Research Article

Statistics in Medicine

Received 10 January 2011, Accepted 1 November 2011 Published online 23 February 2012 in Wiley Online Library

(wileyonlinelibrary.com) DOI: 10.1002/sim.4471

Augmented generalized estimating equations for improving efficiency and validity of estimation in cluster randomized trials by leveraging cluster-level and individual-level covariates

Alisa J. Stephens,*† Eric J. Tchetgen Tchetgen and Victor De Gruttola

Recent methodological advances in covariate adjustment in randomized clinical trials have used semiparametric theory to improve efficiency of inferences by incorporating baseline covariates; these methods have focused on independent outcomes. We modify one of these approaches, augmentation of standard estimators, for use within cluster randomized trials in which treatments are assigned to groups of individuals, thereby inducing correlation. We demonstrate the potential for imbalance correction and efficiency improvement through consideration of both cluster-level covariates and individual-level covariates. To improve small-sample estimation, we consider several variance adjustments. We evaluate this approach for continuous and binary outcomes through simulation and apply it to data from a cluster randomized trial of a community behavioral intervention related to HIV prevention in Tanzania. Copyright © 2012 John Wiley & Sons, Ltd.

Keywords: cluster randomized trials; covariate adjustment; double robustness; generalized estimating equations; HIV intervention; longitudinal data

1. Introduction

1.1. Traditional and cluster randomized trials

Randomized clinical trials (RCTs) are recognized as the gold standard in medical research for evaluating new treatments. Cluster or group randomized trials (GRTs), which assign treatment to groups of individuals, are advantageous when interaction among subjects within a group may impact their respective outcomes. GRTs are therefore especially relevant for assessing prevention and treatment methods for infectious diseases, where subjects within a geographical unit such as a neighborhood, school, or workplace may infect each other. For example, in vaccine studies, a subject's vaccination status may impact health outcomes not only for that subject but for others as well. Clustered designs also have the advantage of reducing the potential for contamination of effects caused by sharing of information or medication between treated and control subjects. Similarly, group treatment assignment can enhance compliance as subjects within a group are given the same regimen to follow. In some cases, the intervention may be administered at the cluster level, such as in studies involving schools or medical practices. Klar and Donner provided several examples of intervention trials in which groups were randomized for medical, political, or logistical reasons [1].

Department of Biostatistics, Harvard School of Public Health, Boston, MA, U.S.A.
**Correspondence to: Alisa J. Stephens, Department of Biostatistics, Harvard School of Public Health, Boston, MA, U.S.A.*
†*E-mail: astephen@hsph.harvard.edu*

Although intervention is assigned at the group level, interest often lies in performing inference on the individual. Generally, subjects within a group are expected to be more similar than subjects in different groups, inducing dependence across study subjects. Cluster randomized designs thus present the additional challenge of accounting for correlation among group members. Standard approaches for estimating treatment effects when responses are correlated include maximum likelihood for generalized linear mixed models (GLMM) and generalized estimating equations (GEE) for restricted mean models [2, 3]. To estimate the marginal effect in a binary treatment setting, one typically fits a model including an intercept and treatment term. The relevant GLMM is defined by the model $E(Y_{ij}|A_i, b_i) = g(\beta_0 + \beta_{1^c} A_i + b_i)$, where Y_{ij} denotes the outcome for the jth individual in the ith cluster, A_i is an indicator for treatment, b_i is a random effect inducing correlation among subjects within a cluster, and g(.) is a monotone link function. The outcome Y_{ij} and random effect b_i are assumed to follow a particular distribution. We note here that β_{1^c} is interpreted as a cluster-specific treatment effect but marginalizes over all other covariates. In the analogous GEE approach, estimating equations are constructed following the mean model

$$E(Y_{ij}|A_i) = g(A_i;\beta) = g(\beta_0 + \beta_1 A_i), \tag{1}$$

where correlation is accounted for by incorporating a working covariance matrix $\mathbf{V_w}$. For cluster randomized designs, independence or exchangeable structure is generally assumed. An advantage of the GEE approach is that consistency of $\hat{\beta}$, the estimate of β, only requires that the mean $g(A_i;\beta)$ is correctly specified, in which case, $\hat{\beta}$ is asymptotically normal for all $\mathbf{V_w}$ and efficient when $\mathbf{V_w}$ takes the true form of $\mathbf{V}$, the variance of response vector $\mathbf{Y_i}$. We review the exact form of the GEE in the following section. GEE differ from maximum likelihood estimation in mixed models by treating correlation as a nuisance parameter. Additionally, GLMM require full specification of the distribution of Y_{ij}, whereas GEE follow from semiparametric theory and only specify the first moment of Y_{ij} while requiring the second moment to be finite. Unlike GLMM, GEE do not make any assumptions about cluster effects and thus provide a population-averaged effect estimate in contrast to the GLMM cluster-specific estimate. In either approach, treatment is evaluated through inference on β_1.

A second challenge presented by cluster randomized designs is that the number of available experimental units may be fairly small. Inference for model-based methods relies on asymptotic theory, which may not be applicable in trials with relatively few clusters. For GEE, several studies have shown that the sandwich variance estimator typically underestimates the variability of parameter estimates and consequently results in inference that is too liberal [4]. Researchers [5–9] have proposed a number of adjustment methods for small sample analysis. These adjustments generally take one of two strategies; they account for the variability in the sandwich estimator or correct for its small-sample bias. None of these methods have been uniformly adopted.

The number of available experimental units also affects the degree to which randomization successfully balances baseline characteristics across treatment groups. RCTs with large sample sizes assure a reasonable degree of balance in covariate profiles with high probability, but GRTs often have smaller numbers of experimental units and therefore provide less assurance of balance [10]. GRTs are also likely to contain subject heterogeneity in cluster-level and individual-level characteristics that can influence estimated treatment effects. Clustered designs therefore require methods that permit controlling for imbalances at the cluster and subject levels.

1.2. Methods for covariate adjustment in randomized trials

Traditionally, adjustment for residual imbalance has been achieved by adding covariates Z_i, X_{ij} to a model for the effect of treatment on some outcome. The adjusted model for Y_{ij} is defined by $E(Y_{ij}|A_i, Z_i, X_{ij}) = g(\beta_0 + \beta_{1^\star} A_i + \beta_Z Z_i + \beta_X X_{ij})$, where Z_i is a vector of covariates shared by all subjects within the ith cluster and X_{ij} is a subject-specific vector of measurements. Standard approaches such as mixed models and GEE can incorporate adjustment at both levels. With the exception of linear and log-linear models, the conditional model differs from the marginal model (1) in the interpretation of $\beta_{1^\star}$. Inference on $\beta_{1^\star}$ is also affected by the presence of baseline covariates. For uncorrelated continuous outcomes and an identity link function relating covariates to the mean, it has been shown that when X and Y are correlated, $\beta_{1^\star}$ is more precise than the unadjusted estimator [11, 12]. No direct relationship

between the efficiency of β_1 (1) and $\beta_{1\star}$ has been established for nonlinear models [13] or correlated outcomes. To provide an alternative that makes fewer parametric assumptions, Gail *et al.* [14] proposed a permutation approach to covariate adjustment in GRTs. Parametric models are used for adjustment, and permutation inference is conducted on the cluster-averaged model-based residuals. Permutation tests are guaranteed to be valid even for small samples, unlike modeling approaches. Braun and Feng [15] developed a similar model-based permutation approach using an optimally weighted combination of residuals.

Recent methodological developments in covariate adjustment for RCTs include van der Laan's targeted maximum likelihood [16] and Tsiatis' augmentation approach [12, 17]. These methods adapt semiparametric theory developed by Robins [18] and Robins *et al.* [19] for observational studies with time-varying exposures and missing data problems, respectively. RCTs may be conceptualized theoretically in either framework, with counterfactual outcomes under the treatment not received considered missing, or as observational studies with a known probability of point exposure. Robins *et al.* [19] and Robins [18] characterized the efficient influence function in these settings. Van der Laan and Tsiatis solved the set of estimating functions determined by the efficient score using two different approaches, which are equivalent in the absence of model misspecification.

Targeted maximum likelihood estimation (tMLE) is an iterative procedure that involves adding a cleverly defined covariate to standard regression models. Upon convergence, the tMLE estimator solves the efficient influence function for the parameter of interest, resulting in bias reduction and efficiency improvement relative to maximum likelihood. tMLE is currently available for independent binary, continuous, time-to-event outcomes [20, 21], and most recently, clustered or longitudinal outcomes. Tsiatis' approach involves directly solving a set of augmented estimating equations determined by the efficient influence function for the marginal treatment effect [17]. Researchers [12, 17, 22, 23] have explored this method for continuous, binary, and discrete survival outcomes. Current applications of the augmentation method have focused on independent outcomes, with the exception of a simulation study based on the linear mixed model [17].

Whereas tMLE simultaneously uses baseline covariates from treatment and control groups to target treatment effect estimation, Tsiatis' method separates covariate adjustment and treatment evaluation. It also has the added advantage of allowing separate adjustment for baseline covariates within treatment arms. If carried out by separate statistical groups that do not share data, this approach reduces the risk that adjustment models are chosen to yield the most significant result. Even without decoupling of adjustment and treatment effect estimation, all covariate adjustment methods can be made objective by prespecifying the adjustment strategy.

