and analysis of CRTs (Murray 1998). The exchangeable correlation structure is assumed for observations within the same cluster, while the observations from different clusters are assumed to be uncorrelated.

Let $\boldsymbol{Y}_i = (Y_{i1}, \ldots, Y_{im_i})'$ and $\boldsymbol{\mu}_i = (\mu_{i1}, \ldots, \mu_{im_i})'$ be the $m_i \times 1$ vector of outcomes and marginal means for cluster i, respectively, where m_i is the ith cluster size. The GEE method is used to estimate the parameter $\boldsymbol{\beta}$ from the marginal mean model with a specified working correlation matrix (Liang and Zeger 1986). We define $\boldsymbol{D}_i = \partial \boldsymbol{\mu}_i / \partial \boldsymbol{\beta}'$ and let $\boldsymbol{V}_i = \boldsymbol{A}_i^{1/2} \boldsymbol{R}_i \boldsymbol{A}_i^{1/2}$ be a working covariance matrix for $\boldsymbol{Y}_i$, where $\boldsymbol{A}_i$ is the m_i-dimensional diagonal matrix with elements $\phi v(\mu_{ij})$, ϕ is the dispersion parameter, and v is the variance function; $\boldsymbol{R}_i(\alpha)$ is a working correlation matrix whose dimension may vary across clusters but is specified by the common parameter α. With the exchangeable working correlation structure, we can succinctly write $\boldsymbol{R}_i(\alpha) = (1 - \alpha)\boldsymbol{I}_{m_i} + \alpha \boldsymbol{J}_{m_i}$, where $\boldsymbol{I}_{m_i}$ is the $m_i \times m_i$ identity matrix and $\boldsymbol{J}_{m_i}$ is an $m_i \times m_i$ matrix of ones. From the results given in Li, Turner, and Preisser (2018) and Li et al. (2019), $\boldsymbol{R}_i(\alpha)$ has two distinct eigenvalues, $\lambda_1 = 1 - \alpha$ and $\lambda_{i2} = 1 + (m_i - 1)\alpha$. Valid values of α guarantee a positive definite correlation matrix and can be easily determined from the set of linear constraints given by $\min\{\lambda_1, \lambda_{12}, \ldots, \lambda_{n2}\} > 0$. In other words, the plausible range of ICC is provided by $-\left(\max_{i=1}^{n}\{m_i\} - 1\right)^{-1} < \alpha < 1 \; \forall \; m_i \geq 2$.

The GEE estimators $\hat{\boldsymbol{\beta}}$, $\hat{\alpha}$, and $\hat{\phi}$ are jointly obtained by solving the set of estimating equations

$$\sum_{i=1}^{n} \boldsymbol{D}_i' \boldsymbol{V}_i^{-1} (\boldsymbol{Y}_i - \boldsymbol{\mu}_i) = \boldsymbol{0}$$

with a Newton-type algorithm implemented in the xtgee command. Furthermore, when the number of clusters is sufficiently large ($n \geq 30$), the variance–covariance of $\widehat{\boldsymbol{\beta}}$ can be consistently estimated by

$$\widehat{\Sigma} = \widehat{\boldsymbol{\Omega}}\left(\sum_{i=1}^{n} \boldsymbol{D}_i' \boldsymbol{V}_i^{-1} \boldsymbol{r}_i \boldsymbol{r}_i' \boldsymbol{V}_i^{-1} \boldsymbol{D}_i\right)\widehat{\boldsymbol{\Omega}} \tag{1}$$

where $\widehat{\Omega} = \left(\sum_{i=1}^{n} \boldsymbol{D}_i' \boldsymbol{V}_i^{-1} \boldsymbol{D}_i\right)^{-1}$ is the model-based variance (what Stata terms the "conventional" variance) and $\boldsymbol{r}_i = \boldsymbol{Y}_i - \widehat{\boldsymbol{\mu}}_i$ is the residual vector of cluster i. Equation (1) is referred to as the robust sandwich variance. Under mild regularity conditions, the sandwich variance estimator is consistent even if the correlation structure is misspecified (Liang and Zeger 1986). In practice, the sandwich variance is often preferred over the model-based variance (whose consistency is dictated by the correct specification of the working correlation) because of this robustness property.

2.2 Bias-corrected sandwich variance estimators

A practical limitation of CRTs is that fewer than 30 to 40 clusters are often randomized, mainly because of availability or resource constraints (Ivers et al. 2011; Fiero et al. 2016). When the number of clusters is small, it is known that the residuals, r_i, tend to be too small, and therefore the sandwich variance tends to underestimate the true variability of $\widehat{\boldsymbol{\beta}}$ (Mancl and DeRouen 2001). One simple correction is known as the degrees-of-freedom (DF) correction, defined as $\widehat{\boldsymbol{\Sigma}}_{\mathrm{DF}} = K\widehat{\boldsymbol{\Sigma}}/(K - p)$, where K is the number of clusters and p is the number of parameters. Such an ad hoc correction lacks theoretical motivation and does not provide satisfactory performance in empirical simulation studies designed to reflect characteristics expected in cluster randomized designs (Li and Redden 2015).[1] To improve finite-sample variance estimation, we consider four additional bias-corrected sandwich variance estimators that facilitate the implementation of the state-of-the-art recommendations for the analysis of CRTs (Li and Redden 2015; Ford and Westgate 2017).

