Supplementary material for manuscript

For manuscript: Properties and pitfalls of weighting as an alternative to multilevel multiple imputation in cluster randomized trials with missing binary outcomes under covariate-dependent missingness

Elizabeth L. Turner [1,2*], Lanqiu Yao[3], Fan Li[1], and Melanie Prague[4,5]

1. Department of Biostatistics and Bioinformatics, Duke University, Durham, NC
2. Duke Global Health Institute, Duke University, Durham, NC, USA
3. Department of Population Health, New York University
4. Inria, project team SISTM, Bordeaux, France
5. Department of Biostatistics, Harvard T.H. Chan School of Public Health, Boston, MA, USA

1 Definitions of missingness mechanisms

Here we outline missing data definitions in order to link to the broader missing data literature. Specifically, we refer to Chapter 1 of the book by Molenburghs et al. (2015) [1], where the authors remind us that missingness mechanism is defined with respect to the true underlying data-generating models. Like us, in Section 1.2 of Chapter 1, the authors focus on missing outcomes assuming that all covariates are fully observed.

Using their notation: $\boldsymbol{Y}_i = (Y_{i1}, Y_{i2}, \ldots, Y_{in_i})$ is an $n_i \times 1$ vector of responses (e.g., in our case, for n_i individuals in the ith cluster) and $\boldsymbol{X_i}$ is an $n_i \times p$ matrix of covariates. The response vector $\boldsymbol{Y}_i$ can be partitioned in to the portion that is observed ($\boldsymbol{Y}_i^o$) and the portion that is not observed ($\boldsymbol{Y}_i^m$), i.e. that is missing. We, like them, use the indicator $R_{ij} = 1$ to indicate that a response is observed and a value of 0 to indicate that it is missing.

Molenburghs et al. [1], like many authors, use the term MCAR (missing completely at random) for situations where:

$$\Pr(\boldsymbol{R_i}|\boldsymbol{Y_i^o}, \boldsymbol{Y_i^m}, \boldsymbol{X}_i) = \Pr(\boldsymbol{R_i}), \text{ Definition MCAR - see [1]}$$

that is where the probability of being observed (or being missing) is independent of all data, including covariates. They use the term MAR (missing at random) for situations where the probability of being observed (or missing) may depend on the observed outcome data and/or covariates:

$$\Pr(\boldsymbol{R_i}|\boldsymbol{Y_i^o}, \boldsymbol{Y_i^m}, \boldsymbol{X}_i) = \Pr(\boldsymbol{R_i}|\boldsymbol{Y_i^o}, \boldsymbol{X}_i), \text{ Definition MAR - see [1]}$$

They also point out that some other authors use the term MCAR to include situations where the probability of being observed (or being missing) may depend on covariates i.e. so that:

$$\Pr(\boldsymbol{R_i}|\boldsymbol{Y_i^o}, \boldsymbol{Y_i^m}, \boldsymbol{X}_i) = \Pr(\boldsymbol{R_i}|\boldsymbol{X}_i), \text{ Alternative definition MCAR - see [1]}$$

The authors point out that Little (1995) [2] suggested that this setting be called covariate-dependent missingness.

Importantly, this is the setting that we consider in our manuscript, specifically in a trial setting with fully observed $\boldsymbol{X_i}$, where those covariates, $\boldsymbol{X_i}$, are baseline covariates. The fact that there is disagreement as to whether covariate-dependent missingness with fully observed covariates is a type of MCAR or MAR mechanism is not relevant to the results presented in the current manuscript. What is important is that these definitions do not depend on the analysis models that are chosen but instead are defined by the true, underlying mechanisms. What is also important, and which is noted by Molenberghs et al. [1], is that a subtle issue arises for the CDM case, namely that the full set of covariates that are predictive of $\boldsymbol{R_i}$ must be accounted for in the analysis using the correct functional form. We also note that were there to be missingness in the baseline covariates that are predictive of missingness outcomes, this setting would be a type of MNAR.

2 Supplemental materials for analysis of motivating data set

Table S1: Results of analysis of high literacy outcome from HALI motivating data set using five different GEE methods with robust standard errors under two different working correlation matrices (exchangeable and independent).

