



An evaluation of quadratic inference functions for estimating intervention effects in cluster randomized trials

Hengshi Yu^{a,*}, Fan Li^{b,c}, Elizabeth L. Turner^{d,e}

^a Department of Biostatistics, University of Michigan, Ann Arbor, MI, 48109, USA

^b Department of Biostatistics, Yale University, New Haven, CT, 06510, USA

^c Center for Methods in Implementation and Prevention Science, Yale University, New Haven, CT, 06510, USA

^d Department of Biostatistics and Bioinformatics, Duke University, Durham, NC, 27710, USA

^e Duke Global Health Institute, Durham, NC, 27710, USA

ARTICLE INFO

Keywords:

Generalized estimating equations
Generalized method of moments
Marginal intervention effect estimation
Population-average intervention effect
Quadratic inference function

ABSTRACT

Cluster randomized trials (CRTs) usually randomize groups of individuals to interventions, and outcomes are typically measured at the individual level. Marginal intervention effects are frequently of interest in CRTs due to their population-averaged interpretations. Such effects are estimated using generalized estimating equations (GEE), or a recent alternative called the quadratic inference function (QIF). However, the performance of QIF relative to GEE have not been extensively evaluated in the CRT context, especially when the marginal mean model includes additional covariates. Motivated by the HALI trial, we conduct simulation studies to compare the finite-sample operating characteristics of QIF and GEE. We demonstrate that QIF and GEE are equivalent under some conditions. When the marginal mean model includes individual-level covariates, QIF shows an efficiency improvement over GEE with overall larger power, but its test size may be more liberal than GEE and GEE achieves better coverage than QIF. The test size inflation may not be fully addressed from using finite-sample bias corrections. The estimates of QIF tend to be closer to GEE in the HALI data, although the former presents a small standard error. Overall, we confirm that the QIF approach generally has potentially better efficiency than GEE in our simulation studies but might be more cautiously used as a viable approach for the analysis of CRTs. More research is needed, however, to address the finite-sample bias in the variance estimation of the QIF to better control its test size.

1. Introduction

A cluster randomized trial (CRT), also referred to as a group randomized trial or a community-randomized trial, is a randomized controlled trial where a cluster (e.g. hospital, village) of individuals is the unit of randomization [1]. The outcomes are usually measured for each individual member within a cluster. As a consequence of the cluster randomization design, the outcomes of individuals within the same cluster are expected to be correlated and this correlation, namely the intraclass correlation coefficient (ICC), must be accounted for in both the design and analysis [2,3].

During the analysis stage, the ICC is commonly accounted for using one of two modeling approaches for individual-level outcome data. One approach is the mixed-effects model, which estimates the cluster-specific (conditional) intervention effect. The other approach is the

generalized estimating equations (GEE), which estimate the population-averaged (marginal) intervention effect [4]. The conditional and marginal intervention effects are equal to each other with the identity or log link, while they could differ with the logit link for binary outcomes [5]. The choice of marginal or conditional models depends on the research objectives and each model has its pros and cons. In particular, while the mixed-effects model requires the correct specification of the random-effects distribution to obtain consistent model-based variance [6], the marginal model is more robust with the sandwich variance being consistent even under misspecification of the correlation structure. For this reason, the marginal model coupled with GEE has been commonly used in the analysis of CRTs [7].

Despite the robustness of the marginal model, misspecification of the correlation structure could lead to reduced efficiency in estimating the intervention effect. To improve the efficiency of GEE, an alternative

* Corresponding author.

E-mail address: hengshi@umich.edu (H. Yu).

<https://doi.org/10.1016/j.conctc.2020.100605>

Received 22 March 2020; Received in revised form 15 June 2020; Accepted 28 June 2020

Available online 5 July 2020

2451-8654/© 2020 The Authors.

Published by Elsevier Inc.

This is an open access article under the CC BY-NC-ND license

(<http://creativecommons.org/licenses/by-nc-nd/4.0/>).

approach – the quadratic inference function (QIF) – has been developed [8,9]. Although it has been shown that QIF may have an efficiency advantage over GEE with correlated data that arise in the longitudinal data settings [8,9], it is not as commonly used in the analysis of CRTs, with a few exceptions [10–12]. The purpose of this article is to provide additional empirical evidence by comparing the performance of QIF and GEE in the CRT scenarios, and discuss several analytical insights of these two methods.

2. Motivating study: the HALI trial

The HALI (Health and Literacy Intervention) trial is a 2 × 2 factorial CRT conducted in Kenya to evaluate the impact of a malaria intervention and a literacy intervention on child health and educational outcomes [13,14]. We focus on the literacy intervention and do not address features of the factorial design in this paper. In other words, we treat the trial as a regular parallel-arm CRT. A multilevel data structure is present because children are nested in schools, which are nested in teacher advisory center; the randomization is carried out at the school level and stratified by each teacher advisory center. We focus on the 9-month outcome and only consider school-level clustering as previous analysis suggested minimal clustering at the teacher advisory center level [15]. There are 51 schools in the literacy intervention arm and 50 schools in the control arm, with approximately 25 children in each school.

The descriptive statistics in the HALI trial are provided in Table 1. The primary outcome is the spelling score, which ranges from 0 to 20 and is treated as a continuous outcome. Other baseline variables included in the trial are age, sex, baseline spelling score (individual-level variables) and the availability of handwashing facilities (school-level variable), which might have a direct or indirect relationship with the primary outcome. All baseline variables (with age log-transformed) are reasonably balanced between the intervention and the control arms due to randomization [15].

In the context of the HALI study, we consider CRTs with a continuous outcome measured at a single follow-up time point. We also assume a relatively large number of clusters (e.g., > 40) with equal cluster size. We first describe the GEE and QIF approaches to estimate the marginal intervention effect and analytically study their connections. The finite-sample operating characteristics of GEE and QIF are evaluated using simulations, and both approaches are applied to the HALI trial for

empirical illustrations.

3. Statistical methods

3.1. Generalized estimating equations (GEE)

For correlated data arising from N clusters and m individuals in each cluster, we use X_{ij} , Y_{ij} , μ_{ij} to denote the p -dimensional design vector, individual-level outcome, and marginal mean of the j th individual in the i th cluster. We define the covariate matrix $X_i = (X_{i1}, X_{i2}, \dots, X_{im})^T$, outcome vector $Y_i = (Y_{i1}, Y_{i2}, \dots, Y_{im})^T$ and mean vector $\mu_i = (\mu_{i1}, \mu_{i2}, \dots, \mu_{im})^T$ for the i th cluster. We write V_i as the working covariance matrix of the i th cluster, β as the p -dimensional regression parameter vector, and define a generalized linear model $g(\mu_{ij}) = X_{ij}^T \beta$, where $g(\cdot)$ is a monotonic and differentiable link function. The marginal variance is defined as $\varphi\nu(\mu_{ij})$, where $\nu(\cdot)$ is a parametric variance function and φ is the common dispersion. We also define the gradient matrix $D_i = \partial\mu_i/\partial\beta^T$.

The generalized estimating equations (GEE) approach was originally developed for longitudinal data analysis based on quasi-likelihood [4, 16] and formulated as

$$\sum_{i=1}^N D_i^T V_i^{-1} (Y_i - \mu_i) = 0.$$

To specify the unknown covariance matrix V_i , a working correlation matrix $R_i(\alpha)$ is assumed and frequently parameterized by a common parameter α . Define the variance matrix as $A_i = \text{diag}\{\nu(\mu_{i1}), \dots, \nu(\mu_{im})\}$, and the covariance matrix V_i can be written as $V_i = \varphi A_i^{1/2} R_i(\alpha) A_i^{1/2}$. The estimation of β and (φ, α) is carried out via the modified Fisher-scoring algorithm [4]. It is known that the GEE estimator $\hat{\beta}$ is consistent for any choice of working correlation structure and has a robust “sandwich” covariance matrix given by

$$V_R = \left(\sum_{i=1}^N D_i^T V_i^{-1} D_i \right)^{-1} \left\{ \sum_{i=1}^N D_i^T V_i^{-1} \text{cov}(Y_i) V_i^{-1} D_i \right\} \left(\sum_{i=1}^N D_i^T V_i^{-1} D_i \right)^{-1}. \tag{1}$$

A consistent estimator for the “sandwich” covariance matrix is obtained by replacing $\text{cov}(Y_i)$ in equation (1) with its empirical version

Table 1

Baseline descriptive statistics in the HALI trial. The total sample size includes $n = 2230$ children in $N = 101$ schools (137 participants have missing baseline spelling scores and excluded in the analysis).

Level	Variable		Overall	Intervention	Control	
			(101 schools)	(51 schools)	(50 schools)	
Child-level	log-Age	n	2230	1103	1127	
		Min	1.6	1.6	1.6	
		Mean	2.0	2.0	2.0	
		Median	2.1	2.1	2.1	
		IQR	1.9~2.2	1.9~2.2	1.9~2.2	
		Max	2.7	2.7	2.6	
		SD	0.22	0.21	0.22	
		Sex	n	2230 (100%)	1103 (100%)	1127 (100%)
			Male	1136 (50.9%)	575 (52.1%)	561 (49.8%)
			Female	1094 (49.1%)	528 (47.9%)	566 (50.2%)
		Baseline spelling score	n	2193	1089	1104
			Min	0	0	0
			Mean	8.2	8.5	7.9
			Median	7	8	7
IQR	5 ~ 11		5 ~ 12	5 ~ 11		
Max	20		20	19		
SD	4.50		4.66	4.33		
School-level	Handwash facilities in school	N	101 (100%)	51 (100%)	50 (100%)	
		Yes	26 (25.7%)	17 (33.3%)	9 (18.0%)	
		No	75 (74.3%)	34 (66.7%)	41 (82.0%)	

$(Y_i - \mu_i)(Y_i - \mu_i)^T$. In the context of CRTs, two most frequently used correlation structure is the independence and exchangeable structure. Interestingly, under some conditions, the GEE estimator provides identical point estimates using either correlation structure [8], as summarized in the following result.