Define the cluster leverage to be $\boldsymbol{H}_i = \boldsymbol{D}_i \widehat{\boldsymbol{\Omega}} \boldsymbol{D}_i' \boldsymbol{V}_i^{-1}$ (Preisser and Qaqish 1996). Kauermann and Carroll (2001) used the cluster-leverage-adjusted residuals to estimate the sandwich variance given by

$$\widehat{\boldsymbol{\Sigma}}_{\mathrm{KC}} = \widehat{\Omega}\left\{\sum_{i=1}^{n} \boldsymbol{D}_i' \boldsymbol{V}_i^{-1}\left(\boldsymbol{I}_{m_i} - \boldsymbol{H}_i\right)^{-1/2} \boldsymbol{r}_i \boldsymbol{r}_i' \left(\boldsymbol{I}_{m_i} - \boldsymbol{H}_i'\right)^{-1/2} \boldsymbol{V}_i^{-1} \boldsymbol{D}_i\right\}\widehat{\boldsymbol{\Omega}} \tag{2}$$

Because elements of $\boldsymbol{H}_i$ are between zero and one, $\widehat{\boldsymbol{\Sigma}}_{\mathrm{KC}}$ is expected to inflate the uncorrected sandwich variance $\widehat{\boldsymbol{\Sigma}}$. In practice, because the calculation of $(\boldsymbol{I} - \boldsymbol{H}_i)^{-1/2}$ tends to be unstable compared with $(\boldsymbol{I} - \boldsymbol{H}_i)^{-1}$, we approximate the summation within the curly brackets of (2) by

[1] We note that Stata allows a somewhat similar correction in xtgee but only for Gaussian distributions (that is, when the family(gaussian) option is specified) through the use of the rgf option. However, this correction multiplies the robust standard error by $(n - 1)/(n - p)$, where n is the number of individual observations rather than the number of clusters. So this "correction" does not match the DF correction as defined by Li and Redden (2015), or as implemented in our newly created command.

$$\left\{\sum_{i=1}^{n} D_i' V_i^{-1}\left(I_{m_i} - H_i\right)^{-1} r_i r_i' V_i^{-1} D_i + \sum_{i=1}^{n} D_i' V_i^{-1} r_i r_i' \left(I_{m_i} - H_i'\right)^{-1} V_i^{-1} D_i\right\}/2$$

Mancl and DeRouen (2001) devised a similar bias correction by using

$$\widehat{\Sigma}_{\mathrm{MD}} = \widehat{\Omega}\left\{\sum_{i=1}^{n} D_i' V_i^{-1}\left(I_{m_i} - H_i\right)^{-1} r_i r_i' \left(I_{m_i} - H_i'\right)^{-1} V_i^{-1} D_i\right\}\widehat{\Omega} \tag{3}$$

Because elements of the cluster leverage H_i are less than one, $\widehat{\Sigma}_{\mathrm{MD}}$ further inflates $\widehat{\Sigma}_{\mathrm{KC}}$. Fay and Graubard (2001) corrected the finite-sample bias in variance estimation by scaling the contribution from each cluster to the empirical variance

$$\widehat{\Sigma}_{\mathrm{FG}} = \widehat{\Omega}\left(\sum_{i=1}^{n} C_i D_i' V_i^{-1} r_i r_i' V_i^{-1} D_i C_i\right)\widehat{\Omega} \tag{4}$$

where $C_i = \mathrm{diag}([1-\min\{r,(Q_i)_{jj}\}]^{-1/2})$ and $Q_i = D_i' V_i^{-1} D_i \widehat{\Omega}$. The bound parameter $r < 1$ can be specified by the user but usually takes the default value 0.75 to avoid overcorrection of the bias. Finally, we implement the bias correction proposed by Morel, Bokossa, and Neerchal (2003). Their bias-corrected variance is given by

$$\widehat{\Sigma}_{\mathrm{MBN}} = \frac{(N-1)n}{(N-p)(n-1)}\widehat{\Omega}\left(\sum_{i=1}^{n} D_i' V_i^{-1} r_i r_i' V_i^{-1} D_i\right)\widehat{\Omega} + \delta_n \varphi \widehat{\Omega} \tag{5}$$

where $N = \sum_{i=1}^{n} m_i$ is the total sample size, $\delta_n = \min\{0.5, p/(n-p)\}$ is the correction factor that converges to zero as n increases to infinity, and

$$\varphi = \max\left[1, \mathrm{tr}\left\{\left(\sum_{i=1}^{n} D_i' V_i^{-1} r_i r_i' V_i^{-1} D_i\right)\widehat{\Omega}\right\}/p\right]$$

quantifies the design effect (Morel 1989). Of note, the additive bias correction (5) ensures a positive-definite covariance matrix, while the multiplicative bias corrections (2), (3), and (4) do not guarantee the positive definiteness of the estimated covariance (Morel, Bokossa, and Neerchal 2003), which was argued to be an additional benefit of (5). Once the variance estimator for the intervention effect is obtained using one of these bias-corrected variance formulas, we could conduct a test of no intervention effect by using the standard Wald z test or the Wald t test with DF $n - p$.