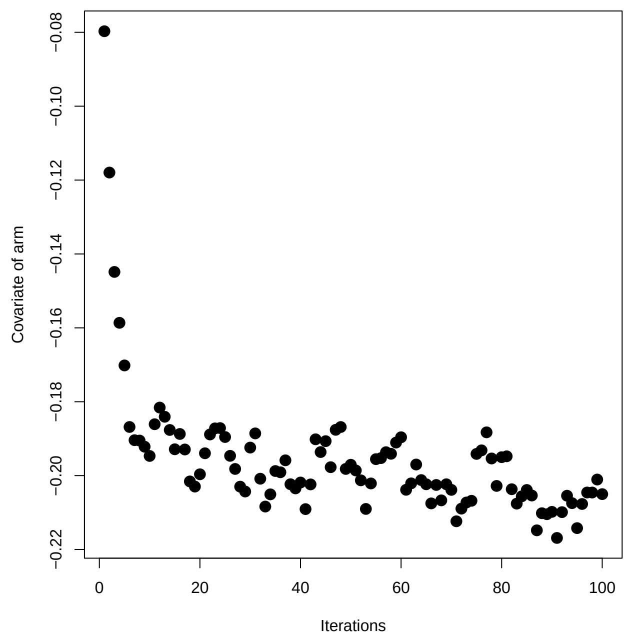
	Exchangeable			Independent		
Method	Odds ratio	95% confidence interval	p-value	Odds ratio	95% confidence interval	p-value
CRA-GEE	1.82	(1.26, 2.65)	0.002	1.84	(1.27, 2.66)	0.001
A-CRA-GEE	1.93	(1.38, 2.71)	<0.001	1.92	(1.37, 2.68)	<0.001
W-GEE	1.80	(1.23, 2.62)	0.002	1.81	(1.24, 2.63)	0.002
CW-GEE	1.83	(1.26, 2.65)	0.001	1.83	(1.26, 2.65)	0.002
MMI-GEE	1.81	(1.25, 2.63)	0.002	1.79	(1.23, 2.60)	0.002

Table S2: Predictors of outcome (high literacy) and of missingness (of high literacy outcome) from HALI motivating data set.

	Predictors of outcome (high literacy)			Predictors of outcome missingness		
	Estimate[a]	SE	p-value	Estimate[a]	SE	p-value
Intercept	-3.61	0.38	<0.001	-1.51	0.39	<0.001
Intervention	0.75	0.19	<0.001	0.09	0.15	0.538
Baseline literacy score	0.35	0.02	<0.001	-0.03	0.01	0.105
Age	0.06	0.04	0.094	-0.02	0.04	0.544
Sex (Female)	-0.04	0.11	0.733	-0.00	0.13	0.991
Household head education[b]			0.002			0.021
Primary	0.13	0.13		-0.34	0.15	
Secondary	0.61	0.22		-0.17	0.24	
College/degree	1.16	0.38		0.46	0.32	
Household SES[c]			0.357			0.920
Poor	-0.04	0.17		0.04	0.19	
Median Poor	-0.14	0.17		-0.02	0.19	
Less Poor	-0.08	0.18		-0.10	0.21	
Least Poor	0.24	0.20		-0.15	0.22	

[a]: Log-odds ratio; [b]: With reference level of "Primary school not completed"; [c]: Socioeconomic status (SES) derived via principal components analysis of household assets with reference level of "Poorest".

Figure S1: MCMC traceplot for MMI of the high literacy outcome from the HALI data set



Footnote: This figure was generated using the jomo1rancat function in the jomo package for burn-in of 100 iterations, thinning rate of 25 and for 15 imputed data sets. The five covariates of Table S2 (i.e. baseline literacy score, age, sex, household head education level and household SES) and treatment arm were included as additive terms in the random effects logistic imputation model.

3 Supplemental materials for GEE finite-sample correction

Using the notation used in the main manuscript text, the uncorrected sandwich variance estimator is given by (ignoring the variability of weights)

$$\sum_{i=1}^{M} \hat{\boldsymbol{\Omega}} \left(\boldsymbol{D}_i^T \boldsymbol{V}_i^{-1} \boldsymbol{W}_i \boldsymbol{r}_i \boldsymbol{r}_i^T \boldsymbol{W}_i \boldsymbol{V}_i^{-1} \boldsymbol{D}_i \right) \hat{\boldsymbol{\Omega}},$$

where $\boldsymbol{W}_i$ is the diagonal matrix of weights, $\hat{\boldsymbol{\Omega}} = (\sum_{i=1}^{M} \boldsymbol{D}_i^T \boldsymbol{V}_i^{-1} \boldsymbol{W}_i \boldsymbol{D}_i)^{-1}$ is the model-based variance with weighting, and $\boldsymbol{r}_i = \boldsymbol{Y}_i - \hat{\boldsymbol{\mu}}_i$ is the residual vector for cluster i. With a limited number of clusters, the residual vector, $\boldsymbol{r}_i$, tends to be biased towards zero, and the uncorrected sandwich variance is likely to underestimate the true variability. Define the cluster leverage as $\boldsymbol{H}_i = \boldsymbol{D}_i \hat{\boldsymbol{\Omega}} \boldsymbol{D}_i^T \boldsymbol{V}_i^{-1} \boldsymbol{W}_i$. Following the bias-corrected variance due to Kauermann and Carroll [3], the corresponding sandwich variance with weighting is