2. The simple augmented generalized estimating equation

This section demonstrates the use of augmented estimating equations in analyses of cluster randomized trials. In such a trial, m clusters of size n_i, $i = 1, \ldots, m$, are randomized to either treatment ($A_i = 1$) or control ($A_i = 0$) with probability $P(A_i = 1) = \pi$. To motivate the augmented GEE, we first review standard GEE. Let Y_{ij} denote the response for the jth individual in the ith cluster. $\mathbf{Y_i} = (Y_{i1}, Y_{i2}, \ldots, Y_{in_i})^{\mathrm{T}}$, where n_i is the number of subjects within the ith cluster. GEE for the marginal treatment effect are defined by the mean model (1), where β is a p-dimensional parameter. An estimator for β is obtained by solving the estimating equations

$$\sum_{i=1}^{m} \psi_i(\mathbf{Y}, A; \beta) = \sum_{i=1}^{m} \mathbf{D_i}^{\mathrm{T}} \mathbf{V_i}^{-1} \{\mathbf{Y_i} - \mathbf{g}(A_i; \beta)\} = \mathbf{0}, \tag{2}$$

where $\mathbf{D_i} = \frac{\partial \mathbf{g}(A_i;\beta)}{\partial \beta^{\mathrm{T}}}$, $\mathbf{V_i} = \mathbf{V}_{\phi}^{1/2} \mathbf{R}\{\alpha(A_i)\} \mathbf{V}_{\phi}{}^{1/2}$, and $\mathbf{g}(A_i; \beta)$ denotes the n_i-dimensional link function for the outcome vector $\mathbf{Y_i}$. Covariance matrix $\mathbf{V_i}$ is determined by the $n_i \times n_i$ matrix function $v(A_i)$. The variance function $v(A_i)$ is a product of the diagonal matrix $\mathbf{V}_{\phi}$, where $V_{\phi_{i,i}}$ is the variance of Y_{ij} and correlation matrix $\mathbf{R}\{\alpha(A_i)\}$, where we allow α to be treatment specific. This differs from the usual presentation of GEE, in which $\mathbf{V_i}$ is constant and does not depend on A_i. Because our model does not

place any restrictions on $\mathbf{V_i}$, we generalize the usual approach to allow $\mathbf{V_i}$ to be more flexible. Variance parameters ϕ and α_k, where k indexes treatment, are estimated by the method of moments using $\hat{\beta}_{\text{init}}$, an initial estimator of β. To recover the GEE fit in standard software, the aforementioned expressions simplify such that $v(A_i) = v(1) = v(0) = \mathbf{V}$ and a single correlation parameter α is estimated across all clusters. In a slight abuse of notation, we take $\mathbf{V_i}$ to be the matrix function $v(A_i)$ and $\mathbf{V}$ the constant variance matrix.

For continuous outcomes with the identity link $\mathbf{g}(A_i;\beta) = \mathbf{A}_i^{\star}\beta$, where $\mathbf{A}_i^{\star}$ is the $n_i \times 2$ design matrix composed of rows $(1, A_i)$, the solution to (2), $\hat{\beta}$, exists in closed form, with

$$\hat{\beta} = \left(\sum_{i=1}^{m} \mathbf{A}_i^{\star\mathrm{T}} \mathbf{V_i}^{-1} \mathbf{A}_i^{\star}\right)^{-1} \left(\sum_{i=1}^{m} \mathbf{A}_i^{\star\mathrm{T}} \mathbf{V_i}^{-1} \mathbf{Y_i}\right).$$

For simple designs, a closed-form solution for $\hat{\beta}$ can also be derived for non-identity link functions $g(A_i;\beta)$ using the discreteness of A. The solution to (2) for the logit link is given in Appendix A of the Supplementary material.[‡] Generally, for more complex models, GEE coefficient estimates are found using an iterative procedure such as the Newton–Raphson method or iteratively reweighted least squares.

Robins *et al.* [19] and Robins [18] established that in a model for data $O = (Y, A, X)$ in which $\pi_k = P(A = k|X)$ is known, any regular and asymptotically linear estimator for β can be found as the solution to $\sum_i \psi_{i_{\text{aug}}}(Y, A, X; \bar{\gamma}_K) = \mathbf{0}$ for a specific choice of $\bar{\gamma}_K(X)$. Zhang *et al.* [17] demonstrated the use of this theory in RCTs with univariate outcomes. With the application of these results to multivariate settings, $\hat{\beta}$ may be improved by augmenting the standard GEE with a function of baseline covariates X. The general form of the augmented GEE for a K-level treatment is

$$\sum_{i=1}^{m} \psi_{i_{\text{aug}}}(\mathbf{Y}, A, \mathbf{X}; \beta, \bar{\gamma}_K) = \sum_{i=1}^{m} \left[\mathbf{D_i}^{\mathrm{T}} \mathbf{V_i}^{-1} \{\mathbf{Y_i} - \mathbf{g}(A_i;\beta)\} - \sum_{k=1}^{K} \{I(A_i = k) - \pi_k\}\gamma_k(\mathbf{X_i})\right] = \mathbf{0}, \quad (3)$$

where $\gamma_k(\mathbf{X_i})$ is a p-dimensional function of $\mathbf{X_i}$.

It was further shown that, for the class of estimating functions $\{\psi_{\text{aug}}(\bar{\gamma}_K) : \bar{\gamma}_K \in \Gamma_K\}$, where Γ_K is the set of all functions of X such that $E[\psi_{\text{aug}}(\bar{\gamma}_K)^{\mathrm{T}}\psi_{\text{aug}}(\bar{\gamma}_K)] < \infty$, the optimal estimator within this class for a fixed $\psi(Y, A;\beta)$ is obtained by setting $\gamma_{k_{\text{opt}}}(X_i) = E\{\psi_i(Y, A;\beta)|A_i = k, X_i\}$ [17–19]. When only two treatment arms are considered, the augmentation term $\sum_{k=1}^{K}\{I(A_i = k) - \pi_k\}\gamma_k(\mathbf{X_i})$ can be written as $(A_i - \pi_1)\left[\mathrm{D_i}(1)^{\mathrm{T}}\mathbf{V_i}(1)^{-1}\{E(\mathbf{Y_i}|A_i = 1, \mathbf{X_i}) - \mathbf{g}(1;\beta)\} - \mathbf{D_i}(0)^{\mathrm{T}}\mathbf{V_i}(0)^{-1}\{E(\mathbf{Y_i}|A_i = 0, \mathbf{X_i}) - \mathbf{g}(0;\beta)\}\right]$. The simple augmented GEE is thus

$$\sum_{i=1}^{m} \psi_{i_{\text{opt}}}(\mathbf{Y}, A, \mathbf{X};\beta) = \sum_{i=1}^{m} \mathbf{D_i}^{\mathrm{T}} \mathbf{V_i}^{-1} \{\mathbf{Y_i} - \mathbf{g}(A_i;\beta)\} - (A_i - \pi)\gamma(\mathbf{X_i}) = \mathbf{0}, \quad (4)$$

where $\gamma(\mathbf{X_i}) = \left[\mathbf{D_i}(\mathbf{1})^{\mathrm{T}}\mathbf{V_i}(1)^{-1}\{E(\mathbf{Y_i}|A_i = 1, \mathbf{X_i}) - \mathbf{g}(1;\beta)\} - \mathbf{D_i}(0)^{T}\mathbf{V_i}(0)^{-1}\{E(\mathbf{Y_i}|A_i = 0, \mathbf{X_i}) - \mathbf{g}(0;\beta)\}\right]$.

Solving for $\hat{\beta}_{\text{aug}}$ therefore requires knowledge of ϕ, α_k, π, and $E(\mathbf{Y_i}|\mathbf{X_i}, A_i = k)$ for $k = 0, 1$. Following standard practice, we estimate ϕ and α_k using the residuals from a generalized linear model (GLM) fit under independence. Specifically, ϕ is estimated by the Pearson chi-square statistic, and α_k is obtained by solving the treatment-specific moment equations $\sum_{i=1}^{m} I(A_i = k)\left\{\hat{\epsilon}_{ij}\hat{\epsilon}_{ij} - h(\alpha_k)\right\} = 0$, where $\hat{\epsilon}_{ij} = Y_{ij} - g(A_i;\hat{\beta}_{\text{init}})$, and $h(\alpha_k)$ is determined by the correlation structure assumed.

[‡]*We refer the reader to the online Supplementary material for closed-form solutions of parameter estimates in the standard and augmented logistic GEE for evaluation of cluster randomized trials with binary outcomes and for additional simulation results.*

Statistics in Medicine

For fixed $\psi_i(\mathbf{Y}, A; \beta)$, the optimality of the augmentation depends on correct estimation of $E(\mathbf{Y_i}|\mathbf{X_i}, A_i = k) = f_k(\mathbf{X_i}; \eta_k)$. When $E(\mathbf{Y_i}|\mathbf{X_i}, A_i = k)$ is misspecified, asymptotic normality and consistency hold, but the resulting estimator does not achieve maximum asymptotic efficiency. Several options are available for estimating the conditional mean $E(\mathbf{Y_i}|\mathbf{X_i}, A_i)$. We propose a strategy in the following discussion, which in large samples is guaranteed to improve on standard GEE. Following Tsiatis' approach of estimating $E(\mathbf{Y_i}|\mathbf{X_i}, A_i)$ separately within each arm, estimation proceeds via ordinary least squares (OLS) or maximum likelihood (ML) on an appropriately defined GLM. Although the observations within a cluster are not independent, the predicted values from OLS and ML fits remain consistent. For treatment-specific estimation, the argument in [22] may be generalized to GEE, guaranteeing that when $E(\mathbf{Y_i}|\mathbf{X_i}, A_i)$ is estimated with OLS, the augmented estimator is at least as efficient as the unaugmented estimator for continuous and discrete outcomes. This property holds even if models are misspecified. To more correctly specify the mean function, one may opt to fit an appropriate GLM, such as logistic regression for a binary outcome. We explore both approaches through simulation. It is also worthwhile to note that if the probability of treatment depends on baseline covariates $\mathbf{X_i}$ such that $\pi_k = P(A_i = k|\mathbf{X_i})$, the simple augmented GEE does provide a valid estimate of treatment effects, but OLS is no longer sufficient to guarantee efficiency improvement over unaugmented methods. For continuous Y_{ij} and identity link $g(A_i; \beta)$, the improved estimator is

$$\hat{\beta}_{\text{aug}} = \left[\sum_{i=1}^{m} \mathbf{A}_i^{*\mathrm{T}} \mathbf{V_i}^{-1} \mathbf{A}_i^* - (A_i - \pi)\left\{\mathbf{A}_i^*(1)^{\mathrm{T}} \mathbf{V}_i(1)^{-1} \mathbf{A}^*(1) - \mathbf{A}_i^*(0)^{\mathrm{T}} \mathbf{V}_i(0)^{-1} \mathbf{A}^*(0)\right\}\right]^{-1}$$
$$\times \left[\sum_{i=1}^{m} \mathbf{A}_i^{*\mathrm{T}} \mathbf{V_i}^{-1} \mathbf{Y_i} - (A_i - \pi)\Big\{\mathbf{A}_i^*(1)^{\mathrm{T}} \mathbf{V}_i(1)^{-1} \hat{E}(\mathbf{Y_i}|A_i = 1, \mathbf{X}_i) - \mathbf{A}_i^*(0)^{\mathrm{T}} \mathbf{V}_i(0)^{-1} \hat{E}(\mathbf{Y_i}|A_i = 0, X_i)\Big\}\right].$$

As in the unaugmented case, a closed-form solution can be derived for non-identity links under a simple design. Solutions for the logit link may be found in Appendix A of the Supplementary material.