2.3 Computations with large cluster sizes

When the cluster sizes m_i become large (greater than 1,000), calculation of the bias-corrected variance estimators may become computationally inefficient because of numerical inversion of large matrices. To alleviate such a concern, we first note that a closed-form

expression is available for the inverse of the exchangeable correlation structure (Li, Turner, and Preisser 2018; Li et al. 2019) and is given by

$$\boldsymbol{R}^{-1}(\alpha) = \frac{1}{1-\alpha}\boldsymbol{I}_{m_i} - \frac{\alpha}{(1-\alpha)\{1+(m_i-1)\alpha\}}\boldsymbol{J}_{m_i}$$

Furthermore, Preisser, Qaqish, and Perin (2008) noted that inverting the asymmetric matrix $\boldsymbol{I}_{m_i} - \boldsymbol{H}_i$ is computationally demanding with large cluster sizes. Instead, they recommend working with its equivalent form $\left(\boldsymbol{V}_i - \boldsymbol{D}_i\widehat{\boldsymbol{\Omega}}\boldsymbol{D}_i'\right)\boldsymbol{V}_i^{-1}$ and efficiently calculate the inverse of the symmetric matrix $\boldsymbol{V}_i - \boldsymbol{D}_i\widehat{\boldsymbol{\Omega}}\boldsymbol{D}_i'$ by iteratively applying the Sherman–Morrison–Woodbury formula (Sherman and Morrison 1950; Henderson and Searle 1981). Preisser, Qaqish, and Perin (2008) demonstrated huge computational advantage of their algorithm over standard numeric inversions, and therefore we implement their algorithm in obtaining the multiplicative bias-correction factor $\left(\boldsymbol{I}_{m_i} - \boldsymbol{H}_i\right)^{-1}$ for $\widehat{\boldsymbol{\Sigma}}_{\mathrm{KC}}$ and $\widehat{\boldsymbol{\Sigma}}_{\mathrm{MD}}$. See Preisser, Qaqish, and Perin (2008) for additional computational details.

3 The xtgeebcv command

The xtgeebcv command was created to provide easy computation of finite-sample bias-corrected variances (hence the “bcv” in xtgeebcv) in Stata. In this section, we explain the available options in detail and examine the inner workings of the command.

The user should first specify a variable list (*varlist*) with an outcome (dependent) variable followed by predictor (independent) variables, just as one would do with the xtgee command. The user must tell xtgeebcv what the outcome variable and cluster indicator variable are by using the options outcome() and cluster(), respectively. Options are also available to specify the distribution family, link function, and type of finite-sample correction, as described in section 3.2.

Inside the command, the user-supplied data are passed to the xtgee command, with the command running xtset on the variable provided in the cluster() option before running xtgee. The xtgee command is specified with the option nmp. The nmp option tells xtgee to divide the scale parameter by $n - p$, where n is the number of clusters and p is the number of coefficients estimated. Although without the nmp option, Stata defaults to dividing only by n, $n - p$ is the form of the divisor used in Liang and Zeger (1986), so we use this option by default for the first set of output produced by xtgee, which reports the conventional (model-based) standard errors.

xtgeebcv allows use of either the independence or exchangeable working correlation matrices using the corr() option. Exchangeable is usually the most appropriate correlation structure to characterize the similarity between individual responses within each cluster in a cluster randomized design.

The design matrix, coefficient estimates, and variance–covariance matrix of the parameters output by the xtgee command are then passed to a mata command, which is used to compute

and output the desired finite-sample corrected standard errors of the parameter estimates. As described below, the option stderr() is used to specify which of five finite-sample bias-corrected standard errors ($\widehat{\boldsymbol{\Sigma}}_{\mathrm{DF}}$, $\widehat{\boldsymbol{\Sigma}}_{\mathrm{MD}}$, $\widehat{\boldsymbol{\Sigma}}_{\mathrm{FG}}$, $\widehat{\boldsymbol{\Sigma}}_{\mathrm{KC}}$, or $\widehat{\boldsymbol{\Sigma}}_{\mathrm{MBN}}$) to use for the output of standard errors, confidence intervals, and *p*-values.

3.1 Syntax

```
xtgeebcv varlist, outcome(varname) cluster(varname) [family(string)
link(string) stderr(string) statistic(string) corr(string) xtgee options]
```

varlist contains the regression specification: the dependent variable (outcome) followed by independent variables (predictors). Note that all categorical variables with more than two levels will need to be dummy coded by the user before supplying them to the command.

3.2 Options

outcome(*varname*) specifies the name of the outcome variable. outcome() is required.

cluster(*varname*) specifies the name of the cluster indicator variable. cluster() is required.

family(*string*) specifies the distributional family. The default is family(binomial).

link(*string*) specifies the link function. The following table gives more information on the available family() and link() combinations. The default depends on the specification of family(). The default for Gaussian, binomial, and Poisson are link(identity), link(logit), and link(log), respectively.

family()	link()
inomial	logit
binomial	log
binomial	identity
poisson	log
poisson	identity
gaussian	identity

stderr(*string*) gives the standard error to compute; the default is Kauermann–Carroll (stderr(kc)). The table below gives a complete list of specifications. Note that the robust standard errors provided by xtgeebcv will differ from Stata's default robust standard errors by a factor of $(K-1)/K$, where K is the number of clusters. This is because Stata automatically applies a correction of $K/(K-1)$ to the robust standard errors produced by xtgee when using the vce(robust) option. We do not follow this Stata-specific convention of applying this correction in this command, because 1) the robust sandwich variance of Liang and Zeger (1986) does not involve this correction; 2) this robust variance of Liang and Zeger (1986) is the one upon which the literature on bias-corrected sandwich variances is built (Mancl and DeRouen 2001; Kauermann and Carroll 2001; Fay and Graubard 2001); and 3)

other statistical software programs do not apply this $K/(K-1)$ correction to their robust standard errors. Thus, all the bias-corrected standard errors we implement in this command are based on the robust standard error without the $K/(K-1)$ correction.