$$\hat{\boldsymbol{\Omega}} \Big(\sum_{i=1}^{M} \boldsymbol{D}_i^T \boldsymbol{V}_i^{-1} \boldsymbol{W}_i (\boldsymbol{I} - \boldsymbol{H}_i)^{-1/2} \boldsymbol{r}_i \boldsymbol{r}_i^T (\boldsymbol{I} - \boldsymbol{H}_i^T)^{-1/2} \boldsymbol{W}_i \boldsymbol{V}_i^{-1} \boldsymbol{D}_i \Big) \hat{\boldsymbol{\Omega}}.$$

Following Mancl and DeRouen [4], a similar bias-corrected sandwich variance is given by

$$\hat{\boldsymbol{\Omega}} \Big(\sum_{i=1}^{M} \boldsymbol{D}_i^T \boldsymbol{V}_i^{-1} \boldsymbol{W}_i (\boldsymbol{I} - \boldsymbol{H}_i)^{-1} \boldsymbol{r}_i \boldsymbol{r}_i^T (\boldsymbol{I} - \boldsymbol{H}_i^T)^{-1} \boldsymbol{W}_i \boldsymbol{V}_i^{-1} \boldsymbol{D}_i \Big) \hat{\boldsymbol{\Omega}}.$$

Notice that both approaches estimate $\text{cov}(\boldsymbol{Y}_i)$ based on the leverage-adjusted residuals, correcting for the finite-sample bias in the raw residuals in a multiplicative fashion. We also considered the finite-sample correction by Fay and Graubard [5], given by

$$\hat{\boldsymbol{\Omega}} \Big(\sum_{i=1}^{M} \boldsymbol{F}_i \boldsymbol{D}_i^T \boldsymbol{V}_i^{-1} \boldsymbol{W}_i \boldsymbol{r}_i \boldsymbol{r}_i^T \boldsymbol{W}_i \boldsymbol{V}_i^{-1} \boldsymbol{D}_i \boldsymbol{F}_i \Big) \hat{\boldsymbol{\Omega}},$$

where $\boldsymbol{F}_i = \text{diag}\{(1 - \min\{0.75, [\boldsymbol{Q}_i]_{jj}\})^{-1/2}\}$ and $\boldsymbol{Q}_i = \boldsymbol{D}_i^T \boldsymbol{V}_i^{-1} \boldsymbol{W}_i \boldsymbol{D}_i \hat{\boldsymbol{\Omega}}$. In brief, the multiplicative correlation factor $\boldsymbol{F}_i$ is motivated by expanding the estimating function around the truth and assuming the working variance is approximately proportional to the true variance.

4 Supplemental materials for simulation study reported in the Results section

The following pages contain all supplemental tables and figures referred to in the text of the main manuscript. More specifically, they contain more comprehensive results of the simulation study with analyses using both exchangeable and independent working correlation matrix. For simulation study details and methods, see the main manuscript text.

Table S3: Simulation study results summary with exchangeable working correlation matrix