We summarize the implementation of the simple augmented GEE for inclusion of baselines covariates in analysis of a cluster randomized trial in the following steps:

1. Determine $\hat{E}(\mathbf{Y_i}|\mathbf{X_i}, A_i = k) = f_k(\mathbf{X_i}; \hat{\eta}_k)$ from OLS or ML regression of Y onto baseline covariates X within each treatment arm.
2. Fit a GLM under independence to obtain $\hat{\beta}_{\text{init}}$.
3. Estimate ϕ and α_k from $\hat{\epsilon}_{ij}$ of the initial fit.
4. Construct the augmented estimating equations $\psi_{\text{aug}}(\mathbf{Y}, A, \mathbf{X}; \beta)$.
5. Solve for $\hat{\beta}_{\text{aug}}$.

The GEE was initially proposed as an iterative procedure, in which fitting involved repeatedly estimating correlation parameters α and mean parameters β until convergence. Since its inception, however, theoretical development and simulation studies have shown that the one-step procedure, as we have proposed for the augmented estimator, provides asymptotically equivalent estimates to the fully iterated approach, with similar finite sample properties [24].

3. Variance estimation

The asymptotic variance of $\hat{\beta}_{\text{aug}}$, under $m \to \infty$, is derived through the usual M-estimator Taylor expansion, accounting for the nuisance parameters $\hat{\eta}_k$ involved in estimating $E(\mathbf{Y_i}|\mathbf{X_i}, A_i = k) = f_k(\mathbf{X_i}; \eta_k)$. We let $\hat{\psi}_{i_{\text{opt}}}(\mathbf{Y}, A, \mathbf{X}; \beta)$ be an estimate of (4) evaluated at $\hat{\eta}$. The familiar sandwich variance estimator $\text{var}(\hat{\beta}_{\text{aug}}) = \Gamma^{-1}\Delta\Gamma^{-1^{\mathrm{T}}}$ is obtained, where $\Gamma = E\left[\frac{\partial \psi_{i_{\text{opt}}}(\mathbf{Y}, A, \mathbf{X}; \beta)}{\partial \beta^{\mathrm{T}}}\right]$, and $\Delta = E[\psi_{i_{\text{opt}}}(\mathbf{Y}, A, \mathbf{X}; \beta)^{\otimes 2}]$, where $U^{\otimes 2} = UU^{\mathrm{T}}$. By randomization, the augmentation term has mean zero and does not contribute to Γ. We therefore estimate Γ by $\hat{\Gamma} = m^{-1}\sum_i \mathbf{D_i}^{\mathrm{T}} \mathbf{V_i}^{-1} \mathbf{D_i}$. Estimation of η_k results in additional terms in our expansion of $\hat{\psi}_{i_{\text{opt}}}(\mathbf{Y}, A, \mathbf{X}; \beta)$:

$$\sum_{i=1}^{m}\hat{\psi}_{i_{\text{opt}}}(\mathbf{Y},A,\mathbf{X};\beta)=\sum_{i=1}^{m}\Big\{\mathbf{D_i}^{\mathrm{T}}\mathbf{V_i}^{-1}\{\mathbf{Y_i}-\mathbf{g}(A_i;\beta_0)\}-(A_i-\pi)\left[\mathbf{D_i}(1)^{\mathrm{T}}\mathbf{V_i}(1)^{-1}\{f_1(\mathbf{X_i};\hat{\eta}_1)\right.$$

$$\left.-\mathbf{g}(1;\beta_0)\}-\mathbf{D_i}(0)^{\mathrm{T}}\mathbf{V_i}(0)^{-1}\{f_0(\mathbf{X_i};\hat{\eta}_0)-\mathbf{g}(0;\beta_0)\}\right]\Big\}$$

$$=\sum_{i=1}^{m}\tilde{\psi}_{i_{\text{opt}}}=\sum_{i=1}^{m}\Bigg\{\mathbf{D_i}^{\mathrm{T}}\mathbf{V_i}^{-1}\{\mathbf{Y_i}-\mathbf{g}(A_i;\beta_0)\}-(A_i-\pi)\Bigg[\mathbf{D_i}(1)^{\mathrm{T}}\mathbf{V_i}(1)^{-1}$$

$$\times\{f_1(\mathbf{X_i};\eta_1^*)-\mathbf{g}(1;\beta_0)\}-\mathbf{D_i}(0)^{\mathrm{T}}\mathbf{V_i}(0)^{-1}\{f_0(\mathbf{X_i};\eta_0^*)-\mathbf{g}(0;\beta_0)\}\Bigg] \quad \text{(a)}$$

$$-(A_i-\pi)\Bigg[\mathbf{D_i}(1)^{\mathrm{T}}\mathbf{V_i}(1)^{-1}\{f_1'(\mathbf{X_i};\eta_1^*)\}(\hat{\eta}_1-\eta_1^*)$$

$$\left.\left.-\mathbf{D_i}(0)^{\mathrm{T}}\mathbf{V_i}(0)^{-1}\{f_0'(\mathbf{X_i};\eta_0^*)\}(\hat{\eta}_0-\eta_0^*)\right]\right\}+o_p(1)\Bigg\} \quad \text{(b)}$$

where $\hat{\eta}_k \xrightarrow{P} \eta_k^*$. By randomization (b) $\xrightarrow{P} 0$ as $m \to \infty$, showing that asymptotically there is no additional variability associated with estimating η_k, even when $f_k(\mathbf{X_i};\eta_k)$ is misspecified. For cluster randomized designs, however, the asymptotics may not hold, as the number of experimental units may be fairly small. In small sample settings, it is likely that $\text{var}(\hat{\beta}_{\text{aug}})$ is affected by estimation of η_k. We therefore estimate Δ by $\hat{\Delta}=m^{-1}\sum_i \tilde{\psi}_{i_{\text{opt}}}^{\otimes 2}$, with and without term (b), and evaluate our variance estimator through simulation. Specifically, we estimate $(\hat{\eta}_k-\eta_k^*)$ by its first-order approximation and substitute estimated parameter values for the truth. Inclusion of (b) is not guaranteed to increase the estimated variance but does provide a more unbiased estimate by accounting for estimation of η_k.

The sandwich variance estimator of standard GEE is known to often be biased downward for inference involving relatively few independent units. We examine Fay's bias-correction approach to recover loss. We choose this approach because, unlike other methods that were derived for standard GEE, Fay's method is generalizable to any M-estimator, including our augmented estimating equations. We apply Fay's first correction, in which Δ is estimated by $\hat{\Delta}^*=m^{-1}\sum_i(\mathbf{H_i}\hat{\psi}_i)^{\otimes 2}$, where $\mathbf{H_i}$ is a diagonal matrix with $H_{i_{jj}}=\left[1-\min\{q,(\frac{\partial\psi_i(\mathbf{Y},A,\mathbf{X};\beta)}{\partial\beta^{\mathrm{T}}}\times\hat{\Gamma})_{jj}\}\right]^{-1/2}$. Lower bound q is typically set to 0.75 to prevent gross inflation [5]. In total, we consider four standard error estimators for the simple augmented GEE: (1) unadjusted sandwich (SE_1); (2) nuisance ($\hat{\eta}_k$)-adjusted sandwich [term (b)] (SE_2); (3) sandwich with Fay's small-sample bias correction (SE_3); and (4) sandwich with Fay's small-sample bias correction and nuisance adjustment (SE_4), and evaluate each through simulation. We provide formulas for each estimator in Appendix B of the Supplementary material.

An alternative estimate $\hat{\text{var}}(\hat{\beta}_{\text{aug}})$ can be computed through a resampling technique such as the non-parametric bootstrap. To preserve the number of treated and control clusters within any bootstrap sample, we resample clusters within treatment arm. We use strategy 1 described by Davidson and Hinkley [25], in which the composition of resampled clusters is maintained and demonstrate this approach through data analysis.

4. Application: *Young Citizens Study*

We applied the simple augmented GEE to data from the *Young Citizens* study [26]. This trial involved a behavioral intervention designed to train children ages 10–14 years to educate their communities about HIV. To facilitate randomization, 30 communities were grouped into 15 pairs using a clustering algorithm involving several demographic characteristics. One community per pair was randomly assigned to treatment and the other to control. Residents within each community were surveyed post-intervention regarding their beliefs about the ability of children to effectively teach their peers and families about HIV. The primary outcome was a composite score reflecting the strength of this belief (Y_1). A secondary outcome measured residents' beliefs regarding whether or not the AIDS problem was getting worse in their communities. Residents responded on a 4-point scale with values 'strongly disagree', 'disagree', 'agree', and 'strongly agree'. Responses were dichotomized by collapsing 'strongly

agree' and 'agree' into one category labeled 'agree'; 'strongly disagree' and 'disagree' were collapsed similarly (Y_2). The number of residents surveyed per community ranged from 16–80 by multiples of 16.

We implemented standard and augmented GEE using the customary single correlation parameter as well as the less restrictive approach of allowing treatment-specific correlation. For augmented estimators, we also included estimation under independence to further examine the impact of covariance selection on the efficiency of augmented inference. We determined adjustment models separately for treatment and control groups by various model selection procedures. We selected the final models used in analysis via cross validation. For child efficacy (Y_1), the adjustment model in the treatment arm included the baseline covariates employment status, residential or urban community, the number of relatives living in the community, age, religion, population density, and whether or not the household had a flushing toilet, which was an indicator of household wealth. Among control communities, employment, age, and flushing toilet were included. The baseline covariates that entered the adjustment models for beliefs about the state of the HIV problem (Y_2) were mean community wealth, ethnic group, and household wealth for the intervention arm, and only mean community wealth for the control arm.