string	Description
rb	Robust (sandwich) standard errors
df	DF correction
md	Mancl and DeRouen (2001) correction
fg	Fay and Graubard (2001) correction
kc	Kauermann and Carroll (2001) correction
mbn	Morel, Bokossa, and Neerchal (2003) correction

statistic(*string*) specifies the test. Specifying statistic(t) requests the Wald *t* test (the default). Alternatively, the user may specify statistic(z) to report the Wald *z* test instead of the Wald *t* test.

corr(*string*) specifies the type for the working correlation. The default is corr(exch) (the exchangeable correlation). The user may instead specify ind (the independent correlation matrix).

xtgee_options are any of the options documented in [XT] **xtgee**. For example, the option eform will provide exponentiated coefficients. Note that invoking the Stata command xtset (used to declare the clustering variable) is not necessary, because the command will automatically run xtset based on the variable supplied to the cluster() option.

4 Illustrative examples

In this section, we illustrate the use of xtgeebcv with two example datasets that are available to download along with the command. In the first example, we analyze synthetic data simulated from a CRT with clusters of equal size; in the second example, we analyze a real CRT evaluating the effect of a sexual health intervention on outcomes related to HIV.

4.1 Equal-sized clusters

First, we simulated correlated binary data using the method of Lunn and Davies (1998). We created a dataset with 80 clusters, 2 treatment arms (treatment and control), and exactly 14 individuals per cluster. The data were simulated so that the probability of outcome in the treatment group would be approximately 65%, while the probability in the control group would be 45%. This corresponds to a risk ratio of 1.44 or an odds ratio of 2.08, comparing treatment with control. After this, 20 clusters were randomly sampled from the dataset, 10 in treatment and 10 in control, to mimic a CRT with few clusters. To obtain an estimate of the risk ratio with Mancl–DeRouen finite-sample correction to the standard error, we use a log-binomial regression model by specifying a binomial distribution with a log-link function.

```
. use dat_sim
. xtgeebcv yij t, family(binomial) link(log) outcome(yij) cluster(cluster) >
stderr(md) statistic(z) eform nolog
Note: Family is binomial and link is log
Using exchangeable working correlation
with scale parameter divided by K - p
```

GEE population-averaged model		Number of obs	=	280
Group variable:	cluster	Number of groups	=	20
Link:	log	Obs per group:		
Family:	binomial	min	=	14
Correlation:	exchangeable	avg	=	14.0
		max	=	14
		Wald chi2(1)	=	4.62
Scale parameter:	1	Prob > chi2	=	0.0316

yij	exp(b)	Std. Err.	z	P>|z|	[95% Conf.	Interval]
treatment	1.460317	.257182	2.15	0.032	1.034044	2.062318
_cons	.45	.0663456	−5.42	0.000	.3370667	.6007713

```
Mancl-DeRouen bias-corrected standard errors
```

	exp(b)	Std. Err.	z	P>|z|	[95% Conf.	Interval]
treatment	1.460317	.3027435	1.83	0.068	.9727063	2.192365
_cons	.45	.0840296	−4.28	0.000	.3120797	.6488726

The first set of estimates comes from the GEE model with the scale parameter estimated using the $n - p$ DF, as discussed in section 3, and uses the conventional (model-based) standard errors. The second table gives the parameter estimates and Mancl–DeRouen corrected standard errors. We chose this bias correction because Lu et al. (2007) suggested that it performs adequately along with a z test if the number of clusters is in the range of 10 to 20.

The variance–covariance matrix of the parameter estimates for the chosen finite-sample correction is stored in e(V). All other variance–covariance matrices are stored in e(var*name*), where *name* is the name of the correction. Names of matrices can be retrieved using ereturn list.

```
. matrix list e(V)
```

```
symmetric e(V)[2,2]
           treatment     _cons
treatment  .04297887
    _cons   −.034869   .034869
. matrix list e(varfg)
symmetric e(varfg)[2,2]
```

	c1	c2
r1	.04054132	
r2	−.03236501	.03148758

```
. matrix list e(varkc)
symmetric e(varkc)[2,2]
```

	c1	c2
r1	.03868099	
r2	−.0313821	.0313821

Below, we also output the robust standard errors not multiplied by $K/(K-1)$, where K is the number of clusters.[2] Because the bias corrections are applied to this robust (sandwich) variance, we want to compare the standard-error estimates of the Mancl–DeRouen finite-sample correction with this robust variance, rather than with the conventional (model-based) standard-error estimates output from xtgee by default.

```
. xtgeebcv yij t, family(binomial) link(log) outcome(yij) cluster(cluster) >
stderr(rb) statistic(z) eform nolog
(output omitted)
Robust standard errors not multiplied by K/(K−1)
```

	exp(b)	Std. Err.	z	P>|z|	[95% Conf.	Interval]
treatment	1.460317	.2724691	2.03	0.042	1.013044	2.105069
_cons	.45	.0756266	−4.75	0.000	.3237131	.6255539

In this instance, if the researchers were using a strict 0.05 cutoff for significance, their conclusion about the statistical significance of the treatment effect would change if using the bias-corrected standard errors compared with the robust standard-error estimates.

[2] Multiplying by $K/(K-1)$ is the default in Stata when requesting robust standard errors in xtgee through the vce(robust) option. Please see the discussion of this point in section 3.2.