Result 1. *Under the independence and exchangeable working correlations, GEE produces identical point estimate and the robust sandwich covariance if the following two conditions are satisfied: (1) the marginal mean model only includes cluster-level covariates; (2) equal cluster sizes.*

The proof of [Result 1](#) is provided in [Appendix I](#).

In CRTs with correlated binary outcomes, Pan [17] demonstrated the same result assuming a logistic marginal mean model, and [Result 1](#) can be viewed as a generalization of this earlier result to arbitrary link and variance functions. In practice, this result implies that when only the intervention indicator is included in the marginal analysis of CRTs, the point estimate for β is the same regardless of the working correlation specification as long as the cluster sizes are the same. Moreover, [Result 1](#) has importantly implications for sample size and power calculations in the design stage. The equal cluster-size assumption is frequently assumed in designing CRTs [18], in which case the sample size requirements become identical under either working correlation specification.

3.2. Quadratic inference function (QIF)

The approach of quadratic inference function (QIF) was introduced to improve the efficiency of GEE in longitudinal data analysis under correlation misspecification [19]. The QIF approach expresses the inverse of the working correlation matrix as a linear combination of K basis matrices: $R^{-1}(\alpha) = \sum_{k=1}^K \gamma_k M_k$ with M_k as the k th basis matrix and γ_k the weight. The first basis matrix is usually specified as the identity matrix $M_1 = I$. A (Kp) -dimensional score vector is then defined as

$$g_i(\beta) = \begin{pmatrix} D_i^T A_i^{-1} (Y_i - \mu_i) \\ D_i^T A_i^{-1/2} M_2 A_i^{-1/2} (Y_i - \mu_i) \\ \vdots \\ D_i^T A_i^{-1/2} M_K A_i^{-1/2} (Y_i - \mu_i) \end{pmatrix}.$$

Let $\bar{g}_N(\beta) = N^{-1} \sum_{i=1}^N g_i(\beta)$ be the list of extended score equations and $C_N(\beta) = N^{-1} \sum_{i=1}^N g_i(\beta) g_i^T(\beta)$ be the empirical covariance matrix of $g_i(\beta)$. The QIF is written as $Q_N(\beta) = N \bar{g}_N^T(\beta) C_N^{-1}(\beta) \bar{g}_N(\beta)$. Based on the generalized method of moments (GMM) [20], the estimator $\hat{\beta} = \text{argmin}_{\beta} Q_N(\beta)$ can be more efficient than the GEE estimator in large samples when the working correlation is misspecified. From the first derivative of Q_N , the QIF estimator obtained by minimizing the Q_N function is asymptotically equivalent to solving $W_N(\beta) = N(\partial \bar{g}_N / \partial \beta^T) C_N^{-1} \bar{g}_N = \mathbf{0}$ [19], and the Newton-Raphson algorithm can be used to iteratively update the estimator for β until convergence [9]. A consistent variance estimator of the QIF estimator $\hat{\beta}$ then has a sandwich form

$$\begin{aligned} \widehat{\text{cov}}(\hat{\beta}) &= N^{-1} (\partial W_N / \partial \beta)^{-1} \widehat{\text{cov}}(W_N) (\partial W_N / \partial \beta)^{-1} \Big|_{\beta=\hat{\beta}} \\ &= N^{-1} \left[\left\{ \partial \bar{g}_N(\hat{\beta}) / \partial \beta^T \right\} C_N^{-1}(\hat{\beta}) \left\{ \partial \bar{g}_N(\hat{\beta}) / \partial \beta \right\} \right]^{-1}. \end{aligned}$$

Due to the theoretical efficiency improvement of QIF over GEE, there has been increasing efforts in developing the theory of QIF for analyzing correlated data, including the following examples concerning longitudinal data analysis. A QIF likelihood-ratio test statistic, with asymptotic Chi-squared distribution, was proposed and shown useful to test for goodness-of-fit in longitudinal studies [19]. A penalized version of QIF can further accommodate the variable selection [21]. The Godambe Information (TGI) criterion and the trace of the empirical covariance matrix were developed to select the appropriate correlation structure

[22,23]. The QIF approach has also been shown to automatically down-weight outlying observations [8], while GEE has unbounded influence function and can be sensitive to outliers. In addition, the QIF approach can also be utilized for meta-analysis with a flexible joint estimation procedure [24]. In small to moderately-sized samples, it was found that the standard errors of parameters can be severely downward-biased and two biased-corrected covariance estimators have shown to provide adequate finite-sample adjustments [9]. Furthermore, it was shown that imbalance of covariate distributions and of cluster sizes can also lead to larger variability of the QIF estimator [25].

In contrast to the longitudinal data setting, in a CRT with a single follow-up time-point, there is no natural ordering or structure of individuals in the same cluster. That is, decay-type structures are not appropriate but the exchangeable working correlation structure with a common ICC parameter is a natural choice. For the exchangeable working correlation structure, we can write $R(\alpha) = (1 - \rho)I + \rho J$, where J is a $(m \times m)$ -dimensional matrix of 1's. Following Li et al. [26], we have $R^{-1}(\alpha) = (1 - \rho)^{-1}I - \rho(1 - \rho)^{-1}\{1 + (m - 1)\rho\}^{-1}J$. As J is not a full-rank matrix, a better use is to specify I and $J - I$ as two full-rank basis matrices for exchangeable correlation structure of QIF [9]. We can also use QIF with an independence correlation structure with only one basis matrix of the identity matrix I . Parallel to [Result 1](#), we provide an additional insight that the results obtained from GEE and QIF are equal under conditions assumed below.

Result 2. *Under the exchangeable working correlation, QIF and GEE produce identical point estimates and robust covariances if two conditions are satisfied: (1) the marginal mean model only includes cluster-level covariates; (2) equal cluster sizes.*

The proof of [Result 2](#) is provided in [Appendix I](#).

Furthermore, it has been pointed out previously that QIF and GEE are identical when the independence working correlation is used [8,25]. Combining [Result 1](#) and [Result 2](#), we further summarize an additional result as follows.

Result 3. *The point estimates and robust covariances are identical using either GEE or QIF with either independence or exchangeable working correlation, if the following two conditions are satisfied: (1) the marginal mean model only includes cluster-level covariates; (2) equal cluster sizes.*

The proof of [Result 3](#) is also provided in [Appendix I](#).

As previously indicated, these results have implications for sample size procedures assuming equal cluster sizes, in which cases the sample requirements will be equivalent using either GEE or QIF coupled with either independence and exchangeable correlation structures.

4. Simulation studies

To study the empirical performance of QIF in CRTs, we carry out a series of simulation studies. Inspired by the HALI trial in our motivating example, we assume 100 clusters with 50 clusters randomized to the treatment and control arms. For simplicity, we specify the cluster sizes to be 25 for all 100 clusters. We consider two mean models and four types of correlation structures to form different data generating process (DGP) in CRTs. We then implement GEE and QIF with the correctly-specified mean model and exchangeable working correlation structure. In each simulation, we use a multivariate normal model to simulate outcomes in each cluster with individual-level variance $\varphi = 4$. The four correlation structures used in the DGP include: fixed-regular exchangeable (CS0), regular exchangeable (CS1), cluster-specific exchangeable (CS2) and cluster-specific exchangeable with sub-clustering (CS3). The two selected mean models are: intervention-only model (MM1) and covariate-adjusted model (MM2).

In the intervention-only mean model (MM1), we use X_i to indicate whether the i th cluster is in the intervention arm ($X_i = 1$) or not ($X_i = 0$). The MM1 is given by $\mu_{ij} = \beta_0 + \beta_1 X_i$. We fix $\beta_0 = 1$ and allow the marginal intervention effect β_1 to adopt a range of different values in the

DGP. For MM2, we simulate four covariates whose distributions are informed by the HALI trial. Specifically, we simulate three individual-level covariates B_{ij} , C_{ij} and D_{ij} mimicking the age, sex and baseline spelling score; we assume B_{ij} follows a log-normal distribution with mean of 2 and standard deviation 0.2 (both on the log scale), C_{ij} follows a binomial distribution with probability of 0.5 and D_{ij} follows a normal distribution with mean 8 and standard deviation 5. We also simulate a cluster-level covariate E_i from a binomial distribution with probability of 0.26 based on the school-level prevalence of handwash facilities. We write MM2 as $\mu_{ij} = \beta_0 + \beta_1 X_i + \beta_B B_{ij} + \beta_C C_{ij} + \beta_D D_{ij} + \beta_E E_i$. We specify different values for the intervention effect β_1 in the DGP, while keeping all other covariate effects constant as $\beta_0 = \beta_B = \beta_C = \beta_D = \beta_E = 1$.

For the correlation structures, we first assume the fixed-regular exchangeable correlation structure (CS0) that all clusters have the same exchangeable correlation structure with ICC $\rho = 0.05$. For the regular exchangeable correlation structure (CS1), we assume all the clusters have the same exchangeable correlation matrices in the same simulation iteration but might vary in different iterations. The cluster-specific exchangeable correlation structure (CS2) further allows possibly different exchangeable correlation matrices for each cluster in each simulation iteration. Both CS1 and CS2 assume the ICC ρ in the exchangeable correlation structure is sampled from a uniform distribution from 0.01 to 0.2. For the cluster-specific exchangeable correlation structure with sub-clustering (CS3), we first generate a variable F_{ij} from a discrete uniform distribution in $\{1, 2, 3, 4\}$. For the j th and k th individuals in the i th cluster, we specify their pairwise correlation value to be $R_{ijk} = 0.5$ if F_{ij} and F_{ik} are the same, indicating most closely-correlated. If F_{ij} and F_{ik} differ in 1, 2 or 3, we specify the correlation R_{ijk} to be $0.5^2, 0.5^3$ or 0.5^4 , respectively, in order to model different degrees of pairwise correlations. Particularly, except for CS0, we have assumed the true correlation matrix is sampled from a population-level distribution; the purpose of this additional step is to provide a data generating process with the desired marginal mean and a complex

correlation structure, but without the multivariate normality assumption. This type of DGP is less restrictive than the usual multivariate normal DGP, and we provide additional technical details on this type of DGP in Appendix II. For CS0, CS1 and CS2, we use five different values for $\beta_1 \in \{0, 0.40, 0.45, 0.50, 0.55\}$ in the DGP, and use six values for $\beta_1 \in \{0, 0.40, 0.50, 0.60, 0.70, 0.80\}$ under CS3, leading to a total of 42 scenarios.