In analyzing our continuous outcome Y_1, we evaluated the marginal treatment effect by considering model (1), where $g(A_i;\beta)$ was the identity link function. We computed the standard error of $\hat{\beta}_1$ by the sandwich estimator and the nonparametric bootstrap for each estimation procedure. We applied the standard error modifications in Section 3, namely: (1) unadjusted sandwich (SE_1); (2) nuisance ($\hat{\eta}$)-adjusted sandwich [term (b)] (SE_2); (3) sandwich with small-sample bias correction (SE_3); and (4) sandwich with bias correction and nuisance adjustment (SE_4). In our second application, we evaluated the marginal treatment effect on the binary secondary outcome Y_2 and fit model (1) with the inverse logit link. We compared estimates obtained from standard GEE, the simple augmented GEE, adjusted logistic GEE with standardization, that is, the G-formula [27], and inverse probability of treatment weighted (IPTW) methods. In the IPTW approach, we ignore that the treatment probability is known and estimate $P(A=1|X)$ using a logistic regression model in which covariates were entered linearly.

In standard and augmented analyses, the intervention had a highly significant impact on the perceived ability of children to be peer educators {95%CI standard (0.182, 0.526), 95%CI augmented (0.245, 0.482)}. The adjusted sandwich variance estimator suggested over a 70% increase in efficiency resulting from covariate adjustment (RE, Table I). Bootstrap estimates showed a similar efficiency gain under common correlation (58%) and a much more modest gain using treatment-specific correlation (5%) ('RE boot', Table I). Comparing within unaugmented or augmented estimators, we observed little difference in standard error between estimators allowing for treatment-specific correlation versus estimators assuming common correlation. Estimates of β_1 were similar across standard and augmented estimators with either correlation structure.

With the examination of our binary outcome, Y_2, there was a marked difference in the estimated parameters when comparing standard and augmented GEE (Table II). The estimate $\hat{\beta}_1$ was -0.238 {95%CI $(-0.777, 0.300)$ } using standard methods, compared with values in the range $(-0.079, -0.023)$ for all augmented GEE estimates. In either approach, the effect of treatment on the perception of the

Table I. Marginal treatment effect analysis: parameter estimates, sandwich standard errors, and bootstrap standard error.

	Estimate	Sandwich standard error with adjustments						Bootstrap SE	
	$\hat{\beta}_1$	SE_1	SE_2	SE_3	SE_4	95%CI	RE	SE boot	RE boot
Std(Exch)	0.354	0.082	–	0.088	–	(0.182, 0.526)	1.000	0.081	1.000
Std(Exch-TS)	0.355	0.082	–	0.086	–	(0.186, 0.525)	1.034	0.085	0.910
Aug(Ind)	0.360	0.066	0.060	0.071	0.064	(0.236, 0.485)	1.528	0.062	1.681
Aug(Exch)	0.364	0.063	0.057	0.067	0.060	(0.245, 0.482)	1.730	0.064	1.582
Aug(Exch-TS)	0.360	0.063	0.061	0.066	0.064	(0.234, 0.485)	1.769	0.082	0.966

Std, unaugmented GEE; Aug, augmented GEE; Ind, independence; Exch, exchangeable correlation with single parameter; Exch-TS, exchangeable with treatment-specific correlation parameters; SE_1, unadjusted sandwich SE; SE_2, sandwich with nuisance parameter adjustment; SE_3, sandwich with small-sample adjustment; SE_4, sandwich with small-sample and nuisance adjustments; RE, square of the sandwich SE of the Std(Exch) estimator divided by the square of the sandwich SE of the indicated estimator; RE boot, square of the bootstrap SE of the Std(Exch) estimator divided by the square of the bootstrap SE of the indicated estimator. RE and confidence intervals (CIs) are based on SE_3 and SE_4 for unaugmented and augmented estimators, respectively.

Table II. Marginal treatment effect analysis with binary outcome: parameter estimates and sandwich standard errors.

Estimator	$\hat{\beta}_1$	SE	95%CI	RE
Standard GEE				
Std Exch	−0.238	0.275	(−0.777, 0.300)	1.000
Std Exch TS	−0.219	0.266	(−0.74, 0.301)	1.069
Augmented GEE				
Aug Ind-GLM	−0.062	0.215	(−0.484, 0.361)	1.627
Aug Exch-GLM	−0.079	0.21	(−0.491, 0.332)	1.716
Aug Exch-TS-GLM	−0.065	0.204	(−0.465, 0.335)	1.811
Aug Ind-OLS	−0.023	0.22	(−0.454, 0.408)	1.561
Aug Exch-OLS	−0.062	0.214	(−0.481, 0.358)	1.648
Aug Exch-TS-OLS	−0.047	0.206	(−0.451, 0.358)	1.773
Adjusted logistic GEE				
Model 1	−0.093	0.167	(−0.420, 0.234)	2.712
Model 2	−0.044	0.179	(−0.396, 0.308)	2.349
IPTW logistic GEE				
Model 3	−0.293	0.236	(−0.756, 0.170)	1.354

Std, unaugmented GEE; Aug, augmented GEE; Ind, independence; Exch, exchangeable correlation with single parameter; Exch-TS, exchangeable with treatment-specific correlation parameters; GLM or OLS, generalized linear model or ordinary least squares augmentation; RE, square of the sandwich SE of the Std(Exch) estimator divided by the square of the sandwich SE of the indicated estimator. Confidence intervals (CIs) and RE based on adjusted sandwich standard errors. Adjusted logistic GEE—model 1: $\text{logit}(P(Y_{ij} = 1)) = \eta_0 + \eta_1 \text{Mean\_wealth}_i$; model 2: $\text{logit}(P(Y_{ij} = 1)) = \eta_0 + \eta_1 \text{Mean\_wealth}_i + \eta_2 I(\text{Ethnic}_{ij} = 1) + \eta_3 I(\text{Wealth}_{ij} = 0)$. IPTW: $\text{logit}(P(A_i = 1)) = \eta_0 + \eta_1 \text{Know\_leader}_i + \eta_2 \text{Good\_floor}_i$.

AIDS epidemic was not significant at the $p = 0.05$ level {95%CI augmented GEE = (−0.491, 0.332)}. Estimates from the standardized adjusted logistic GEE were also closer to 0. Although effects were not significant at the $p = 0.05$ level for any of the approaches, confidence intervals for the augmented GEE were somewhat tighter, as were those using standard methods of covariate adjustment.

Considering efficiency, for both outcomes the estimated variability was lower for the augmented estimator compared to standard GEE. Although there is some uncertainty regarding the behavior of the sandwich estimator in small samples, these results suggest that when the asymptotics hold, augmented GEE is a valid approach that may be substantially more efficient than standard GEE. Randomized trials involving longitudinal data with many subjects or clustered designs with many small units, such as households or offspring, are therefore ideal candidates for this method. We evaluate our method and the behavior of the sandwich variance estimator through simulation in the following section.

5. Simulations

We assessed the performance of the simple augmented GEE in two sets of simulations. The first investigates continuous outcomes with an identity link, and the second set explores binary outcomes using an inverse logit link. We considered the impact of misspecification of the augmentation term and working covariance structure on the performance of our estimator. We base the results on 1000 simulated datasets.

5.1. Simulations 1

Cluster level covariates were treatment, density, wealth, and community type (e.g., urban/rural). We completed treatment assignment by first fixing the number of treated and control clusters to $m/2$, where m is the total number of clusters. We then randomly assigned clusters to treatment or control. We generated community type from a multinomial distribution. We generated density and wealth from the exponential and normal distributions, respectively. We simulated individual-level covariates

age, employment, security1, and security2 from normal and multinomial distributions with age treated as continuous and other covariates categorical. We generated data following the means and variances of covariates in the *Young Citizens* data. Intracluster correlation was induced by cluster-specific random effects and community-level covariates. We considered varying levels of correlation for treated versus control clusters. To assess small-sample performance, we compared scenarios of $m = 30$ and $m = 100$ clusters.

We generated outcomes from the following models:

$(Y_{ij}|X_{ij}, A_i = 1) = 7.23 + 0.599\text{employed}_{ij} + 0.44\text{mean\_wealth}_i - 0.22I(\text{security1}_{ij} = 3) - 0.06\text{age}_{ij} + 49.702\text{density}_i + b_{1i} + \epsilon_{ij}$, and $(Y_{ij}|X_{ij}, A_i = 0) = 2.56 + 0.245\text{employed}_{ij} + 0.691I(\text{community\_type}_i = 4) + 0.921I(\text{security2}_{ij} = 4) + 0.055\text{age}_{ij} + b_{0i} + \epsilon_{ij}$, where $b_{ki} \sim N(0, \sigma_k^2)$, and $\epsilon_{ij} \sim N(0, \sigma^2)$. Community-level and individual-level covariates therefore contributed to heterogeneity in subject responses. Values of σ_1^2 and σ_0^2 were selected to yield the desired within-cluster marginal correlation (ρ_k). For treatment and control clusters alike, $\sigma^2 = 1$.

We evaluated the effect of working covariance and augmentation misspecification by estimating β_1 and its variance under different covariance structures and augmentation models. We considered two

Table III. Standard versus augmented GEE, continuous outcome: 30 and 100 clusters, $\rho_0 = 0.05$, $\rho_1 = 0.10$, $\beta_1 = 1.3239$.