4.2 Unequal-sized clusters

In this section, we use data from the *MEME kwa Vijana* (MKV) CRT in Tanzania, which is described in Hayes and Moulton (2009, 23) and is also published in Ross et al. (2007). The data are publicly available online (Hayes and Moulton 2016). In brief, the goal of the trial was to evaluate the impact of a sexual health intervention on various HIV-related outcomes. The publicly available dataset includes data from male participants at follow-up, with the main outcome provided being "good knowledge of HIV acquisition", a binary variable. In this dataset, there are 20 communities that were randomized to receive either intervention or "standard activities". The number of participants per community ranges from 169 to 257, with a mean of 205 and a standard deviation of 26.3. The coefficient of variation of cluster sizes is 0.128. In this dataset, 65.3% of the intervention group has good knowledge of HIV acquisition at follow-up versus 44.9% in control, corresponding to an (unadjusted) odds ratio of 2.32 and risk ratio of 1.46.

The goal of the analysis is to estimate the odds ratio comparing intervention with control, while demonstrating the use of the Kauermann–Carroll finite-sample correction. In addition to including intervention group (arm) in the statistical model, we adjust for strata defined based on community HIV risk (three levels: high, medium, and low) on which the randomization was stratified (stratum, a community-level covariate with three levels, which is dummy coded before being included in the list of variables) and ethnic group (ethnicgp, a binary individual-level covariate).

```
. use mkvtrial, clear
. quietly tabulate stratum, generate(stratum)
. xtgeebcv know arm stratum2 stratum3 ethnicgp, family(binomial) link(logit)
> outcome(know) cluster(community) stderr(kc) eform nolog
Note: Family is binomial and link is logit
Using exchangeable working correlation
with scale parameter divided by K - p
```

GEE population-averaged model		Number of obs	=	4,100
Group variable:	community	Number of groups	=	20
Link:	logit	Obs per group:		
Family:	binomial	min	=	169
Correlation:	exchangeable	avg	=	205.0
		max	=	257
		Wald chi2(4)	=	43.75
Scale parameter:	1	Prob > chi2	=	0.0000

know	Odds Ratio	Std. Err.	z	P>|z|	[95% Conf.	Interval]
arm	2.286608	.338949	5.58	0.000	1.710079	3.057506

know	Odds Ratio	Std. Err.	z	P>|z|	[95% Conf.	Interval]
stratum2	1.051687	.1885727	0.28	0.779	.7400511	1.494552
stratum3	1.133454	.2181231	0.65	0.515	.7773161	1.652761
ethnicgp	.737854	.0648754	−3.46	0.001	.6210536	.8766209
_cons	.9892138	.1624665	−0.07	0.947	.7169527	1.364865

```
Note: _cons estimates baseline odds (conditional on zero random effects).
Kauermann-Carroll bias-corrected standard errors
t-statistic with K - p degrees of freedom
```

	exp(b)	Std. Err.	t	P>|t|	[95% Conf.	Interval]
arm	2.286608	.3808982	4.97	0.000	1.603225	3.261287
stratum2	1.051687	.2259479	0.23	0.818	.6652899	1.662501
stratum3	1.133454	.2373591	0.60	0.559	.7253637	1.771136
ethnicgp	.737854	.0759666	−2.95	0.010	.5924699	.9189134
_cons	.9892138	.1957989	−0.05	0.957	.6487352	1.508387

```
. xtgeebcv know arm stratum2 stratum3 ethnicgp, family(binomial) link(logit)
> outcome(know) cluster(community) stderr(rb) eform nolog
(output omitted)
Robust standard errors not multiplied by K/(K−1)
t-statistic with K - p degrees of freedom
```

	exp(b)	Std. Err.	t	P>|t|	[95% Conf.	Interval]
arm	2.286608	.3406765	5.55	0.000	1.664475	3.141277
stratum2	1.051687	.2018406	0.26	0.796	.6986018	1.583227
stratum3	1.133454	.2094824	0.68	0.508	.7644029	1.680681
ethnicgp	.737854	.0728592	−3.08	0.008	.5978122	.9107016
_cons	.9892138	.1760339	−0.06	0.952	.6769598	1.445498

In this case, with 20 clusters and many participants per cluster, although the finite-sample correction inflates the standard error by about 12% above the robust standard errors, any conclusion about significance of the effect based on the *p*-value would not change.

To see the potential impact of finite-sample corrections, suppose the researchers are interested in the intervention effect only in stratum 2. To this end, we subset the dataset to the 8 communities in stratum 2. This dataset has cluster sizes ranging from 187 to 243, with a mean of 214 and standard deviation of 21.1, which gives a coefficient of variation of cluster sizes of 0.099. In this dataset, 63.2% of the intervention group has good knowledge

of HIV acquisition at follow-up versus 45.7% in control. Because we have subset on the stratum, we no longer adjust for this variable.

```
. keep if stratum == 2
(2,388 observations deleted)
. xtgeebcv know arm ethnicgp, family(binomial) link(logit) outcome(know) >
cluster(community) stderr(kc) eform nolog
Note: Family is binomial and link is logit
Using exchangeable working correlation
with scale parameter divided by K - p
```