We generate $S = 3000$ replicates for each of 8 combinations of the 2 mean models and 4 correlation structures. Two models are fit for each simulated data, namely GEE with exchangeable working correlation and QIF with exchangeable working correlation. As we explain in Appendix II, the working correlation matrix will be incorrectly specified under DGP with CS2 and CS3. In all scenarios, the correct mean model is specified in the analysis, namely an intervention-only mean model is fit for data generated under MM1 and the covariate-adjusted mean model is fit for data generated under MM2.

The following metrics are used to compare the performance of GEE and QIF: (1) relative bias (RBS), which equals the bias relative to the true effect, (2) empirical standard error (ESE) of the estimates, and, (3) the mean robust standard error (MRSE) over all 3000 replicates. In addition, we set the nominal type I error rate to be 0.05 to obtain power for the Wald-type Z-test. We also calculate the power ratio (PR) defined by the power of QIF analysis relative to that of GEE analysis (Q/G). We additionally calculate the 95% coverage probability (Coverage) for both GEE and QIF. Finally, the empirical type I error rate is estimated assuming a true null intervention effect ($\beta_1 = 0$) in the DGP.

We present results for DGP with MM1 and four different correlation structures in Table 2. The results indicate that GEE and QIF lead to the exact same estimates, and confirms the analytical insights from Result 2. With the exchangeable working correlation structure and balanced cluster sizes, MM1 only includes the cluster-level intervention variable, satisfying the conditions listed in Result 2. In this setting, the relative biases are small for both GEE and QIF, and the coverage probabilities are all close to nominal. The type I error rate, which is the power when $\beta_1 =$

Table 2

The GEE and QIF results for data generated under mean model MM1 with covariance structures CS0, CS1, CS2, CS3 and different true β_1 's based on data from 3000 replicates for each scenario ($S = 3000$).

DGP: MM1		RBS		ESE		MRSE		Power		Coverage	
Analysis: MM1 ^a		GEE	QIF	GEE	QIF	GEE	QIF	GEE	QIF	GEE	QIF
CS0	$\beta_1 = 0$	-	-	-	-	-	-	5.57%	5.57%	94.43%	94.43%
	$\beta_1 = 0.40$	- 0.98%	- 0.98%	0.118	0.118	0.117	0.117	91.97%	91.97%	94.67%	94.67%
	$\beta_1 = 0.45$	- 0.87%	- 0.87%	0.118	0.118	0.117	0.117	96.70%	96.70%	94.67%	94.67%
	$\beta_1 = 0.50$	- 0.78%	- 0.78%	0.118	0.118	0.117	0.117	98.60%	98.60%	94.67%	94.67%
	$\beta_1 = 0.55$	0.20%	0.20%	0.119	0.119	0.117	0.117	99.60%	99.60%	94.27%	94.27%
CS1	$\beta_1 = 0$	-	-	-	-	-	-	5.87%	5.87%	94.13%	94.13%
	$\beta_1 = 0.40$	- 0.35%	- 0.35%	0.147	0.147	0.143	0.143	78.17%	78.17%	95.20%	95.20%
	$\beta_1 = 0.45$	- 0.80%	- 0.80%	0.148	0.148	0.145	0.145	84.93%	84.93%	94.77%	94.77%
	$\beta_1 = 0.50$	0.22%	0.22%	0.150	0.150	0.145	0.145	89.77%	89.77%	94.60%	94.60%
	$\beta_1 = 0.55$	- 0.53%	- 0.53%	0.148	0.148	0.146	0.146	94.03%	94.03%	94.93%	94.93%
CS2	$\beta_1 = 0$	-	-	-	-	-	-	5.30%	5.30%	94.70%	94.70%
	$\beta_1 = 0.40$	1.00%	1.00%	0.152	0.152	0.148	0.148	77.43%	77.43%	93.93%	93.93%
	$\beta_1 = 0.45$	- 0.40%	- 0.40%	0.150	0.150	0.148	0.148	85.60%	85.60%	94.10%	94.10%
	$\beta_1 = 0.50$	- 1.03%	- 1.03%	0.149	0.149	0.148	0.148	91.10%	91.10%	94.60%	94.60%
	$\beta_1 = 0.55$	- 0.57%	- 0.57%	0.146	0.146	0.148	0.148	95.53%	95.53%	94.80%	94.80%
CS3	$\beta_1 = 0$	-	-	-	-	-	-	5.30%	5.30%	94.70%	94.70%
	$\beta_1 = 0.40$	- 0.60%	- 0.60%	0.215	0.215	0.211	0.211	47.70%	47.70%	94.20%	94.20%
	$\beta_1 = 0.50$	0.30%	0.30%	0.219	0.219	0.212	0.212	65.30%	65.30%	94.00%	94.00%
	$\beta_1 = 0.60$	- 0.82%	- 0.82%	0.216	0.216	0.212	0.212	78.97%	78.97%	94.53%	94.53%
	$\beta_1 = 0.70$	- 0.26%	- 0.26%	0.212	0.212	0.212	0.212	90.63%	90.63%	94.60%	94.60%
	$\beta_1 = 0.80$	1.13%	1.13%	0.213	0.213	0.212	0.212	96.53%	96.53%	94.10%	94.10%

^a The analysis utilizes mean model MM1 and exchangeable working correlation matrix with robust SE.

0, are also close to 0.05. In addition, in each scenario, the ESE and the MRSE are close to each other, indicating that the robust variance estimators are consistent for both GEE and QIF. Notably, the results in scenarios with CS3 and $\beta_1 = 0.4$ or $\beta_1 = 0.5$ have smaller power than those with CS0 and CS1, suggesting that both QIF and GEE could be less efficient with a misspecified correlation structure [19].

Table 3 presents the results for the DGP with MM2 and four correlation structures. Our analytical finding from Result 2 does not apply with MM2 which now includes multiple individual-level covariates. As a result, the GEE and QIF results differ, although both approaches give small relative bias across all scenarios. Interestingly, while GEE has smaller ESE compared to QIF, QIF presents smaller MRSE than GEE. This finding suggests that the robust variance of QIF tends to be biased towards zero under this complex mean model. The downward bias of QIF variance estimator further leads to under-coverage of the interval estimator, and a type I error inflation especially when the correlation structure deviates from CS0. In contrast, the type I error rate and coverage of the GEE estimator are more close to nominal throughout. The results in Table 3 also allow us to compare the efficiency between QIF and GEE, by comparing the ESE and the power. Under CS0 and CS1, GEE and QIF have almost identical power under the alternative, confirming that their results are asymptotically equivalent under correct

correlation specification. When the working correlation model is misspecified, the power of both GEE and QIF will decrease, and QIF appears to be slightly more efficient, as evidenced by the results under CS3. Throughout, the power ratio of QIF over GEE is at most slightly larger than 1 (the largest increase in power is 4% for QIF over GEE under CS3), and becomes closer to 1 as the effect size increases. However, one should be cautious in interpreting the efficiency advantage of QIF because (a) QIF carries an inflated type I error rate under the null and (b) QIF interval estimator frequently leads to under-coverage due to the negative bias in its robust variance estimator.

In an effort to potentially reduce the bias in QIF variance estimator under MM2, we additionally explore two bias-corrections for the variance proposed by Westgate [9]. These two bias-corrections are extensions of the MancI-DeRouen (MD) and Kauermann-Carroll (KC) methods proposed in the GEE literature [27,28], and we denote them by QIF_{MD} and QIF_{KC}. The details of the two bias-correction methods for QIF are given in Appendix III. Table 4 shows that the type I error rates from the two bias-correction methods are closer to the nominal 0.05 level than that from QIF without correction. However, the type I error rate of bias-corrected QIF remains liberal and consistently larger than that of GEE, cautioning the use of QIF in CRTs when the marginal mean model is complex with multiple individual-level covariates.

Table 3

The GEE and QIF results for MM2 with CS0, CS1, CS2, CS3 and different true β_1 's, each of which is from 3000 simulations ($S = 3000$).