	$\hat{\beta}_1$	Bias	SE_1	SE_2	SE_3	SE_4	MC SE	MC RE	Cov. U	Cov. A
$m = 30$										
Std(Exch)	1.324	0.000	0.156	–	0.165	–	0.165	1.000	0.933	0.949
Std(Exch-TS)	1.325	−0.001	0.158	–	0.166	–	0.164	1.004	0.934	0.949
C(Ind)	1.324	−0.001	0.135	0.138	0.140	0.144	0.151	1.194	0.912	0.929
F(Ind)	1.325	−0.001	0.132	0.136	0.137	0.142	0.153	1.162	0.895	0.922
B(Ind)	1.325	−0.001	0.132	0.137	0.137	0.142	0.154	1.149	0.893	0.918
W(Ind)	1.321	0.003	0.143	0.151	0.150	0.158	0.160	1.057	0.911	0.937
C(Exch)	1.328	−0.004	0.130	0.132	0.134	0.137	0.147	1.248	0.909	0.927
F(Exch)	1.328	−0.004	0.128	0.131	0.132	0.135	0.149	1.219	0.898	0.917
B(Exch)	1.328	−0.004	0.128	0.131	0.132	0.135	0.150	1.205	0.893	0.915
W(Exch)	1.326	−0.002	0.139	0.145	0.144	0.150	0.156	1.110	0.904	0.928
C(Exch-TS)	1.329	−0.005	0.134	0.138	0.137	0.143	0.149	1.223	0.914	0.925
F(Exch-TS)	1.329	−0.005	0.133	0.137	0.135	0.142	0.151	1.198	0.895	0.915
B(Exch-TS)	1.329	−0.005	0.132	0.137	0.135	0.142	0.151	1.186	0.890	0.910
W(Exch-TS)	1.328	−0.004	0.142	0.148	0.146	0.154	0.159	1.078	0.900	0.927
$m = 100$										
Std(Exch)	1.3261	−0.002	0.088	–	0.090	–	0.090	1.000	0.943	0.946
Std(Exch-TS)	1.3265	−0.003	0.088	–	0.089	–	0.090	1.001	0.942	0.945
C(Ind)	1.3276	−0.004	0.077	0.077	0.078	0.078	0.078	1.333	0.942	0.947
F(Ind)	1.3280	−0.004	0.076	0.077	0.077	0.078	0.079	1.296	0.937	0.943
B(Ind)	1.3279	−0.004	0.076	0.077	0.077	0.078	0.079	1.297	0.937	0.944
W(Ind)	1.3266	−0.003	0.083	0.084	0.084	0.085	0.088	1.048	0.935	0.945
C(Exch)	1.3281	−0.004	0.074	0.074	0.075	0.075	0.075	1.449	0.944	0.945
F(Exch)	1.3286	−0.005	0.073	0.074	0.074	0.075	0.076	1.409	0.933	0.941
B(Exch)	1.3286	−0.005	0.073	0.074	0.074	0.075	0.076	1.410	0.932	0.94
W(Exch)	1.3269	−0.003	0.080	0.081	0.080	0.082	0.083	1.166	0.936	0.943
C(Exch-TS)	1.3282	−0.004	0.074	0.074	0.074	0.075	0.074	1.452	0.946	0.95
F(Exch-TS)	1.3288	−0.005	0.073	0.074	0.074	0.075	0.075	1.412	0.937	0.943
B(Exch-TS)	1.3287	−0.005	0.073	0.074	0.074	0.075	0.075	1.413	0.936	0.943
W(Exch-TS)	1.3273	−0.003	0.079	0.080	0.080	0.081	0.083	1.168	0.934	0.943

Std, unaugmented; C, F, B, or W, augmented with 'correct', 'forward' selected, 'backward' selected, or 'wrong' model; Ind, independence; Exch, exchangeable with single correlation parameter; Exch-TS, exchangeable with treatment-specific parameter; SE_1, average unadjusted sandwich SE; SE_2, average sandwich SE with nuisance parameter adjustment; SE_3, average sandwich SE with small-sample adjustment; SE_4, average sandwich SE with small-sample and nuisance adjustments; MC SE, Monte Carlo standard deviation; MC RE, square of the Monte Carlo SE of the Std(Exch) estimator divided by the Monte Carlo SE for the indicated estimator; Cov. U, SE_1 coverage; Cov. A, SE_3 and SE_4 coverage for unaugmented and augmented GEE, respectively.

variations of standard GEE: standard GEE with common exchangeable correlation {Std(Exch)} and standard GEE with treatment-specific exchangeable correlation {Std(Exch-TS)}. For the class of augmented GEE, we estimated $\hat{\beta}_{\text{aug}}$ with independence, exchangeable, and treatment-specific exchangeable correlation structures. We evaluated each estimator under several augmentation models. The estimator resulting from fitting the true form of $E(Y_{ij}|X_{ij}, A_i = k)$ is denoted by 'C' for 'correct'. Alternative augmentation models were defined by forward (F) and backward (B) selections and a wrong (W) model. The wrong models were given by $E(Y_{ij}|X_{ij}, A_i = 1) = \eta_0 + \eta_1 \text{mean\_wealth}_i + \eta_2 I(\text{community\_type}_i = 2) + \eta_3 I(\text{security1}_{ij} = 2)$ and $E(Y_{ij}|X_{ij}, A_i = 0) = \eta_0 + \eta_1 \text{density}_i + \eta_2 \text{age}_{ij} + \eta_3 I(\text{community\_type}_i = 1) + \eta_4 I(\text{security2}_{ij} = 4)$. Augmentation under the 'correct' model illustrates the largest possible efficiency gain. Alternative model fitting techniques were chosen to be representative of methods commonly used when performing covariate adjustment in analyzing clinical trial data. Analysts may use forward or backward stepwise selection favoring more parsimonious or larger models, respectively. We included the 'wrong' model for comparison using models that contain some relevant covariates but omit others. To correct for small-sample variance underestimation, we applied several modifications to the sandwich estimator as detailed in Section 3. For the unaugmented estimators, we calculated the sandwich variance

Table IV. Standard versus augmented GEE, continuous outcome: 30 and 100 clusters, $\rho_0 = 0.13$, $\rho_1 = 0.17$, $\beta_1 = 1.3239$.

	$\hat{\beta}_1$	Bias	SE_1	SE_2	SE_3	SE_4	MC SE	MC RE	Cov. U	Cov. A
$m = 30$										
Std(Exch)	1.312	0.012	0.217	–	0.229	–	0.233	1.000	0.933	0.937
Std(Exch-TS)	1.313	0.011	0.217	–	0.227	–	0.233	1.000	0.929	0.934
C(Ind)	1.317	0.007	0.209	0.219	0.221	0.232	0.234	0.989	0.907	0.934
F(Ind)	1.317	0.007	0.201	0.216	0.211	0.226	0.255	0.834	0.863	0.908
B(Ind)	1.317	0.007	0.202	0.216	0.211	0.227	0.256	0.828	0.861	0.911
W(Ind)	1.310	0.014	0.213	0.226	0.226	0.240	0.250	0.864	0.897	0.922
C(Exch)	1.319	0.005	0.199	0.206	0.207	0.214	0.215	1.170	0.912	0.934
F(Exch)	1.319	0.005	0.194	0.205	0.201	0.213	0.237	0.963	0.881	0.916
B(Exch)	1.319	0.005	0.194	0.206	0.201	0.213	0.238	0.954	0.879	0.915
W(Exch)	1.314	0.010	0.204	0.214	0.213	0.224	0.232	1.008	0.910	0.94
C(Exch-TS)	1.320	0.004	0.198	0.205	0.206	0.213	0.215	1.172	0.911	0.935
F(Exch-TS)	1.319	0.004	0.193	0.205	0.200	0.212	0.237	0.964	0.884	0.914
B(Exch-TS)	1.320	0.004	0.193	0.205	0.200	0.212	0.238	0.955	0.882	0.914
W(Exch-TS)	1.315	0.009	0.203	0.214	0.212	0.223	0.232	1.008	0.909	0.938
$m = 100$										
Std(Exch)	1.318	0.006	0.123	–	0.125	–	0.124	1.000	0.947	0.948
Std(Exch-TS)	1.318	0.006	0.123	–	0.125	–	0.123	1.004	0.947	0.949
C(Ind)	1.322	0.002	0.121	0.123	0.123	0.125	0.124	0.997	0.943	0.949
F(Ind)	1.321	0.003	0.118	0.121	0.120	0.123	0.125	0.979	0.935	0.941
B(Ind)	1.321	0.003	0.118	0.121	0.120	0.123	0.125	0.978	0.935	0.941
W(Ind)	1.320	0.004	0.124	0.127	0.127	0.129	0.128	0.931	0.935	0.944
C(Exch)	1.319	0.005	0.113	0.114	0.114	0.116	0.116	1.128	0.939	0.943
F(Exch)	1.318	0.006	0.112	0.114	0.113	0.115	0.118	1.102	0.929	0.937
B(Exch)	1.318	0.006	0.112	0.114	0.113	0.115	0.118	1.101	0.929	0.937
W(Exch)	1.317	0.007	0.117	0.118	0.118	0.120	0.120	1.056	0.934	0.944
C(Exch-TS)	1.319	0.005	0.113	0.114	0.114	0.115	0.116	1.133	0.941	0.945
F(Exch-TS)	1.318	0.006	0.111	0.114	0.113	0.115	0.118	1.106	0.931	0.939
B(Exch-TS)	1.318	0.006	0.111	0.114	0.113	0.115	0.118	1.105	0.931	0.939
W(Exch-TS)	1.317	0.007	0.116	0.118	0.118	0.120	0.120	1.061	0.935	0.942

Std, unaugmented; C, F, B, or W, augmented with 'correct', 'forward' selected, 'backward' selected, or 'wrong' model; Ind, independence; Exch, exchangeable with single correlation parameter; Exch-TS, exchangeable with treatment-specific parameter; SE_1, average unadjusted sandwich SE; SE_2, average sandwich SE with nuisance parameter adjustment; SE_3, average sandwich SE with small-sample adjustment; SE_4, average sandwich SE with small-sample and nuisance adjustments; MC SE, Monte Carlo standard deviation; MC RE, square of the Monte Carlo SE of the Std(Exch) estimator divided by the Monte Carlo SE for the indicated estimator; Cov. U, SE_1 coverage; Cov. A, SE_3 and SE_4 coverage for unaugmented and augmented GEE, respectively.