GEE population-averaged model		Number of obs	=	1,712
Group variable:	community	Number of groups	=	8
Link:	logit	Obs per group:		
Family:	binomial	min	=	187
Correlation:	exchangeable	avg	=	214.0
		max	=	243
		Wald chi2(2)	=	18.07
Scale parameter:	1	Prob > chi2	=	0.0001

know	Odds Ratio	Std. Err.	z	P>|z|	[95% Conf.	Interval]
arm	1.870034	.4094975	2.86	0.004	1.217459	2.872397
ethnicgp	.6190309	.0988164	−3.00	0.003	.4527249	.8464285
_cons	1.337813	.2828975	1.38	0.169	.8838899	2.02485

```
Note: _cons estimates baseline odds (conditional on zero random effects).
Kauermann-Carroll bias-corrected standard errors
t-statistic with K - p degrees of freedom
```

	exp(b)	Std. Err.	t	P>|t|	[95% Conf.	Interval]
arm	1.870034	.4815623	2.43	0.059	.964633	3.625241
ethnicgp	.6190309	.1072287	−2.77	0.039	.3965803	.9662591
_cons	1.337813	.4277156	0.91	0.404	.588128	3.043121

```
. xtgeebcv know arm ethnicgp, family(binomial) link(logit) outcome(know) >
cluster(community) stderr(rb) eform nolog
(output omitted)
```

```
Robust standard errors not multiplied by K/(K-1)
t-statistic with K - p degrees of freedom
```

	exp(b)	Std. Err.	t	P>|t|	[95% Conf.	Interval]
arm	1.870034	.4214707	2.78	0.039	1.047698	3.337819
ethnicgp	.6190309	.0959661	−3.09	0.027	.4155684	.9221089
_cons	1.337813	.3798498	1.03	0.352	.6447855	2.77572

From the GEE model with robust standard errors, we estimate an adjusted odds ratio of 1.87 (95% confidence interval [1.05, 3.34]). This estimate is significant at the 0.05 level. After we apply the Kauermann–Carroll bias correction to the robust standard errors, inflating the standard error of the intervention effect by 14.3%, the 95% confidence interval widens to [0.96,3.63]. The Kauermann–Carroll correction and the *t*-test statistic were chosen in this case given that Li and Redden (2015) suggested that they maintain close to the nominal type I error rate when the coefficient of variation of cluster sizes is less than 0.6. Compared with the *p*-value associated with the robust standard errors ($p = 0.039$), this estimate is not significant at the 0.05 level ($p = 0.059$).

5 Discussion

Many CRTs randomize fewer than 40 clusters, and cluster size is often highly variable. Many researchers use Stata to analyze their CRTs. Current GEE routines in Stata may not properly account for the small-sample bias in the robust standard errors and so may risk an inflated type I error rate when used in the analysis of small CRTs. We have introduced the xtgeebcv command to facilitate the analysis of CRTs with few clusters. This command is simple to use and does not require advanced programming skills, making it accessible to many researchers.

Although we have enabled the implementation of bias-corrected sandwich variance estimators in Stata, we have not attempted to make specific recommendations as to which correction works best in small CRTs. Several suggestions have been put forward in the statistical literature. For example, Li et al. (2017) found that the Wald *t* test with $\widehat{\boldsymbol{\Sigma}}_{\mathrm{KC}}$ carries the nominal type I error rate under both simple and constrained randomization designs with binary outcomes and equal cluster sizes. Lu et al. (2007) showed in a simulation study that the 95% Wald *z* confidence interval with $\widehat{\boldsymbol{\Sigma}}_{\mathrm{MD}}$ provides close to the nominal coverage when cluster sizes are balanced and the number of clusters is small to moderate (for example, 10 to 20). Li and Redden (2015) found that a Wald *t* test with $\widehat{\boldsymbol{\Sigma}}_{\mathrm{KC}}$ maintains the correct test size (that is, a type I error rate) when the coefficient of variation of cluster sizes is below 0.6, while a Wald *t* test with $\widehat{\boldsymbol{\Sigma}}_{\mathrm{FG}}$ maintains the nominal test size otherwise in small CRTs with binary outcomes. Ford and Westgate (2017) further demonstrated that the *t* test based on the average of $\widehat{\boldsymbol{\Sigma}}_{\mathrm{MD}}$ and $\widehat{\boldsymbol{\Sigma}}_{\mathrm{KC}}$ achieves the nominal test size in CRTs with both continuous and binary outcomes. These specific recommendations may be informative for analyzing small CRTs. In any case, as the bias-corrected sandwich variance becomes closer to the uncorrected variance with increasing numbers of clusters, it should preferably

always be reported along with the uncorrected sandwich variance as a sensitivity check. The investigation of finite-sample corrections in various small CRT settings is currently an area of active research, and our programs may also facilitate future simulation studies to generate recommendations specific to a research study.

There are some limitations to xtgeebcv. We have specifically designed xtgeebcv to accommodate the exchangeable working correlation structure most commonly used in parallel CRTs while also allowing for the simpler independent working correlation matrix. In more complex cluster randomization designs with multiple levels of clustering, nested exchangeable working correlation structures may be more appropriate (Li, Turner, and Preisser 2018; Li et al. 2019; Teerenstra et al. 2010), and we may extend our command accordingly as a next step. In terms of variance estimation in small CRTs, these authors have found that a z test with $\widehat{\boldsymbol{\Sigma}}_{\mathrm{MD}}$ or a t test with $\widehat{\boldsymbol{\Sigma}}_{\mathrm{KC}}$ carries a correct type I error rate in CRTs, although the former generally requires many clusters (at least 20) to work well. On the other hand, the extension requires additional efforts because estimating more than one correlation parameter requires an additional set of estimating equations (Prentice 1988; Preisser, Qaqish, and Perin 2008) and is not accommodated by standard xtgee routines. Another future extension of our command is to incorporate the first-order autoregressive correlation structure to enable the appropriate analysis of longitudinal studies with a limited number of subjects. The GEE analysis of longitudinal data is generally similar to the analysis of CRTs, although the cluster size (defined as the number of repeated measurements per individual) is frequently much smaller than that in CRTs, and finite-sample corrections may require additional considerations. Recent empirical studies (Ford and Westgate 2018; Wang et al. 2016) have already found that bias-corrected variance works reasonably well in this setting, so such an extension is an important avenue for future research.