DGP: MM2		RBS		ESE		MRSE		Power		PR	Coverage	
Analysis: MM2 ^a		GEE	QIF	GEE	QIF	GEE	QIF	GEE	QIF	Q/G	GEE	QIF
CS0	$\beta_1 = 0$	-	-	-	-	-	-	5.27%	6.93%	-	94.73%	93.07%
	$\beta_1 = 0.40$	- 0.83%	- 1.03%	0.120	0.123	0.117	0.114	91.47%	91.67%	1.0022	94.50%	92.97%
	$\beta_1 = 0.45$	- 0.09%	0.02%	0.122	0.127	0.117	0.114	96.47%	96.30%	0.9983	93.53%	91.90%
	$\beta_1 = 0.50$	0.62%	0.64%	0.120	0.124	0.117	0.114	98.67%	98.57%	0.9990	94.40%	92.67%
	$\beta_1 = 0.55$	0.18%	0.22%	0.116	0.120	0.117	0.114	99.70%	99.70%	0.9990	94.37%	93.13%
CS1	$\beta_1 = 0$	-	-	-	-	-	-	6.07%	7.83%	-	93.93%	92.17%
	$\beta_1 = 0.40$	- 0.33%	- 0.10%	0.150	0.154	0.146	0.142	77.00%	78.37%	1.0178	94.87%	93.47%
	$\beta_1 = 0.45$	0.58%	0.73%	0.154	0.159	0.146	0.142	84.67%	84.73%	1.0001	93.27%	91.80%
	$\beta_1 = 0.50$	0.56%	0.48%	0.151	0.156	0.145	0.141	90.10%	90.13%	1.0003	94.17%	92.93%
	$\beta_1 = 0.55$	- 0.35%	- 0.64%	0.157	0.163	0.146	0.142	92.87%	92.90%	1.0003	93.77%	91.73%
CS2	$\beta_1 = 0$	-	-	-	-	-	-	5.67%	7.30%	-	94.33%	92.70%
	$\beta_1 = 0.40$	0.48%	0.65%	0.151	0.153	0.148	0.143	77.17%	78.73%	1.0202	94.40%	93.20%
	$\beta_1 = 0.45$	0.58%	0.69%	0.154	0.158	0.148	0.143	85.33%	86.57%	1.0145	94.37%	92.03%
	$\beta_1 = 0.50$	0.01%	0.06%	0.151	0.154	0.148	0.143	91.80%	92.23%	1.0003	94.93%	93.53%
	$\beta_1 = 0.55$	- 0.35%	- 0.64%	0.152	0.155	0.148	0.143	95.30%	95.40%	1.0010	94.10%	92.43%
CS3	$\beta_1 = 0$	-	-	-	-	-	-	5.93%	7.17%	-	94.07%	92.83%
	$\beta_1 = 0.40$	1.05%	1.13%	0.221	0.228	0.212	0.206	47.93%	50.07%	1.0446	94.07%	92.10%
	$\beta_1 = 0.50$	- 0.14%	- 0.22%	0.220	0.227	0.212	0.206	64.90%	66.60%	1.0262	93.67%	91.87%
	$\beta_1 = 0.60$	1.26%	1.15%	0.215	0.221	0.211	0.205	81.97%	82.50%	1.0065	94.10%	92.67%
	$\beta_1 = 0.70$	- 0.04%	- 0.09%	0.216	0.223	0.212	0.206	90.67%	90.73%	1.0007	94.13%	92.93%
	$\beta_1 = 0.80$	- 0.25%	- 0.29%	0.213	0.220	0.212	0.206	96.37%	96.37%	1.0000	94.53%	93.33%

^a The analysis utilizes mean model MM2 and exchangeable working correlation matrix with robust SE.

Table 4

The GEE, QIF and two corrected QIF results for MM2 with CS0, CS1, CS2, CS3, each of which is from 3000 simulations ($S = 3000$).

DGP: MM2	Method	ESE	MRSE	Power (Type I error)	
Analysis: MM2 ^a					
CS0	$\beta_1 = 0$	GEE	0.119	0.117	5.27%
		QIF	0.123	0.114	6.93%
		QIF _{MD}	0.123	0.117	5.93%
		QIF _{KC}	0.123	0.115	6.33%
CS1	$\beta_1 = 0$	GEE	0.152	0.145	6.07%
		QIF	0.159	0.141	7.83%
		QIF _{MD}	0.159	0.146	7.27%
		QIF _{KC}	0.159	0.143	7.50%
CS2	$\beta_1 = 0$	GEE	0.152	0.148	5.67%
		QIF	0.155	0.143	7.30%
		QIF _{MD}	0.155	0.148	6.43%
		QIF _{KC}	0.155	0.145	6.90%
CS3	$\beta_1 = 0$	GEE	0.215	0.212	5.93%
		QIF	0.220	0.206	7.17%
		QIF _{MD}	0.220	0.212	6.33%
		QIF _{KC}	0.220	0.209	6.73%

^a The analysis utilizes mean model MM2 and exchangeable working correlation matrix with robust SE and bias-corrected SEs.

As pointed out by a reviewer, a possible explanation for the bias of the variance estimator for QIF and the associated inflated type I error could be the large variability of the empirical weighting or covariance matrix $C_N(\beta)$. Previous studies have suggested that including additional covariates in the mean model could lead to increased variability in estimating $C_N(\beta)$, which affects the efficiency of the QIF estimator [23, 25, 29, 30]. It remains to be explored whether improved estimation of $C_N(\beta)$ along the lines of Westgate [29] and Westgate [30] coupled with the bias-corrected variances could reduce the negative bias in the robust variance of QIF and improve the coverage rate of the interval estimator.

5. Analysis of HALI trial

We apply GEE and QIF to the analysis of the HALI trial and focus on the continuous outcome of spelling score at the 9-month follow-up [15]. From the descriptive statistics, we observe the baseline covariates are approximately the same across arms, but the mean value of the spelling score at 9-month follow-up differs between the arms, indicating a potential intervention effect due to the literacy intervention. In the marginal mean model, we sequentially include the cluster-level intervention X_i , age B_{ij} , sex C_{ij} , presence of handwash facilities E_i and baseline spelling score D_{ij} . We utilize both independence and exchangeable correlation structures based on three models on the marginal mean outcome μ_{ij} in Table 5 to estimate the unadjusted and covariate-adjusted literacy intervention effects. The intervention parameter is denoted as β_1 . Specifically, the first mean model is an unadjusted model with intervention indicator as the only covariate, which corresponds to MM1 in simulation studies. The second and third mean models are covariate-adjusted models with three and four other covariates, with the third model corresponding to MM2 in simulation studies. We denote GEE_{ind} and GEE_{exc} as GEE with the independence and exchangeable correlation structures, respectively. The notations of QIF_{ind} and QIF_{exc}

Table 5
Mean models of the 9-month spelling score.

Model	Formulation
1	$\mu_{ij} = E(Y_{ij}) = \beta_0 + \beta_1 X_i$
2	$\mu_{ij} = E(Y_{ij}) = \beta_0 + \beta_1 X_i + \beta_B B_{ij} + \beta_C C_{ij} + \beta_E E_i$
3	$\mu_{ij} = E(Y_{ij}) = \beta_0 + \beta_1 X_i + \beta_B B_{ij} + \beta_C C_{ij} + \beta_D D_{ij} + \beta_E E_i$

are similarly used for QIF. In addition, we also consider the bias-correction techniques of QIF [9] for the two correlation structures and denote them as QIF_{ind-MD} , QIF_{ind-KC} , QIF_{exc-MD} and QIF_{exc-KC} .

We summarize the results in Table 6. The intervention effect estimates from GEE and QIF are generally close to each other under mean model 1, but may be slightly different under mean model 2 and 3. Specifically, the intervention effect estimate tends to be larger using QIF and assuming working exchangeable correlation compared to the rest of methods. Although the standard error of the uncorrected variance of QIF appears to be the smallest, it may carry negative bias as suggested in simulation studies. The two bias-corrected variances of QIF could slightly reduce the bias and improve the variance estimator. For example, the MD corrected QIF standard error is close to the GEE standard error with the exchangeable working correlation. Overall, the standard error estimates are similar across methods under each specific mean model, suggesting that the application of QIF may have limited efficiency improvement over GEE in this data example. However, across the mean models, the standard error estimates for all methods sharply decrease when mean model 3 is considered compared to the rest of mean models, suggesting a strong predictive effect of the baseline spelling score.

6. Discussion

In this paper, we compare the QIF approach with the more commonly-used GEE approach for the estimation of intervention effect in CRTs. In particular, we focus on CRTs with continuous outcomes at

Table 6
The results of the intervention effects on the 9-month spelling score.

Model	Method	Estimate	S.E.	95% confidence interval	p-value
1	GEE_{ind}	1.766	0.4813	(0.823, 2.709)	0.00024
	GEE_{exc}	1.758	0.4819	(0.813, 2.703)	0.00026
	QIF_{ind}	1.766	0.4813	(0.822, 2.709)	0.00024
	QIF_{ind-MD}	1.766	0.4913	(0.803, 2.728)	0.00033
	QIF_{ind-KC}	1.766	0.4863	(0.812, 2.719)	0.00028
	QIF_{exc}	1.797	0.4748	(0.866, 2.727)	0.00015
	QIF_{exc-MD}	1.797	0.4847	(0.847, 2.747)	0.00021
	QIF_{exc-KC}	1.797	0.4797	(0.856, 2.737)	0.00018
	2	GEE_{ind}	1.811	0.4758	(0.878, 2.744)
GEE_{exc}		1.842	0.4774	(0.906, 2.778)	0.00011
QIF_{ind}		1.811	0.4758	(0.879, 2.744)	0.00014
QIF_{ind-MD}		1.811	0.4943	(0.842, 2.780)	0.00025
QIF_{ind-KC}		1.811	0.4849	(0.861, 2.762)	0.00019
QIF_{exc}		2.056	0.4583	(1.158, 2.955)	0.00001
QIF_{exc-MD}		2.056	0.4734	(1.128, 2.984)	0.00001
QIF_{exc-KC}		2.056	0.4659	(1.143, 2.969)	0.00001
3		GEE_{ind}	1.413	0.2868	(0.851, 1.975)
	GEE_{exc}	1.446	0.2934	(0.871, 2.021)	8.3×10^{-7}
	QIF_{ind}	1.413	0.2868	(0.851, 1.975)	8.4×10^{-7}
	QIF_{ind-MD}	1.413	0.2980	(0.829, 1.997)	2.1×10^{-6}
	QIF_{ind-KC}	1.413	0.2924	(0.840, 1.986)	1.3×10^{-6}
	QIF_{exc}	1.601	0.2817	(1.049, 2.153)	1.3×10^{-8}
	QIF_{exc-MD}	1.601	0.2922	(1.028, 2.173)	4.3×10^{-8}
	QIF_{exc-KC}	1.601	0.2872	(1.038, 2.164)	2.5×10^{-8}

one follow-up time point, a large number of clusters and equal cluster sizes, similar to the motivating data example. We present three analytical results on the equivalence between GEE and QIF under specific conditions to explicitly acknowledge their connections. Our simulation studies also confirm these analytical results. Although our simulations show a potential power advantage of QIF over GEE, the inflated type I error rate of QIF cautions its use when the marginal mean model includes multiple baseline covariate. On the other hand, the GEE approach performs quite stable under complex mean and correlation models in our setting with a large number of clusters.