(SE_1) and the sandwich variance with bias correction (SE_3). We calculated standard errors for augmented estimators using SE_2 and SE_4 as well, which account for η_k-estimation.

Table III shows results for $m = 30$ and 100, $\sigma_1^2 = 0.03$, and $\sigma_0^2 = 0.025$, which correspond to approximately 10% and 5% within-cluster correlation in treated and control clusters, respectively. For Table IV, we raise the level of unexplained similarity among cluster members by setting $\sigma_1^2 = 0.23$ and $\sigma_0^2 = 0.20$ for the sample sizes previously considered.

For small-sample and large-sample inferences, bias was similar across all estimators. Working covariance specification affected the variance of the augmented estimator, with exchangeable (true) correlation structures resulting in smaller average standard errors than independence. With the comparison of estimators calculated with an exchangeable correlation structure, augmented estimators were often more efficient than the standard approach. Monte Carlo relative efficiency estimates suggest that in the small-sample setting with low levels of unexplained intracommunity correlation, considerable improvement (5–19%) is observed even when misspecifying the augmentation model (Table III). When intracluster correlation was larger, additional variability associated with automated model selection resulted in loss of efficiency associated with augmentation (Table IV). Average sandwich standard errors were overly optimistic in comparing augmented GEE with standard GEE in small samples, consistently estimating lower variability with augmentation. For large samples, efficiency gains were not hindered by higher levels of unexplained cluster similarity, with Monte Carlo efficiency improving by 5–40% (Table IV).

Coverage results show that for small samples, the uncorrected sandwich variance underestimates the variability of the augmented estimator (Tables III and IV). Bias correction fully recovered small-sample loss of variance for standard GEE. For augmented estimators, correction was less effective. Coverage was slightly increased by accounting for augmentation in the sandwich variance but did not quite reach nominal levels. For large-sample inference, neither adjustment substantially increased coverage, which was already close to nominal levels for the uncorrected sandwich variance without the nuisance term.

5.2. *Simulations 2*

To explore the performance of the augmented GEE for clustered binary outcomes, we again generated datasets of m clusters with probability of treatment $P(A = 1) = 1/2$. We simulated cluster-level variables X_1 and X_2 from exponential and multinomial distributions with mean 0.002 and probabilities $p = (0.46, 0.27, 0.07, 0.17, 0.03)$, respectively. We generated individual-level covariates X_3, X_4, X_5, and X_6 such that $(X_3, X_4) \sim \text{Normal}\left(\begin{pmatrix} 0 & 0 \end{pmatrix}, \begin{pmatrix} 4 & 6 \\ 6 & 25 \end{pmatrix}\right)$, $X_5 \sim \text{Bernoulli}(p = 0.28)$, and $X_6 \sim \text{Multinomial}\{1, p = (0.45, 0.15, 0.30, 0.10)\}$. We used the random intercept logistic model to simulate correlated binary outcomes $\mathbf{Y}$. We drew random intercepts b_i from the bridge distribution for the logit link [28], $B_l(0, 1-\rho)$, where 0 is the mean and ρ is the desired correlation. We selected the bridge distribution to preserve the logistic shape after marginalizing over random effects and provide a simple scaling relationship between parameters of the models for $E(\mathbf{Y}|\mathbf{X}, A, b)$ and $E(\mathbf{Y}|\mathbf{X}, A)$. Outcome-generating models were as follows: $\text{logit}\{E(Y_{ij}|X_{ij}, A_i = 1, b)\} = \eta_{1_0} + \eta_{1_1} X_{3_{ij}} + \eta_{1_2} X_{4_{ij}} + b_i$ and $\text{logit}\{E(Y_{ij}|X_{ij}, A_i = 0, b)\} = \eta_{0_0} + \eta_{0_1} X_{4_{ij}} + \eta_{0_2} X_{4_{ij}}^2 + \eta_{0_3} X_{5_{ij}} + b_i$.

For low association between Y and X, we set $\eta_0 = (3.4, -0.6, 0.03, 0.5)^{\mathrm{T}}$ and $\eta_1 = (2.5, -0.62, 0.86)^{\mathrm{T}}$. We used coefficients $\eta_0 = c(2.0, -0.9, 0.03, 0.5)^{\mathrm{T}}$ and $\eta_1 = (1.5, -0.62, 0.86)^{\mathrm{T}}$ for a high association. We again compared small-sample versus large-sample performance by implementing standard and augmented GEE under $m = 30$, $m = 100$, and $m = 250$ clusters. We also considered two levels of intracluster correlation ($\rho = 0.05, 0.20$). We included results for $m = 250$ in Appendix C of the Supplementary material.

We applied the augmented GEE under independent and exchangeable correlation structures and evaluated different methods of fitting the augmentation term. To guarantee improved efficiency relative to standard GEE, we fit augmentation models $E(\mathbf{Y}|\mathbf{X}, A = k)$ using OLS. We contrast this approach with logistic regression, which correctly specifies the form of the relationship between Y and X but is not guaranteed to improve efficiency under model misspecification. For each model fitting technique, we fit the correct augmentation model (C), a forward selection model (F), and two wrong models (O and W). Wrong models denoted by 'O' contained one baseline covariate. Specifically, the models fit were $E(Y_{ij}|X_{ij}, A = 1) = g(\alpha_{1_0} + \alpha_{1_1} X_{4_{ij}})$ and $E(Y_{ij}|X_{ij}, A = 0) = g(\alpha_{0_0} + \alpha_{0_1} X_{3_{ij}})$. Wrong models 'W' are given by $E(Y_{ij}|X_{ij}, A = 1)\} = g(\alpha_{1_0} + \alpha_{1_1} X_{5_{ij}} + \alpha_{1_1} X_{2_i})$ and $E(Y_{ij}|X_{ij}, A = 0) = g(\alpha_{0_0} + \alpha_{0_1} X_{4_{ij}} + \alpha_{0_2} X_{1_{ij}} + \alpha_{0_3} X_{5_{ij}})$.

Table V. Standard versus augmented GEE, binary outcome: 30 and 100 clusters, low and high association, $\rho = 0.05$, $\beta_1 = -0.2959$ low association, $\beta_1 = 1.1362$ high association.

	Estimator	$\hat{\beta}_1$	Bias	SE_1	SE_2	SE_3	SE_4	MC SE	MC RE	Cov. U	Cov. A
Low											
$m = 30$	Std	−0.299	0.003	0.196	–	0.209	–	0.220	1.000	0.923	0.945
	C-ML(Ind)	−0.300	0.004	0.193	0.196	0.207	0.210	0.220	1.002	0.924	0.942
	C-ML	−0.299	0.003	0.187	0.189	0.198	0.200	0.210	1.101	0.920	0.937
	C-OLS	−0.300	0.004	0.188	0.191	0.200	0.202	0.213	1.075	0.919	0.936
	F-ML	−0.298	0.003	0.178	0.180	0.187	0.189	0.225	0.956	0.880	0.901
	F-OLS	−0.302	0.006	0.179	0.183	0.188	0.192	0.226	0.948	0.878	0.906
	O-ML	−0.299	0.003	0.190	0.192	0.202	0.204	0.215	1.051	0.918	0.935
	O-OLS	−0.299	0.003	0.190	0.192	0.202	0.205	0.215	1.054	0.914	0.937
	W-ML	−0.295	−0.001	0.191	0.195	0.202	0.206	0.224	0.971	0.904	0.929
	W-OLS	−0.298	0.002	0.191	0.195	0.202	0.207	0.224	0.970	0.902	0.933
$m = 100$	Std	−0.293	−0.003	0.115	–	0.117	–	0.116	1.000	0.944	0.947
	C-ML(Ind)	−0.292	−0.004	0.113	0.113	0.115	0.116	0.117	0.987	0.941	0.946
	C-ML	−0.293	−0.003	0.109	0.109	0.111	0.111	0.112	1.089	0.938	0.940
	C-OLS	−0.293	−0.003	0.110	0.110	0.112	0.112	0.112	1.077	0.943	0.945
	F-ML	−0.293	−0.003	0.107	0.108	0.109	0.110	0.113	1.057	0.934	0.944
	F-OLS	−0.293	−0.003	0.108	0.109	0.109	0.110	0.114	1.052	0.937	0.943
	O-ML	−0.293	−0.003	0.111	0.111	0.113	0.113	0.113	1.070	0.944	0.950
	O-OLS	−0.293	−0.003	0.111	0.111	0.113	0.113	0.113	1.070	0.943	0.948
	W-ML	−0.293	−0.003	0.113	0.114	0.115	0.115	0.116	1.014	0.935	0.945
	W-OLS	−0.293	−0.003	0.113	0.114	0.115	0.116	0.116	1.015	0.937	0.943
High											
$m = 30$	Std	1.137	−0.001	0.153	–	0.163	–	0.169	1.000	0.925	0.946
	C-ML(Ind)	1.132	0.004	0.136	0.138	0.146	0.149	0.151	1.254	0.935	0.948
	C-ML	1.135	0.001	0.132	0.134	0.140	0.142	0.144	1.382	0.932	0.952
	C-OLS	1.134	0.002	0.134	0.136	0.143	0.145	0.146	1.348	0.938	0.951
	F-ML	1.135	0.002	0.125	0.128	0.132	0.135	0.155	1.195	0.901	0.921
	F-OLS	1.137	−0.001	0.127	0.131	0.134	0.139	0.156	1.172	0.896	0.924
	O-ML	1.135	0.001	0.135	0.137	0.144	0.146	0.148	1.306	0.928	0.955
	O-OLS	1.135	0.001	0.136	0.139	0.146	0.148	0.150	1.278	0.930	0.950
	W-ML	1.137	−0.001	0.141	0.145	0.150	0.154	0.161	1.101	0.924	0.947
	W-OLS	1.137	−0.001	0.141	0.146	0.150	0.155	0.163	1.085	0.921	0.945
$m = 100$	Std	1.138	−0.002	0.089	–	0.090	–	0.090	1.000	0.946	0.949
	C-ML(Ind)	1.139	−0.003	0.079	0.079	0.080	0.081	0.083	1.162	0.934	0.935
	C-ML	1.139	−0.003	0.076	0.076	0.078	0.078	0.080	1.257	0.936	0.941
	C-OLS	1.140	−0.004	0.077	0.078	0.079	0.079	0.081	1.234	0.943	0.945
	F-ML	1.138	−0.002	0.075	0.075	0.076	0.076	0.082	1.210	0.931	0.935
	F-OLS	1.140	−0.003	0.076	0.076	0.077	0.078	0.082	1.197	0.934	0.943
	O-ML	1.139	−0.003	0.078	0.078	0.080	0.080	0.081	1.222	0.943	0.954
	O-OLS	1.140	−0.004	0.079	0.079	0.080	0.081	0.082	1.203	0.946	0.952
	W-ML	1.139	−0.003	0.083	0.083	0.084	0.085	0.086	1.098	0.948	0.954
	W-OLS	1.140	−0.003	0.083	0.084	0.084	0.085	0.086	1.095	0.945	0.952