Acknowledgments

The authors would like to thank Alyssa Platt, Joe Egger, and Ryan Simmons of the Duke Global Health Institute Research Design and Analysis Core for testing and providing feedback on the programs. We would also like to thank an anonymous reviewer whose comments on a previous version of this manuscript helped improve the final version. This research was funded in part by National Institutes of Health grant R01 HD075875 (principal investigator [PI]: Dr. Joanna Maselko). In addition, the development of the command xtgeebcv was partly inspired by the studies Evaluation of an Early Childhood Development Intervention for HIV-Exposed Children in Cameroon (PI: Dr. Joy Noel Baumgartner); Evaluation of the iMBC/ECD model on maternal mental health and child development in Kenya (PI: Dr. Joy Noel Baumgartner); and Evaluation of the iMBC/ECD Model in Ghana (PI: Dr. Joy Noel Baumgartner), funded by Catholic Relief Services.

About the authors

John A. Gallis, ScM, currently works as a biostatistician at Duke University in the Department of Biostatistics and Bioinformatics and at the Duke Global Health Institute. His research interests include the design and analysis of CRTs and the analysis of data with other forms of clustering, including longitudinal and spatial data.

Fan Li, PhD, is an assistant professor in the Department of Biostatistics and in the Center for Methods in Implementation and Prevention Science at Yale School of Public Health. His primary research interests include statistical methodology for the design and analysis of

CRTs, causal inference for observational studies, longitudinal and spatial data analysis, and Bayesian methods.

Elizabeth L. Turner, PhD, is an associate professor in the Department of Biostatistics and Bioinformatics and in the Duke Global Health Institute at Duke University. Her primary research interest is the design and analysis of CRTs, with a special focus on translating methods to be accessible to the practitioner. She heads the Duke Global Health Institute Research Design and Analysis Core and has led the design and analysis of a range of CRTs in global mental health, malaria, and cardiovascular disease in multiple settings around the world, including in Kenya, Tanzania, Nepal, China, and Pakistan.

8 References

Baumgartner JN 2017. Evaluation of an early childhood development intervention for HIV-exposed children in Cameroon. https://clinicaltrials.gov/ct2/show/NCT03195036.

——. 2018. Evaluation of the iMBC/ECD model in Ghana. https://clinicaltrials.gov/ct2/show/NCT03665246.

Fay MP, and Graubard BI. 2001. Small-sample adjustments for Wald-type tests using sandwich estimators. Biometrics 57: 1198–1206. 10.1111/j.0006-341X.2001.01198.x. [PubMed: 11764261]

Fiero MH, Huang S, Oren E, and Bell ML. 2016. Statistical analysis and handling of missing data in cluster randomized trials: A systematic review. Trials 17: 72. 10.1186/s13063-016-1201-z. [PubMed: 26862034]

Fitzmaurice GM, Laird NM, and Ware JH. 2011. Applied Longitudinal Analysis. 2nd ed. Hoboken, NJ: Wiley.

Ford WP, and Westgate PM. 2017. Improved standard error estimator for maintaining the validity of inference in cluster randomized trials with a small number of clusters. Biometrical Journal 59: 478–495. 10.1002/bimj.201600182. [PubMed: 28128854]

——. 2018. A comparison of bias-corrected empirical covariance estimators with generalized estimating equations in small-sample longitudinal study settings. Statistics in Medicine 37: 4318–4329. 10.1002/sim.7917. [PubMed: 30073684]

Hayes R, and Moulton L. 2016. Datasets from the book Cluster Randomised Trials by Hayes & Moulton. Harvard Dataverse. 10.7910/DVN/YXMQZM.

Hayes RJ, and Moulton LH. 2009. Cluster Randomised Trials. Boca Raton, FL: Chapman & Hall/CRC.

Henderson HV, and Searle SR. 1981. On deriving the inverse of a sum of matrices. SIAM Review 23: 53–60. 10.1137/1023004.

Ivers NM, Taljaard M, Dixon S, Bennett C, McRae A, Taleban J, Skea Z, Brehaut JC, Boruch RF, Eccles MP, Grimshaw JM, Weijer C, Zwarenstein M, and Donner A. 2011. Impact of CONSORT extension for cluster randomised trials on quality of reporting and study methodology: Review of random sample of 300 trials, 2000–8. British Medical Journal 343: d5886. 10.1136/bmj.d5886. [PubMed: 21948873]

Kauermann G, and Carroll RJ. 2001. A note on the efficiency of sandwich covariance matrix estimation. Journal of the American Statistical Association 96: 1387–1396. 10.1198/016214501753382309.

Li F, Forbes AB, Turner EL, and Preisser JS. 2019. Power and sample size requirements for GEE analyses of cluster randomized crossover trials. Statistics in Medicine 38: 636–649. 10.1002/sim.7995. [PubMed: 30298551]

Li F, Turner EL, Heagerty PJ, Murray DM, Vollmer WM, and DeLong ER. 2017. An evaluation of constrained randomization for the design and analysis of group-randomized trials with binary outcomes. Statistics in Medicine 36: 3791–3806. 10.1002/sim.7410. [PubMed: 28786223]

Li F, Turner EL, and Preisser JS. 2018. Sample size determination for GEE analyses of stepped wedge cluster randomized trials. Biometrics 74: 1450–1458. 10.1111/biom.12918. [PubMed: 29921006]

Li P, and Redden DT. 2015. Small sample performance of bias-corrected sandwich estimators for cluster-randomized trials with binary outcomes. Statistics in Medicine 34: 281–296. 10.1002/sim.6344. [PubMed: 25345738]

Liang K-Y, and Zeger SL. 1986. Longitudinal data analysis using generalized linear models. Biometrika 73: 13–22. 10.1093/biomet/73.1.13.