Our simulation results suggest that the two specific limitations of QIF in the application to CRTs, which may be addressed by further research. First, we found an inflated type I error rate of QIF when the DGP concerns a complex marginal mean model MM2. Surprisingly, the application of the two bias-correction techniques does not fully address this issue, as the empirical type I error rate is still above 7% across 3000 simulations. The type I error inflation is mostly due to the negative bias of the QIF variance estimator. In CRTs, a better control of type I error rate may be achieved by permutation test. In future studies, one could consider developing the marginal-model-based permutation analysis as in Braun and Feng [31] and Li et al. [32] for QIF to achieve better finite-sample properties. Second, despite a potential power advantage of QIF over GEE suggested by Table 3, we additionally saw a larger ESE for QIF compared to GEE. In fact, if we define the relative efficiency based on the ratio of the ESE, then Table 3 suggests that QIF could be slightly less efficient than GEE, even though the asymptotic theory states otherwise. As we explain in the simulations, the reduced efficiency of QIF may be attributed to the large variability in estimating the empirical weighting or covariance matrix $C_N(\beta)$ under a complex mean model

APPENDIX I. Proof of Result 1, 2, 3

Result 1

Suppose that there are only cluster-level covariate vector for each individual in the i th cluster with covariate-vector denoted by T_i . We have $g(\mu_{ij}) = T_i^T \beta$ and the individual-level variance $\sigma_{ij}^2 = \varphi\nu(\mu_{ij})$. With only cluster-level covariates, the individual-level means and also variances are equal in the i th cluster with $\mu_{ij} = \mu_{ik}$ and $\sigma_{ij}^2 = \sigma_{ik}^2$ for $j \neq k$. For convenience, we set $\pi_i = \mu_{ij}$, $\omega_i = \sigma_{ij}^2 = \varphi\nu(\mu_{ij})$ and $\eta_i = \partial\pi_i/\partial\beta$. Thus, we can express the cluster-level mean vector as follows: $\mu_i = (\mu_1, \mu_2, \dots, \mu_m)^T = \pi_i \mathbf{1}$, $D_i = \partial\mu_i/\partial\beta = \mathbf{1}\eta_i^T$ and $A_i = \varphi^{-1}\omega_i \mathbf{I}$.

For an independence working correlation structure, we can simplify the GEE with as follows: $\sum_{i=1}^N D_i^T V_i^{-1} (Y_i - \mu_i) = \sum_{i=1}^N \omega_i^{-1} \eta_i^T (Y_i - \mu_i) = 0$. Furthermore, the ‘‘sandwich’’ covariance matrix for the independence working correlation matrix can be denoted by

$$V_{R,ind} = m^{-2} \left(\sum_{i=1}^N \omega_i^{-1} \eta_i \eta_i^T \right)^{-1} \left(\sum_{i=1}^N \omega_i^{-2} \eta_i \mathbf{1}^T \text{cov}(Y_i) \mathbf{1} \eta_i^T \right) \left(\sum_{i=1}^N \omega_i^{-1} \eta_i \eta_i^T \right)^{-1}$$

which shares the same form with the consistent covariance estimator $\widehat{V}_{R,ind}$ with only $\text{cov}(Y_i)$ being replaced by $(Y - \mu_i)(Y - \mu_i)^T$.

For an exchangeable working correlation structure, there is some value ρ for which $R(\alpha) = (1-\rho)\mathbf{I} + \rho\mathbf{J}$ and $R^{-1}(\alpha) = (1-\rho)^{-1}\mathbf{I} - \rho(1-\rho)^{-1}\{1 + (m-1)\rho\}^{-1}\mathbf{J}$. We thus let $\alpha_1 = (1-\rho)^{-1}$ and $\alpha_2 = -\rho(1-\rho)^{-1}\{1 + (m-1)\rho\}^{-1}$. As a consequence, we can simplify the GEE as follows:

$$\sum_{i=1}^N D_i^T V_i^{-1} (Y_i - \mu_i) = (\alpha_1 + m\alpha_2) \sum_{i=1}^N \omega_i^{-1} \eta_i \mathbf{1}^T (Y_i - \mu_i) = 0.$$

Since $\alpha_1 + m\alpha_2 = \{1 + (m-1)\rho\}^{-1} \neq 0$ for $0 \leq \rho < 1$, we can further drop out the non-zero scalar $\alpha_1 + m\alpha_2$ to give $\sum_{i=1}^N \omega_i^{-1} \eta_i \mathbf{1}^T (Y_i - \mu_i) = 0$. Therefore, the estimating equations reduce to the same form for GEE with independence and exchangeable correlation structures. Furthermore, the ‘‘sandwich’’ covariance matrix for the exchangeable working correlation is

$$V_{R,exc} = m^{-2} \left(\sum_{i=1}^N \omega_i^{-1} \eta_i \eta_i^T \right)^{-1} \left(\sum_{i=1}^N \omega_i^{-2} \eta_i \mathbf{1}^T \text{cov}(Y_i) \mathbf{1} \eta_i^T \right) \left(\sum_{i=1}^N \omega_i^{-1} \eta_i \eta_i^T \right)^{-1}$$

with $\text{cov}(Y_i)$ replaced by $(Y - \mu_i)(Y - \mu_i)^T$ for $\widehat{V}_{R,exc}$. Therefore, GEE with independence and exchangeable correlation structures share the same estimating equation form as well as robust ‘‘sandwich’’ covariance estimator, which further implies numerically equivalent estimators and robust covariance estimators.

[25]. Additionally studies are required to systematically evaluate whether improved estimation of $C_N(\beta)$ [25,29,30] can lead to better efficiency of QIF and eventually address the bias of its variance estimates.

To conclude, our empirical evaluation supports the use of GEE over QIF in CRTs with a large number of clusters. We observe that the QIF could exhibit inflated type I error rate and under-coverage when the marginal mean model includes baseline adjustment variables other than the intervention status, while the GEE approach performs consistently well across all scenarios. The bias-corrected variances of QIF also shows limited improvement in terms of type I error rate, and more empirical evaluations of QIF are required to clearly demonstrate its theoretical efficiency advantage over GEE before recommending its routine applications in CRTs.

Funding

This research did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors.

Acknowledgements

The paper was done in part as a master’s thesis of the first author. We thank Simon Brooker and Matthew Jukes, Principal Investigators of the HALI trial, for agreeing to the use of data. We thank all the participants of the study for providing the data. We also thank Huiman X. Barnhart and Yuliya V. Lokhnygina for reviewing the manuscript and helpful discussions as the thesis committee members at the various stages of this work.

Result 2

For convenience, we set $\boldsymbol{\varepsilon}_i = Y_i - \boldsymbol{\mu}_i$. We have $\mathbf{M}_1 = \mathbf{I}$ and $\mathbf{M}_2 = \mathbf{J} + s\mathbf{I}$ with $s \in \mathbb{R}$ for a more general choice of basis matrices of QIF with the exchangeable correlation structure. We use $s = -1$ specifically in our numerical studies. For general exchangeable correlation structure, we have

$$\bar{\mathbf{g}}_N(\boldsymbol{\beta}) = \frac{\varphi}{N} \sum_{i=1}^N \begin{Bmatrix} \omega_i^{-1} \boldsymbol{\eta}_i \mathbf{1}^T \boldsymbol{\varepsilon}_i \\ \omega_i^{-1} \boldsymbol{\eta}_i (m+s) \mathbf{1}^T \boldsymbol{\varepsilon}_i \end{Bmatrix} = \frac{\varphi}{N} \begin{pmatrix} 1 \\ m+s \end{pmatrix} \otimes \sum_{i=1}^N \omega_i^{-1} \boldsymbol{\eta}_i \mathbf{1}^T \boldsymbol{\varepsilon}_i,$$

$$\frac{\partial \bar{\mathbf{g}}_N(\boldsymbol{\beta})}{\partial \boldsymbol{\beta}} = N^{-1} \varphi \sum_{i=1}^N \begin{Bmatrix} -m \omega_i^{-1} \boldsymbol{\eta}_i \boldsymbol{\eta}_i^T \\ -(m^2 + sm) \omega_i^{-1} \boldsymbol{\eta}_i \boldsymbol{\eta}_i^T \end{Bmatrix} = -N^{-1} \varphi m \begin{pmatrix} 1 \\ m+s \end{pmatrix} \otimes \sum_{i=1}^N \omega_i^{-1} \boldsymbol{\eta}_i \boldsymbol{\eta}_i^T,$$

and

$$\mathbf{C}_N(\boldsymbol{\beta}) = N^{-1} \varphi^2 \sum_{i=1}^N \omega_i^{-2} \begin{Bmatrix} \boldsymbol{\eta}_i \mathbf{1}^T \\ (m+s) \boldsymbol{\eta}_i \mathbf{1}^T \end{Bmatrix} \boldsymbol{\varepsilon}_i \boldsymbol{\varepsilon}_i^T \begin{Bmatrix} \boldsymbol{\eta}_i \mathbf{1}^T \\ (m+s) \boldsymbol{\eta}_i \mathbf{1}^T \end{Bmatrix}^T = N^{-1} \varphi^2 \begin{Bmatrix} 1 & m+s \\ m+s & (m+s)^2 \end{Bmatrix} \otimes \sum_{i=1}^N \omega_i^{-2} \boldsymbol{\eta}_i \mathbf{1}^T \boldsymbol{\varepsilon}_i \boldsymbol{\varepsilon}_i^T \mathbf{1} \boldsymbol{\eta}_i^T.$$