Std, unaugmented; Correlation exchangeable unless denoted by 'Ind' for independence. C, F, O, or W, augmentation with 'correct', 'forward' selected, 'one-variable', or 'wrong' model; ML or OLS, augmentation fit with maximum likelihood or ordinary least squares; SE_1, average unadjusted sandwich SE; SE_2, average sandwich SE with nuisance parameter adjustment; SE_3, average sandwich SE with small-sample adjustment; SE_4, average sandwich SE with small-sample and nuisance adjustments; MC RE, square of the Monte Carlo SE of the Std(Exch) estimator divided by the Monte Carlo SE for the indicated estimator; Cov. U, SE_1 coverage; Cov. A, SE_3 and SE_4 coverage for unaugmented and augmented GEE, respectively.

Results were similar to those obtained for the continuous outcomes in the first set of simulations (Tables V and VI). Bias was similar across all methods of estimation for small-sample and large-sample inference, and correct specification of the working covariance resulted in more efficient estimation for augmented estimators. Small-sample results suggested that for low association of baseline covariates

Statistics in Medicine

Table VI. Standard versus augmented GEE, binary outcome: 30 and 100 clusters, low and high association, $\rho = 0.20$ $\beta_1 = -0.2164$ low association, $\beta_1 = 1.0501$ high association.

	Estimator	$\hat{\beta}_1$	Bias	SE_1	SE_2	SE_3	SE_4	MC SE	MC RE	Cov. U	Cov.A
Low											
$m = 30$	Std	−0.228	0.012	0.317	–	0.335	–	0.344	1.000	0.925	0.935
	C-ML(Ind)	−0.235	0.019	0.328	0.332	0.354	0.358	0.373	0.855	0.908	0.934
	C-ML	−0.229	0.013	0.312	0.313	0.329	0.330	0.338	1.041	0.927	0.939
	C-OLS	−0.230	0.013	0.313	0.314	0.330	0.331	0.339	1.031	0.931	0.937
	F-ML	−0.218	0.001	0.294	0.296	0.308	0.310	0.365	0.891	0.892	0.908
	F-OLS	−0.224	0.008	0.297	0.303	0.310	0.317	0.376	0.838	0.887	0.905
	O-ML	−0.232	0.015	0.314	0.315	0.331	0.332	0.338	1.036	0.924	0.935
	O-OLS	−0.232	0.016	0.314	0.315	0.331	0.332	0.339	1.034	0.926	0.935
	W-ML	−0.221	0.004	0.311	0.314	0.328	0.330	0.350	0.968	0.915	0.925
	W-OLS	−0.225	0.009	0.312	0.315	0.329	0.332	0.352	0.958	0.912	0.926
$m = 100$	Std	−0.213	−0.003	0.183	–	0.186	–	0.188	1.000	0.937	0.943
	C-ML(Ind)	−0.213	−0.003	0.192	0.193	0.196	0.197	0.201	0.868	0.938	0.943
	C-ML	−0.212	−0.004	0.180	0.180	0.182	0.182	0.185	1.031	0.939	0.941
	C-OLS	−0.212	−0.004	0.180	0.180	0.183	0.183	0.185	1.023	0.930	0.935
	F-ML	−0.213	−0.003	0.176	0.177	0.179	0.179	0.190	0.975	0.925	0.930
	F-OLS	−0.213	−0.003	0.176	0.178	0.179	0.180	0.191	0.968	0.930	0.933
	O-ML	−0.213	−0.003	0.181	0.181	0.184	0.184	0.186	1.021	0.930	0.936
	O-OLS	−0.213	−0.004	0.181	0.181	0.184	0.184	0.186	1.018	0.930	0.936
	W-ML	−0.213	−0.004	0.181	0.182	0.184	0.185	0.188	0.996	0.937	0.943
	W-OLS	−0.213	−0.003	0.181	0.182	0.184	0.185	0.188	0.995	0.935	0.943
High											
$m = 30$	Std	1.067	−0.017	0.244	–	0.258	–	0.251	1.000	0.934	0.948
	C-ML(Ind)	1.065	−0.015	0.244	0.247	0.264	0.267	0.258	0.941	0.923	0.949
	C-ML	1.069	−0.019	0.230	0.231	0.243	0.244	0.235	1.132	0.935	0.950
	C-OLS	1.069	−0.019	0.231	0.232	0.245	0.246	0.237	1.120	0.934	0.947
	F-ML	1.065	−0.015	0.217	0.219	0.228	0.230	0.257	0.952	0.889	0.909
	F-OLS	1.069	−0.019	0.219	0.224	0.230	0.235	0.262	0.914	0.895	0.913
	O-ML	1.074	−0.024	0.233	0.234	0.247	0.247	0.243	1.061	0.931	0.941
	O-OLS	1.075	−0.025	0.234	0.235	0.247	0.248	0.244	1.052	0.928	0.943
	W-ML	1.066	−0.016	0.234	0.237	0.248	0.250	0.248	1.021	0.931	0.942
	W-OLS	1.067	−0.017	0.235	0.237	0.248	0.250	0.248	1.018	0.932	0.944
$m = 100$	Std	1.062	−0.012	0.138	–	0.140	–	0.137	1.000	0.946	0.949
	C-ML(Ind)	1.059	−0.009	0.140	0.140	0.143	0.144	0.143	0.928	0.941	0.950
	C-ML	1.059	−0.009	0.130	0.130	0.133	0.133	0.129	1.124	0.952	0.958
	C-OLS	1.059	−0.009	0.131	0.131	0.133	0.133	0.130	1.117	0.953	0.956
	F-ML	1.057	−0.007	0.128	0.129	0.130	0.131	0.132	1.076	0.944	0.948
	F-OLS	1.058	−0.007	0.128	0.130	0.130	0.132	0.133	1.070	0.946	0.953
	O-ML	1.060	−0.010	0.132	0.132	0.134	0.134	0.131	1.096	0.947	0.952
	O-OLS	1.060	−0.010	0.132	0.132	0.134	0.134	0.131	1.095	0.950	0.953
	W-ML	1.061	−0.011	0.134	0.135	0.136	0.137	0.135	1.041	0.943	0.949
	W-OLS	1.062	−0.011	0.134	0.135	0.136	0.137	0.134	1.044	0.945	0.951

Std, unaugmented; Correlation exchangeable unless denoted by 'Ind' for independence. C, F, O, or W, augmentation with 'correct', 'forward' selected, 'one-variable', or 'wrong' model; ML or OLS, augmentation fit with maximum likelihood or ordinary least squares; SE_1, average unadjusted sandwich SE; SE_2, average sandwich SE with nuisance parameter adjustment; SE_3, average sandwich SE with small-sample adjustment; SE_4, average sandwich SE with small-sample and nuisance adjustments; MC RE, square of the Monte Carlo SE of the Std(Exch) estimator divided by the Monte Carlo SE for the indicated estimator; Cov. U, SE_1 coverage; Cov. A, SE_3 and SE_4 coverage for unaugmented and augmented GEE, respectively.

and outcome, small gains are possible for reasonably specified models (2–10%), but for automated model selection and poorly specified models, efficiency loss occurs (−17% to −3%) because of additional variability introduced by model selection and estimation of the augmentation terms. Efficiency

increased by 8–35% when baseline covariates were more strongly related to the outcome, and unexplained intracluster correlation was low. For higher intracluster correlation, efficiency gains were lower (−5% to 12%, Table VI), with loss of efficiency for automated model selection. Similar to the continuous outcome, standard error adjustments were partially effective in recovering nominal coverage. When 100 clusters were sampled, augmentation increased efficiency by 1–35% for high association or low intracluster correlation. With low association between $\mathbf{X}$ and $\mathbf{Y}$ and high intracluster correlation, augmentation decreased efficiency for poorly specified and automated models. Considering 250 clusters, augmented estimators were more efficient than unaugmented estimators across the levels of intracluster correlation, degree of $\mathbf{X}$ and $\mathbf{Y}$ association, and methods of model fitting that were considered (Supplementary material).

In summary, large-sample results suggest improvement with augmentation, whereas results for small-sample estimation are less consistent. Across the number of clusters evaluated, augmentation was less beneficial as the degree of intracluster correlation increased. Regarding augmentation fit, the variability of $\hat{\beta}_1$ was similar when comparing augmented estimators resulting from predictions from OLS and ML.

6. Discussion

This paper demonstrates the use of methodology based on semiparametric theory to improve efficiency of inferences in randomized studies with correlated outcomes through augmenting the standard GEE. This method extends the work of Zhang *et al.* [17] by focusing on multivariate outcomes and is the first application of this approach to a cluster randomized trial.