Lu B, Preisser JS, Qaqish BF, Suchindran C, Bangdiwala SI, and Wolfson M. 2007. A comparison of two bias-corrected covariance estimators for generalized estimating equations. Biometrics 63: 935–941. 10.1111/j.1541-0420.2007.00764.x. [PubMed: 17825023]

Lunn AD, and Davies SJ. 1998. A note on generating correlated binary variables. Biometrika 85: 487–490. 10.1093/biomet/85.2.487.

Mancl LA, and DeRouen TA. 2001. A covariance estimator for GEE with improved small-sample properties. Biometrics 57: 126–134. 10.1111/j.0006-341x.2001.00126.x. [PubMed: 11252587]

Morel JG 1989. Logistic regression under complex survey designs. Survey Methodology 15: 203–223.

Morel JG, Bokossa MC, and Neerchal NK. 2003. Small sample correction for the variance of GEE estimators. Biometrical Journal 45: 395–409. 10.1002/bimj.200390021.

Murray DM 1998. Design and Analysis of Group-Randomized Trials. New York: Oxford University Press.

Murray DM, Pals SL, Blitstein JL, Alfano CM, and Lehman J. 2008. Design and analysis of group-randomized trials in cancer: A review of current practices. Journal of the National Cancer Institute 100: 483–491. 10.1093/jnci/djn066. [PubMed: 18364501]

Preisser JS, and Qaqish BF. 1996. Deletion diagnostics for generalised estimating equations. Biometrika 83: 551–562. 10.1093/biomet/83.3.551.

Preisser JS, Qaqish BF, and Perin J. 2008. A note on deletion diagnostics for estimating equations. Biometrika 95: 509–513. 10.1093/biomet/asn019.

Preisser JS, Young ML, Zaccaro DJ, and Wolfson M. 2003. An integrated population-averaged approach to the design, analysis and sample size determination of cluster-unit trials. Statistics in Medicine 22: 1235–1254. 10.1002/sim.1379. [PubMed: 12687653]

Prentice RL 1988. Correlated binary regression with covariates specific to each binary observation. Biometrics 44: 1033–1048. 10.2307/2531733. [PubMed: 3233244]

Ross DA, Changalucha J, Obasi AI, Todd J, Plummer ML, Cleophas-Mazige B, Anemona A, Everett D, Weiss HA, Mabey DC, Grosskurth H, and Hayes R. 2007. Biological and behavioural impact of an adolescent sexual health intervention in Tanzania: A community-randomized trial. AIDS 21: 1943–1955. 10.1097/QAD.0b013e3282ed3cf5. [PubMed: 17721102]

Sherman J, and Morrison WJ. 1950. Adjustment of an inverse matrix corresponding to a change in one element of a given matrix. Annals of Mathematical Statistics 21: 124–127. 10.1214/aoms/1177729893.

Sikander S, Lazarus A, Bangash O, Fuhr DC, Weobong B, Krishna RN, Ahmad I, Weiss HA, Price L, Rahman A, and Patel V. 2015. The effectiveness and cost-effectiveness of the peer-delivered Thinking Healthy Programme for perinatal depression in Pakistan and India: The SHARE study protocol for randomised controlled trials. Trials 16: 534. 10.1186/s13063-015-1063-9. [PubMed: 26604001]

Teerenstra S, Lu B, Preisser JS, van Achterberg T, and Borm GF. 2010. Sample size considerations for GEE analyses of three-level cluster randomized trials. Biometrics 66: 1230–1237. 10.1111/j.1541-0420.2009.01374.x. [PubMed: 20070297]

Turner EL, Li F, Gallis JA, Prague M, and Murray DM. 2017a. Review of recent methodological developments in group-randomized trials: Part 1—Design. American Journal of Public Health 107: 907–915. 10.2105/AJPH.2017.303706. [PubMed: 28426295]

Turner EL, Prague M, Gallis JA, Li F, and Murray DM. 2017b. Review of recent methodological developments in group-randomized trials: Part 2—Analysis. American Journal of Public Health 107: 1078–1086. 10.2105/AJPH.2017.303707. [PubMed: 28520480]

Turner EL, Sikander S, Bangash O, Zaidi A, Bates L, Gallis J, Ganga N, O'Donnell K, Rahman A, and Maselko J. 2016. The effectiveness of the peer delivered Thinking Healthy Plus (THPP+) Programme for maternal depression and child socio-emotional development in Pakistan:

Study protocol for a three-year cluster randomized controlled trial. Trials 17: 442. 10.1186/s13063-016-1530-y. [PubMed: 27608926]

Wang M, Kong L, Li Z, and Zhang L. 2016. Covariance estimators for generalized estimating equations (GEE) in longitudinal analysis with small samples. Statistics in Medicine 35: 1706–1721. 10.1002/sim.6817. [PubMed: 26585756]

Wedderburn RWM 1974. Quasi-likelihood functions, generalized linear models, and the Gauss–Newton method. Biometrika 61: 439–447. 10.2307/2334725.