We set $\boldsymbol{\Theta} = \sum_{i=1}^N \omega_i^{-1} \boldsymbol{\eta}_i \boldsymbol{\eta}_i^T$, $\boldsymbol{\Delta} = \sum_{i=1}^N \omega_i^{-2} \boldsymbol{\eta}_i \mathbf{1}^T \boldsymbol{\varepsilon}_i \boldsymbol{\varepsilon}_i^T \mathbf{1} \boldsymbol{\eta}_i^T$ and $\boldsymbol{\Sigma} = \sum_{i=1}^N \omega_i^{-1} \boldsymbol{\eta}_i \mathbf{1}^T \boldsymbol{\varepsilon}_i$. As is known, QIF has the asymptotic estimating equation $\mathbf{W}_N(\boldsymbol{\beta}) = \mathbf{0}$, which is implemented for the QIF estimator. We have

$$\mathbf{W}_N(\boldsymbol{\beta}) = N(\partial \bar{\mathbf{g}}_N / \partial \boldsymbol{\beta}^T) \mathbf{C}_N^{-1} \bar{\mathbf{g}}_N = (-m \times 1) \otimes (\boldsymbol{\Theta}^T \boldsymbol{\Delta}^{-1} \boldsymbol{\Sigma}) = \mathbf{0},$$

which further implies that $\boldsymbol{\Theta}^T \boldsymbol{\Delta}^{-1} \boldsymbol{\Sigma} = \mathbf{0}$. Both $\boldsymbol{\Theta}$ and $\boldsymbol{\Delta}$ are positive-definite and non-singular, the estimating equation is equivalent to $\boldsymbol{\Sigma} = \sum_{i=1}^N \omega_i^{-1} \boldsymbol{\eta}_i \mathbf{1}^T \boldsymbol{\varepsilon}_i = \sum_{i=1}^N \omega_i^{-1} \boldsymbol{\eta}_i \mathbf{1}^T (Y_i - \boldsymbol{\mu}_i) = \mathbf{0}$. Therefore, the estimating equations for GEE and QIF are simplified to the same form with exchangeable working correlation structure. Furthermore, the covariance estimator for QIF with exchangeable working correlation is

$$N^{-1} \left\{ \frac{\partial \bar{\mathbf{g}}_N(\hat{\boldsymbol{\beta}})}{\partial \boldsymbol{\beta}^T} \mathbf{C}_N^{-1}(\hat{\boldsymbol{\beta}}) \frac{\partial \bar{\mathbf{g}}_N(\hat{\boldsymbol{\beta}})}{\partial \boldsymbol{\beta}} \right\}^{-1} = m^{-2} \otimes (\boldsymbol{\Theta}^T \boldsymbol{\Delta}^{-1} \boldsymbol{\Theta})^{-1}$$

$$= m^{-2} \left(\sum_{i=1}^N \omega_i^{-1} \boldsymbol{\eta}_i \boldsymbol{\eta}_i^T \right)^{-1} \left(\sum_{i=1}^N \omega_i^{-2} \boldsymbol{\eta}_i \mathbf{1}^T \boldsymbol{\varepsilon}_i \boldsymbol{\varepsilon}_i^T \mathbf{1} \boldsymbol{\eta}_i^T \right) \left(\sum_{i=1}^N \omega_i^{-1} \boldsymbol{\eta}_i \boldsymbol{\eta}_i^T \right)^{-1},$$

which is the same as the ‘‘sandwich’’ covariance estimator $\widehat{\mathbf{V}}_{R,exc}$ of GEE with exchangeable working correlation structure. Therefore, GEE and QIF with exchangeable correlation structures share the same estimating equation form and covariance estimator, which further implies numerically equivalent estimators and covariance estimators.

Result 3

Based on [Result 1](#) and [Result 2](#), we need only to prove that QIF with independence and exchangeable working correlation structures have identical estimating equations and covariance estimators to fulfill [Result 3](#). For QIF with independence correlation structure, we have $\bar{\mathbf{g}}_N(\boldsymbol{\beta}) = N^{-1} \sum_{i=1}^N \mathbf{g}_i(\boldsymbol{\beta}) = N^{-1} \varphi \boldsymbol{\Sigma}$, $\mathbf{C}_N(\boldsymbol{\beta}) = N^{-1} \varphi^2 \boldsymbol{\Delta}$, and $\partial \bar{\mathbf{g}}_N(\boldsymbol{\beta}) / \partial \boldsymbol{\beta} = -m N^{-1} \varphi \boldsymbol{\Theta}$. Therefore, under independence correlation structure, we have the asymptotic estimating equation of QIF to be $\mathbf{W}_N(\boldsymbol{\beta}) = -m \boldsymbol{\Theta}^T \boldsymbol{\Delta}^{-1} \boldsymbol{\Sigma} = \mathbf{0}$, which also implies that $\boldsymbol{\Theta}^T \boldsymbol{\Delta}^{-1} \boldsymbol{\Sigma} = \mathbf{0}$ and is the same as the result in QIF with exchangeable correlation structure. The covariance estimator for QIF with independence working correlation is

$$N^{-1} \left\{ \frac{\partial \bar{\mathbf{g}}_N(\hat{\boldsymbol{\beta}})}{\partial \boldsymbol{\beta}^T} \mathbf{C}_N^{-1}(\hat{\boldsymbol{\beta}}) \frac{\partial \bar{\mathbf{g}}_N(\hat{\boldsymbol{\beta}})}{\partial \boldsymbol{\beta}} \right\}^{-1}$$

$$= m^{-2} \left(\sum_{i=1}^N \omega_i^{-1} \boldsymbol{\eta}_i \boldsymbol{\eta}_i^T \right)^{-1} \left(\sum_{i=1}^N \omega_i^{-2} \boldsymbol{\eta}_i \mathbf{1}^T \boldsymbol{\varepsilon}_i \boldsymbol{\varepsilon}_i^T \mathbf{1} \boldsymbol{\eta}_i^T \right) \left(\sum_{i=1}^N \omega_i^{-1} \boldsymbol{\eta}_i \boldsymbol{\eta}_i^T \right)^{-1}.$$

Thus, both the asymptotic estimating equations and covariance estimators for QIF are simplified to the same forms with independence and exchangeable correlation structures. Combining with [Result 1](#) and [Result 2](#), [Result 3](#) is proven.

APPENDIX II. Additional details of the data generating process

A potential difference between our data generating process (DGP) and previously published simulations (in the GEE or QIF literature) is that we have assumed that the correlation matrix is randomly sampled from some population distribution. Here, we show that this DGP is valid as it corresponds to an induced and fixed marginal correlation matrix. In other words, once we marginalize over the population distribution of the correlation matrix, our DGP follows a specific marginal mean and marginal correlation structure, but dispenses with the multivariate normality assumption. We will explain details of DGP under CS0, CS1, CS2 and CS3 below. Throughout, we let $\sigma^2 = \varphi$ be the individual-level variance, as we are simulating correlated continuous outcomes.

Fixed-regular exchangeable (CS0)

With a fixed off-diagonal value $\rho = 0.05$ in the correlation matrix, we have for each cluster i

$$Y_i | \mu_i, \rho, \sigma^2 \sim N[\mu_i, \sigma^2 \{(1 - \rho)\mathbf{I} + \rho\mathbf{J}\}].$$

For $\forall i \neq j$, we also have $\text{cov}(Y_i, Y_j | \mu_i, \mu_j, \rho, \sigma^2) = 0$. Therefore, under CS0, Y_i follows multivariate normal outcomes, and has a zero covariance matrix with Y_j for $j \neq i$.

Regular exchangeable (CS1)

In this scenario, we assume the common correlation ρ is sampled from a uniform distribution with lower and upper bounds ($l = 0.01, u = 0.2$). Denote $E(\rho) = \bar{\rho} = (l+u)/2$ and observe that conditional on a realized value of ρ , $Y_i | \mu_i, \rho, \sigma^2 \sim N[\mu_i, \sigma^2 \{(1 - \rho)\mathbf{I} + \rho\mathbf{J}\}]$. We can then marginalize over the sampling variability of ρ to get $E(Y_i | \mu_i, \bar{\rho}, \sigma^2) = E\{E(Y_i | \mu_i, \rho, \sigma^2) | \mu_i, \bar{\rho}, \sigma^2\} = \mu_i$ and

$$\begin{aligned} \text{cov}(Y_i | \mu_i, \bar{\rho}, \sigma^2) &= \text{cov}\{E(Y_i | \mu_i, \rho, \sigma^2) | \mu_i, \bar{\rho}, \sigma^2\} + E\{\text{cov}(Y_i | \mu_i, \rho, \sigma^2) | \mu_i, \bar{\rho}, \sigma^2\} \\ &= \sigma^2 \{(1 - \bar{\rho})\mathbf{I} + \bar{\rho}\mathbf{J}\}. \end{aligned}$$

For $\forall i \neq j$, we also have

$$\begin{aligned} \text{cov}(Y_i, Y_j | \mu_i, \mu_j, \bar{\rho}, \sigma^2) &= \text{cov}\{E(Y_i | \mu_i, \rho, \sigma^2), E(Y_j | \mu_j, \rho, \sigma^2) | \mu_i, \mu_j, \bar{\rho}, \sigma^2\} \\ &\quad + E\{\text{cov}(Y_i, Y_j | \mu_i, \mu_j, \rho, \sigma^2) | \mu_i, \mu_j, \bar{\rho}, \sigma^2\} = \mathbf{0}. \end{aligned}$$

Therefore, under CS1, Y_i and Y_j have a zero covariance matrix, and Y_i still follows a distribution with mean μ_i and exchangeable correlation $\sigma^2 \{(1 - \bar{\rho})\mathbf{I} + \bar{\rho}\mathbf{J}\}$, but is no longer multivariate normal (after accounting for the sampling variability of ρ).