The binary outcome analysis illustrates an additional advantage of augmented GEE—double robustness. Results from standard GEE may result in misleading estimates in settings where randomization has led to an imbalance in important predictors. Augmentation involves specifying a conditional model $E(\mathbf{Y}|\mathbf{X}, A)$ that corrects for imbalances and therefore recovers unbiased estimates of treatment effects, even when randomization does not result in independence of $\mathbf{X}$ and A in the observed data. Alternative methods for correction, such as IPTW using a predictive model for the probability of treatment given baseline covariates, may not perform well given the cluster-level assignment. Predictive models for treatment only make use of cluster-level information; individual-level covariates may be averaged by cluster to create cluster-level covariates, but this data-coarsened approach can lead to poorly specified models. Generally, inference on the probability of treatment will be poor given the small number of randomized units. Alternatively, augmentation exploits relationships among individual-level covariates and outcomes. Because there are multiple individuals per cluster, there is more information available for estimating $E(\mathbf{Y}|\mathbf{X}, A)$ compared with $P(A = 1|\mathbf{X})$. Estimation of $E(\mathbf{Y}|\mathbf{X}, A)$ may consequently result in a better estimator of β_1.

Simulation studies explored the possibility of efficiency gains using the augmented GEE in small-sample and large-sample settings. For large samples, the augmented GEE improved efficiency compared with the standard GEE for marginal treatment effects, which ignores baseline covariates. In the small-sample setting, efficiency gain was less consistent; low levels of between-community heterogeneity and high degrees of association between baseline covariates and outcomes were required to benefit from augmentation. Gail *et al.* [14] found a similar trend in their studies of permutation inference, noting that covariate adjustment did not improve efficiency when between-community variability was high. These results highlight the importance of measuring all covariates that contribute to within-community similarities in response. Interpreting the results from the *Young Citizens* study using the insight obtained through simulations, the low intracluster correlation (0.02) suggests improvement in efficiency when adjusting for baseline covariates. The degree of improvement, however, may be overstated by sandwich standard errors. Small-sample estimation also resulted in coverage slightly below nominal levels, even after standard error adjustment. The standard error modifications used only consider first-order approximations to the sandwich variance and nuisance parameter distributions. The simulation results suggest that higher-order effects of nuisance parameter estimation may impact variance underestimation. The shortcomings of this approach in small samples motivate investigation into the use of augmented estimators with permutation-based inference.

We implemented augmentation using separate models for treatment and control, with ML and OLS for binary outcomes and OLS for continuous outcomes. Asymptotically, treatment-specific OLS including an intercept term is guaranteed to be at least as efficient as the unadjusted estimator [12, 22]. As Zhang and Gilbert [23] discussed, data splitting can be inefficient in finite samples compared with fitting a

common model for $E(\mathbf{Y}|\mathbf{X}, A)$. For studies involving relatively few randomized units, fitting a common conditional model may better utilize covariate information. The effect of data splitting in finite sample inference has not yet been examined in practice. To guarantee efficiency gain over unadjusted methods when fitting a common model, we may use van der Laan's empirical efficiency maximization approach [29]. This method estimates nuisance parameters by empirically minimizing the asymptotic variance of a scalar targeted parameter. It results in fitting adjustment models with a weighted least squares procedure, in which weights depend on treatment probabilities.

Although the simple augmented GEE improves estimation in large samples, it is not the semiparametric efficient estimator for our restricted mean model for multivariate outcome data, even under correct specification of $E(\mathbf{Y}|A, \mathbf{X})$. Nonetheless, the simple augmented GEE builds upon standard GEE in an intuitive way and provides insight into how augmentation may be used with multivariate data to improve efficiency. Development of a locally semiparametric efficient estimator for restricted mean models for multivariate data and an understanding of its behavior remain important research questions. A locally efficient estimator is an estimator that remains consistent and asymptotically normal under the restricted mean model and that achieves the semiparametric efficiency bound for the model at the submodel where nuisance parameters are correctly specified. When model misspecification of nuisance parameters is present, it is not clear whether the locally efficient estimator will still improve efficiency compared with standard techniques. Additional modification of the locally efficient estimator is needed to ensure improvement relative to standard GEE. Further research is warranted in this area.

Acknowledgements

The authors are grateful to Felton Earls and Mary Carlson for providing the *Young Citizens* data and to Mark van der Laan for the discussion. The National Institutes of Health (grant nos. AIR0151164 and T32AI007358) supported this research.

References

1. Klar N, Donner A. *Design and Analysis of Cluster Randomization Trials in Health Research*. Hodder Arnold: London, 2000.
2. Laird NM, Ware JH. Random-effects models for longitudinal data. *Biometrics* 1982; **38**:963–974. DOI: 10.2307/2529876.
3. Liang KY, Zeger SL. Longitudinal data analysis for discrete and continuous outcomes. *Biometrics* 1986; **42**:121–130. DOI: 10.2307/2531248.
4. Gunsolley JC, Getchell C, Chinchilli VM. Small sample characteristics of generalized estimating equations. *Communications in Statistics - Simulation and Computation* 1995; **24**:869–878. DOI: 10.1080/03610919508813280.
5. Fay MP, Graubard BI. Small-sample adjustments for Wald-type tests using sandwich estimators. *Biometrics* 2001; **57**:1198–1206. DOI: 10.1111/j.0006-341X.2001.01198.x.
6. Mancl LA, DeRouen TA. A covariance estimator for gee with improved small-sample properties. *Biometrics* 2001; **57**:126–134. DOI: 10.1111/j.0006-341X.2001.00126.x.
7. Pan W, Wall MM. Small-sample adjustments in using the sandwich estimator in generalized estimating equations. *Statistics in Medicine* 2002; **21**:1429–1441. DOI: 10.1002/sim.1142.
8. Thornquist A, Anderson GL. Small-sample properties of generalized estimating equations in group randomized designs with gaussian response, 1992. Retrieved from author.
9. Kauermann G., Carroll R. J. A note on the efficiency of sandwich covariance matrix estimation. *Journal of the American Statistical Association* 2001; **96**:1387–1396. DOI: 10.1198/016214501753382309.
10. Murray DM, Varnell SP, Blitstein JL. Design and analysis of group randomized trials. *American Journal of Public Health* 2004; **94**:423–432. DOI: 10.2105/AJPH.94.3.423.
11. Pocock SJ, Assman SE, Enos LE, Kasten LE. Subgroup analysis, covariate adjustment and baseline comparisons in clinical trial reporting: current practice and problems. *Statistics in Medicine* 2002; **21**:2917–2930. DOI: 10.1002/sim.1296.
12. Tsiatis AA, Davidian M, Zhang M, Lu X. Covariate adjustment for two-sample treatment comparisons for randomized clinical trials: a principled yet flexible approach. *Statistics in Medicine* 2008; **27**:4658–4677. DOI: 10.1002/sim.3113.
13. Robinson LD, Jewell NP. Some surprising results about covariate adjustment in logistic regression models. *International Statistical Review* 1991; **58**:227–240.
14. Gail MH, Mark SD, Carroll RJ, Green SB, Pee D. On design considerations and randomization-based inference for community intervention trials. *Statistics in Medicine* 1996; **15**:1069–1092. DOI: 10.1002/(SICI)1097-0258(19960615)15:11<1069::AID-SIM220>3.3.CO;2-H.
15. Braun TM, Feng Z. Optimal permutation tests for the analysis of group randomized trials. *Journal of the American Statistical Association* 2001; **96**:1424–1432. DOI: 10.1198/016214501753382336.
16. van der Laan MJ, Rubin D. Targeted maximum likelihood learning. *The International Journal of Biostatistics* 2006; **2**:1–40. DOI: 10.2202/1557-4679.1043.
17. Zhang M, Tsiatis AA, Davidian M. Improving efficiency of inferences in clinical randomized trials using auxiliary covariates. *Biometrics* 2008; **64**:707–715. DOI: 10.1111/j.1541-0420.2007.00976.x.

18. Robins JM. Marginal structural models versus structural nested models as tools for causal inference. In *Statistical Models in Epidemiology: The Environment and Clinical Trials*, Vol. 116, Berry D., Halloran M. E. (eds). Springer-Verlag: NY, 1999; 95–134.
19. Robins JM, Rotnitzky A, Zhao LP. Estimation of regression coefficients when some regressors are not always observed. *Journal of the American Statistical Association* 1994; **89**:846–866. DOI: 10.2307/2290910.
20. Moore KL, van der Laan MJ. Covariate adjustment in randomized trials with binary outcomes: targeted maximum likelihood estimation. *Statistics in Medicine* 2009; **28**:39–64. DOI: 10.1002/sim.3445.
21. Moore KL, van der Laan MJ. Application of time-to-event methods in the assessment of safety in clinical trials. In *Design and Analysis of Clinical Trials with Time-to-Event Endpoints*, Peace KE (ed.). Chapman & Hall: Boca Raton, FL, 2009.
22. Leon A, Tsiatis AA, Davidian M. Semiparametric estimation of treatment effect in a pretest-posttest study. *Biometrics* 2003; **59**:1046–1055. DOI: 10.1111/j.0006-341X.2003.00120.x.
23. Zhang M, Gilbert PB. Increasing the efficiency of prevention trials by incorporating baseline covariates. *Statistical Communications in Infectious Disease* 2010; **2**. DOI: 10.2202/1948-4690.1002.
24. Lipsitz SR, Fitzmaurice GM, Orav EJ, Laird NM. Performance of generalized estimating equations in practical situations. *Biometrics* 1994; **50**:270–278. DOI: 10.2307/2533218.
25. Davidson AC, Hinkley DV. *Bootstrap Methods and Their Applications*. Cambridge University Press: Cambridge, UK, 1997.
26. Kamo N, Carlson M, Brennan RT, Earls F. Young citizens as health agents: use of drama in promoting community efficacy for HIV/AIDS. *American Journal of Public Health* 2008; **98**:201–204. DOI: 10.2105/AJPH.2007.113704.
27. Robins JM. A new approach to causal inference in mortality studies with sustained exposure periods—application to control of the healthy worker survivor effect. *Mathematical Modelling* 1986; **7**:1393–1512.
28. Wang A, Louis TA. Matching conditional and marginal shapes in binary random intercept models using a bridge distribution function. *Biometrika* 2003; **90**:765–775. DOI: 10.1093/biomet/90.4.765.
29. Rubin D, van der Laan MJ. Empirical efficiency maximization: improved locally efficient covariate adjustment in randomized experiments and survival analysis. *The International Journal of Biostatistics* 2008; **4**. DOI: 10.2202/1557-4679.1084.