Cluster-specific exchangeable (CS2)

Given that the cluster-specific correlation ρ_i is now sampled from a uniform distribution with lower and upper bounds ($l = 0.01, u = 0.2$) $\forall i$, we similarly have $E(\rho_i) = \bar{\rho} = (l+u)/2$ and $Y_i | \mu_i, \rho_i, \sigma^2 \sim N[\mu_i, \sigma^2 \{(1 - \rho_i)\mathbf{I} + \rho_i\mathbf{J}\}]$. We then marginalize over the variability of ρ_i to get $E(Y_i | \mu_i, \sigma^2) = E\{E(Y_i | \mu_i, \rho_i, \sigma^2) | \mu_i, \bar{\rho}, \sigma^2\} = \mu_i$ and

$$\begin{aligned} \text{cov}(Y_i | \mu_i, \bar{\rho}, \sigma^2) &= \text{cov}\{E(Y_i | \mu_i, \rho_i, \sigma^2) | \mu_i, \bar{\rho}, \sigma^2\} + E\{\text{cov}(Y_i | \mu_i, \rho_i, \sigma^2) | \mu_i, \bar{\rho}, \sigma^2\} \\ &= \sigma^2 \{(1 - \bar{\rho})\mathbf{I} + \bar{\rho}\mathbf{J}\}. \end{aligned}$$

For $\forall i \neq j$, we also have

$$\begin{aligned} \text{cov}(Y_i, Y_j | \mu_i, \mu_j, \bar{\rho}, \sigma^2) &= \text{cov}\{E(Y_i | \mu_i, \rho_i, \sigma^2), E(Y_j | \mu_j, \rho_j, \sigma^2) | \mu_i, \mu_j, \bar{\rho}, \sigma^2\} \\ &\quad + E\{\text{cov}(Y_i, Y_j | \mu_i, \mu_j, \rho_i, \rho_j, \sigma^2) | \mu_i, \mu_j, \bar{\rho}, \sigma^2\} = \mathbf{0}. \end{aligned}$$

Therefore, under CS2, we still have a zero covariance matrix for Y_i and Y_j , and Y_i follows a distribution with mean μ_i and exchangeable correlation $\sigma^2 \{(1 - \bar{\rho})\mathbf{I} + \bar{\rho}\mathbf{J}\}$, but again is no longer multivariate normal (after accounting for the sampling variability of ρ).

Cluster-specific exchangeable with sub-clustering (CS3)

Let $F_i = (F_{i1}, \dots, F_{im})^T$ with F_{ij} sampled from a discrete uniform distribution in $\{1, 2, 3, 4\}$, we have the correlation matrix $\mathbf{R}(F_i)$, which is uniquely determined by F_i . Then, the outcome Y_i has conditional distribution

$$Y_i | \mu_i, \mathbf{R}(F_i), \sigma^2 \sim N\{\mu_i, \sigma^2 \mathbf{R}(F_i)\}.$$

We now denote the population-level mean of $\mathbf{R}(F_i)$ is $\bar{\mathbf{R}} = E\{\mathbf{R}(F_i)\}$, which averages over the sampling variability of F_i . We then marginalize over the variability of F_i to get $E(Y_i | \mu_i, \bar{\mathbf{R}}, \sigma^2) = E\{E(Y_i | \mu_i, \mathbf{R}(F_i), \sigma^2) | \mu_i, \bar{\mathbf{R}}, \sigma^2\} = \mu_i$ and

$$\text{cov}(Y_i | \mu_i, \bar{\mathbf{R}}, \sigma^2) = \text{cov}\{E(Y_i | \mu_i, \mathbf{R}(F_i), \sigma^2) | \mu_i, \bar{\mathbf{R}}, \sigma^2\} + E\{\text{cov}(Y_i | \mu_i, \mathbf{R}(F_i), \sigma^2) | \mu_i, \bar{\mathbf{R}}, \sigma^2\} = \bar{\mathbf{R}}.$$

For $\forall i \neq j$, we also have

$$\begin{aligned} \text{cov}(Y_i, Y_j | \mu_i, \mu_j, \bar{\mathbf{R}}, \sigma^2) &= \text{cov}\{E(Y_i | \mu_i, \mathbf{R}(F_i), \sigma^2), E(Y_j | \mu_j, \mathbf{R}(F_j), \sigma^2) | \mu_i, \mu_j, \bar{\mathbf{R}}, \sigma^2\} \\ &\quad + E\{\text{cov}(Y_i, Y_j | \mu_i, \mu_j, \mathbf{R}(F_i), \mathbf{R}(F_j), \sigma^2) | \mu_i, \mu_j, \bar{\mathbf{R}}, \sigma^2\} = \mathbf{0}. \end{aligned}$$

Therefore, under CS3, we still have a zero covariance matrix for Y_i and Y_j , and Y_i follows a distribution with mean μ_i and marginal common correlation $\bar{\mathbf{R}}$, but again is no longer multivariate normal (after accounting for the sampling variability of F_i).

APPENDIX III. Bias-corrected covariance estimators

The Mancl-DeRouen and Kauermann-Carroll bias-correction methods for QIF [9] are introduced in the similar expressions as the corresponding bias-correction methods for GEE [27,28]. Let

$$\nabla \bar{g}_N(\hat{\beta}) = \partial \bar{g}_N(\hat{\beta}) / \partial \beta,$$

and

$$J_N = \{\nabla \bar{g}_N(\hat{\beta})\}^T C_N^{-1}(\hat{\beta}) \{\nabla \bar{g}_N(\hat{\beta})\},$$

we have the covariance estimator of QIF $\widehat{\text{cov}}(\hat{\beta}) = N^{-1} J_N^{-1}$. The Mancl-DeRouen and Kauermann-Carroll bias-corrected covariance estimators have the general form

$$\widehat{\text{cov}}_{\text{MD/KC}}(\hat{\beta}) = N^{-1} (\mathbf{I} + \mathbf{G}) J_N^{-1} \{\nabla \bar{g}_N(\hat{\beta})\}^T C_N^{-1}(\hat{\beta}) \tilde{C}_{N,\text{MD/KC}}(\hat{\beta}) C_N^{-1}(\hat{\beta}) \{\nabla \bar{g}_N(\hat{\beta})\} J_N^{-1} (\mathbf{I} + \mathbf{G})^T,$$

where

$$\mathbf{G} = - [J_N^{-1} \{\nabla \bar{g}_N(\hat{\beta})\}^T C_N^{-1}(\hat{\beta}) \bar{g}_N(\hat{\beta})] / \partial \beta.$$

Define

$$\mathbf{O}_i = N^{-1} \mathbf{D}_i (\mathbf{I} + \mathbf{G}) J_N^{-1} \{\nabla \bar{g}_N(\hat{\beta})\}^T C_N^{-1}(\hat{\beta}) \begin{Bmatrix} \mathbf{D}_i^T \mathbf{A}_i^{-1} \\ \mathbf{D}_i^T \mathbf{A}_i^{-1/2} \mathbf{M}_2 \mathbf{A}_i^{-1/2} \\ \vdots \\ \mathbf{D}_i^T \mathbf{A}_i^{-1/2} \mathbf{M}_K \mathbf{A}_i^{-1/2} \end{Bmatrix},$$

the Mancl-DeRouen bias-corrected covariance estimator has

$$\tilde{C}_{N,\text{MD}}(\hat{\beta}) = N^{-1} \sum_{i=1}^N \begin{Bmatrix} \mathbf{D}_i^T \mathbf{A}_i^{-1} \\ \mathbf{D}_i^T \mathbf{A}_i^{-1/2} \mathbf{M}_2 \mathbf{A}_i^{-1/2} \\ \vdots \\ \mathbf{D}_i^T \mathbf{A}_i^{-1/2} \mathbf{M}_K \mathbf{A}_i^{-1/2} \end{Bmatrix} (\mathbf{I} + \mathbf{O}_i)^{-1} (\mathbf{Y}_i - \boldsymbol{\mu}_i) (\mathbf{Y}_i - \boldsymbol{\mu}_i)^T (\mathbf{I} + \mathbf{O}_i)^{-1} \begin{Bmatrix} \mathbf{D}_i^T \mathbf{A}_i^{-1} \\ \mathbf{D}_i^T \mathbf{A}_i^{-1/2} \mathbf{M}_2 \mathbf{A}_i^{-1/2} \\ \vdots \\ \mathbf{D}_i^T \mathbf{A}_i^{-1/2} \mathbf{M}_K \mathbf{A}_i^{-1/2} \end{Bmatrix}^T,$$

while the Kauermann-Carroll bias-corrected covariance estimator has

$$\tilde{C}_{N,\text{KC}}(\hat{\beta}) = N^{-1} \sum_{i=1}^N \begin{Bmatrix} \mathbf{D}_i^T \mathbf{A}_i^{-1} \\ \mathbf{D}_i^T \mathbf{A}_i^{-1/2} \mathbf{M}_2 \mathbf{A}_i^{-1/2} \\ \vdots \\ \mathbf{D}_i^T \mathbf{A}_i^{-1/2} \mathbf{M}_K \mathbf{A}_i^{-1/2} \end{Bmatrix} (\mathbf{I} + \mathbf{O}_i)^{-1} (\mathbf{Y}_i - \boldsymbol{\mu}_i) (\mathbf{Y}_i - \boldsymbol{\mu}_i)^T \begin{Bmatrix} \mathbf{D}_i^T \mathbf{A}_i^{-1} \\ \mathbf{D}_i^T \mathbf{A}_i^{-1/2} \mathbf{M}_2 \mathbf{A}_i^{-1/2} \\ \vdots \\ \mathbf{D}_i^T \mathbf{A}_i^{-1/2} \mathbf{M}_K \mathbf{A}_i^{-1/2} \end{Bmatrix}^T.$$

APPENDIX IV. Table of abbreviations

We present the acronyms and abbreviations in [Table 7](#).

Table 7
Table of Abbreviations

Acronym	Full name
CRT	cluster randomized trial
ICC	intracluster correlation coefficient
GEE	generalized estimating equations
QIF	quadratic inference function
GMM	generalized method of moments
HALI	health and literacy intervention
IQR	interquartile range
SD	standard deviation
TGI	the godambe information criterion
DGP	data generating process
CS	correlation structure in the data generating process
MM	mean model in the data generating process
CS0	fixed-regular exchangeable correlation structure
CS1	regular exchangeable correlation structure
CS2	cluster-specific exchangeable correlation structure
CS3	cluster-specific exchangeable with sub-clustering correlation structure
MM1	intervention-arm only model
MM2	intervention-arm with 4 covariates model
RBS	relative bias
SE	standard error
ESE	empirical standard error

(continued on next page)

Table 7 (continued)

Acronym	Full name
MRSE	mean robust standard error
PR	power ratio
Q/G	quadratic inference function relative to generalized estimating equations
QIF _{MD}	quadratic inference function with the Mancl-DeRouen bias-correction method
QIF _{KC}	quadratic inference function with the Kauermann-Carroll bias-correction method
GEE _{ind}	generalized estimating equations using independence working correlation
GEE _{exc}	generalized estimating equations using exchangeable working correlation
QIF _{ind-MD}	quadratic inference function with Mancl-DeRouen bias correction and independence
QIF _{ind-KC}	quadratic inference function with Kauermann-Carroll bias correction and independence
QIF _{exc-MD}	quadratic inference function with Mancl-DeRouen bias correction and exchangeable
QIF _{exc-KC}	quadratic inference function with Kauermann-Carroll bias correction and exchangeable

References

- [1] Richard J. Hayes, Lawrence H. Moulton, Cluster Randomized Trials, CRC Press, 2009, pp. 3–40 (chapter 1-3).
- [2] Elizabeth L. Turner, Fan Li, John A. Gallis, Melanie Prague, David M. Murray, Review of recent methodological developments in group-randomized trials: part 1—design, *Am. J. Publ. Health* 107 (6) (2017) 907–915.
- [3] Elizabeth L. Turner, Melanie Prague, John A. Gallis, Fan Li, David M. Murray, Review of recent methodological developments in group-randomized trials: part 2—analysis, *Am. J. Publ. Health* 107 (7) (2017) 1078–1086.
- [4] Kung-Yee Liang, Scott L. Zeger, Longitudinal data analysis using generalized linear models, *Biometrika* 73 (1) (April 1986) 13–22.
- [5] John Ritz, Donna Spiegelman, Equivalence of conditional and marginal regression models for clustered and longitudinal data, *Stat. Methods Med. Res.* 13 (4) (2004) 309–323.
- [6] Alan E. Hubbard, Jennifer Ahern, Nancy L. Fleischer, Mark Van der Laan, Sheri A. Lippman, Nicholas Jewell, Tim Bruckner, William A. Satariano, To gee or not to gee: comparing population average and mixed models for estimating the associations between neighborhood risk factors and health, *Epidemiology* 21 (4) (July 2010) 467–474.
- [7] John S. Preisser, Mary L. Young, Daniel J. Zaccaro, Mark Wolfson, An integrated population-averaged approach to the design, analysis and sample size determination of cluster-unit trials, *Stat. Med.* 22 (8) (2003) 1235–1254.
- [8] Annie Qu, Peter X.-K. Song, Assessing robustness of generalised estimating equations and quadratic inference functions, *Biometrika* 91 (2) (June 2004) 447–459.
- [9] Philip M. Westgate, A bias-corrected covariance estimate for improved inference with quadratic inference function, *Stat. Med.* 31 (29) (December 2012) 4003–4022.
- [10] Fereshteh Asgari, Akbar Biglarian, Behjat Seifi, Andisheh Bakhshi, Hamid Heidarian Miri, Enayatollah Bakhshi, Using quadratic inference functions to determine the factors associated with obesity: findings from the steps survey in Iran, *Ann. Epidemiol.* 23 (9) (2013) 534–538.
- [11] Enayatollah Bakhshi, Koorosh Etemad, Behjat Seifi, Kazem Mohammad, Akbar Biglarian, Jalil Koohpayehzadeh, Changes in obesity odds ratio among iranian adults, since 2000: quadratic inference functions method, *Comput. Math. Method. Med.* 2016 (2016).
- [12] Kun Yang, Lixin Tao, Gehendra Mahara, Yan Yan, Kai Cao, Xiangtong Liu, Sipeng Chen, Qin Xu, Long Liu, Chao Wang, et al., An association of platelet indices with blood pressure in beijing adults: applying quadratic inference function for a longitudinal study, *Medicine* 95 (39) (2016).
- [13] Simon Brooker, George Okello, Kiambo Njagi, Margaret M. Dubeck, Katherine E. Halliday, Hellen Inyega, Matthew CH. Jukes, Improving educational achievement and anaemia of school children: design of a cluster randomised trial of school-based malaria prevention and enhanced literacy instruction in Kenya, *Trials* 11 (1) (2010) 93.
- [14] Katherine E. Halliday, George Okello, Elizabeth L. Turner, Kiambo Njagi, Carlos Mcharo, Juddy Kengo, Elizabeth Allen, Margaret M. Dubeck, Matthew CH. Jukes, Simon J. Brooker, Impact of intermittent screening and treatment for malaria among school children in Kenya: a cluster randomised trial, *PLoS Med.* 11 (1) (2014).
- [15] Matthew CH. Jukes, Elizabeth L. Turner, Margaret M. Dubeck, Katherine E. Halliday, Hellen N. Inyega, Sharon Wolf, Stephanie Simmons Zuilkowski, Simon J. Brooker, Improving literacy instruction in Kenya through teacher professional development and text messages support: a cluster randomized trial, *J. Res. Edu. Effect.* 10 (3) (2017) 449–481.
- [16] Robert WM. Wedderburn, Quasi-likelihood functions, generalized linear models, and the gauss-Newton method, *Biometrika* 61 (3) (December 1974) 439–447.
- [17] Wei Pan, Sample size and power calculations with correlated binary data, *Contr. Clin. Trials* 22 (3) (2001) 211–227.
- [18] Clare Rutterford, Andrew Copas, Sandra Eldridge, Methods for sample size determination in cluster randomized trials, *Int. J. Epidemiol.* 44 (3) (2015) 1051–1067.
- [19] Annie Qu, Bruce G. Lindsay, Bing Li, Improving generalised estimating equations using quadratic inference functions, *Biometrika* 87 (4) (December 2000) 823–836.
- [20] Lars Peter Hansen, Large sample properties of generalized method of moments estimators, *Econometrica* 50 (4) (July 1982) 1029–1054.
- [21] Lan Wang, Annie Qu, Consistent model selection and data-driven smooth tests for longitudinal data in the estimating equations approach, *J. Roy. Stat. Soc.* 71 (1) (January 2009) 177–190.
- [22] Peter X.-K. Song, Zhichang Jiang, Eunjo Park, Annie Qu, Quadratic inference functions in marginal models for longitudinal data, *Stat. Med.* 28 (29) (December 2009) 3683–3696.
- [23] Philip M. Westgate, Criterion for the simultaneous selection of a working correlation structure and either generalized estimating equations or the quadratic inference function approach, *Biom. J.* 56 (3) (May 2014) 461–476.
- [24] Fei Wang, Lu Wang, Peter X.-K. Song, Quadratic inference function approach to merging longitudinal studies: validation and joint estimation, *Biometrika* 99 (3) (September 2012) 755–762.
- [25] Philip M. Westgate, Thomas M. Braun, The effect of cluster size imbalance and covariates on the estimation performance of quadratic inference functions, *Stat. Med.* 31 (20) (September 2012) 2209–2222.
- [26] Fan Li, Andrew B. Forbes, Elizabeth L. Turner, John S. Preisser, Power and sample size requirements for gee analyses of cluster randomized crossover trials, *Stat. Med.* 38 (4) (2019) 636–649.
- [27] Lloyd A. Mancl, T.A. DeRouen, A covariance estimator for GEE with improved small-sample properties, *Biometrics* 57 (1) (2001) 126–134.
- [28] Göran Kauermann, R.J. Carroll, A note on the efficiency of sandwich covariance matrix estimation, *J. Am. Stat. Assoc.* 96 (456) (2001) 1387–1396.
- [29] Philip M. Westgate, Thomas M. Braun, An improved quadratic inference function for parameter estimation in the analysis of correlated data, *Stat. Med.* 32 (19) (August 2013) 3260–3273.
- [30] Philip M. Westgate, A comparison of utilized and theoretical covariance weighting matrices on the estimation performance of quadratic inference functions, *Commun. Stat. Simulat. Comput.* 43 (10) (2014) 2432–2443.
- [31] Thomas M. Braun, Ziding Feng, Optimal permutation tests for the analysis of group randomized trials, *J. Am. Stat. Assoc.* 96 (456) (dec 2001) 1424–1432.
- [32] Fan Li, Elizabeth L. Turner, Patrick J. Heagerty, David M. Murray, William M. Vollmer, Elizabeth R. DeLong, An evaluation of constrained randomization for the design and analysis of group-randomized trials with binary outcomes, *Stat. Med.* 36 (24) (2017) 3791–3806.