			Mean relative bias (%)					Coverage (%)					Mean SE					MCSD				
k	ρ_O	ρ_M	CRA-GEE	A-CRA-GEE	W-GEE	CW-GEE	MMI-GEE	CRA-GEE	A-CRA-GEE	W-GEE	CW-GEE	MMI-GEE	CRA-GEE	A-CRA-GEE	W-GEE	CW-GEE	MMI-GEE	CRA-GEE	A-CRA-GEE	W-GEE	CW-GEE	MMI-GEE
10	0.05	0	-2.4	3.1	0.0	0.0	0.7	89.9	83.4	91.4	91.6	97.0	0.268	0.217	0.276	0.276	0.294	0.290	0.297	0.298	0.297	0.294
		0.1	-2.3	2.9	-0.1	0.7	0.3	92.1	83.3	92.8	89.9	96.7	0.266	0.214	0.274	0.280	0.294	0.283	0.293	0.290	0.310	0.287
		0.3	-1.4	3.2	0.4	1.8	0.5	90.7	82.7	91.5	83.9	96.2	0.268	0.216	0.274	0.297	0.301	0.292	0.303	0.297	0.378	0.297
		0.5	-1.1	3.0	0.1	2.8	0.3	91.8	81.9	91.8	76.5	97.0	0.269	0.216	0.274	0.315	0.309	0.290	0.300	0.295	0.446	0.296
	0.1	0	-2.6	2.7	-0.2	-0.2	0.5	91.5	75.6	91.4	91.6	95.8	0.313	0.212	0.320	0.320	0.333	0.342	0.352	0.346	0.346	0.348
		0.1	-2.5	2.5	-0.5	0.2	0.3	91.1	76.2	91.2	89.3	96.1	0.313	0.211	0.320	0.327	0.334	0.336	0.349	0.344	0.365	0.337
		0.3	-1.5	2.9	0.0	1.6	0.6	91.1	74.8	91.1	85.4	95.8	0.316	0.212	0.321	0.350	0.341	0.342	0.355	0.348	0.443	0.347
		0.5	-1.1	2.9	-0.1	3.1	0.3	91.8	73.8	92.1	78.2	96.9	0.318	0.212	0.322	0.371	0.348	0.345	0.357	0.350	0.528	0.348
	0.2	0	-2.7	1.8	-0.4	-0.4	0.5	91.0	63.1	91.5	91.4	94.2	0.390	0.203	0.394	0.394	0.402	0.428	0.442	0.429	0.430	0.435
		0.1	-2.6	2.0	-0.6	0.3	0.3	90.8	63.5	91.9	89.8	94.5	0.395	0.205	0.399	0.408	0.408	0.432	0.448	0.436	0.461	0.435
		0.3	-1.7	2.3	-0.3	2.5	0.4	91.6	63.3	91.6	85.1	94.5	0.399	0.205	0.402	0.438	0.414	0.437	0.454	0.440	0.558	0.439
		0.5	-1.5	2.4	-0.6	4.6	0.1	90.8	62.4	90.7	79.9	94.9	0.402	0.205	0.404	0.464	0.421	0.439	0.457	0.443	0.667	0.442
25	0.05	0	-2.7	2.9	-0.3	-0.3	0.6	93.2	85.4	93.3	93.2	96.6	0.174	0.141	0.180	0.180	0.185	0.178	0.183	0.186	0.186	0.180
		0.1	-2.7	2.5	-0.7	-0.8	0.2	94.0	86.0	94.3	93.8	96.5	0.174	0.140	0.179	0.189	0.187	0.176	0.181	0.181	0.194	0.180
		0.3	-2.0	2.6	-0.6	-0.1	0.0	93.5	85.1	93.8	91.1	95.4	0.177	0.141	0.180	0.217	0.190	0.182	0.187	0.186	0.242	0.185
		0.5	-1.5	2.6	-0.6	0.6	-0.0	92.7	84.1	93.8	87.5	95.9	0.178	0.141	0.180	0.245	0.193	0.185	0.189	0.187	0.295	0.189
	0.1	0	-2.8	2.4	-0.6	-0.6	0.4	93.3	75.6	93.6	93.7	95.4	0.205	0.137	0.210	0.210	0.213	0.212	0.219	0.219	0.219	0.214
		0.1	-2.6	2.3	-0.7	-0.8	0.2	94.1	78.8	94.0	94.4	96.1	0.206	0.138	0.210	0.220	0.214	0.210	0.216	0.214	0.227	0.213
		0.3	-2.1	2.4	-0.6	-0.1	0.1	93.2	77.5	93.6	90.9	95.5	0.208	0.139	0.211	0.254	0.218	0.216	0.221	0.218	0.281	0.218
		0.5	-1.5	2.4	-0.6	1.0	0.1	92.8	75.8	93.6	88.3	95.2	0.211	0.139	0.213	0.288	0.222	0.219	0.224	0.220	0.347	0.221
	0.2	0	-3.0	1.5	-0.9	-0.9	0.1	94.0	65.1	94.3	94.0	95.5	0.259	0.131	0.261	0.261	0.265	0.264	0.275	0.267	0.267	0.265
		0.1	-2.8	1.5	-1.1	-1.3	0.0	93.4	62.2	93.8	93.6	94.8	0.259	0.131	0.260	0.273	0.265	0.271	0.278	0.272	0.287	0.276
		0.3	-2.4	1.5	-1.2	-0.6	-0.2	92.5	65.1	93.0	91.5	94.4	0.262	0.133	0.263	0.315	0.268	0.275	0.281	0.275	0.350	0.278
		0.5	-1.8	1.6	-0.9	1.1	-0.1	92.4	63.6	93.1	89.2	94.8	0.264	0.133	0.265	0.361	0.271	0.280	0.284	0.279	0.433	0.285
50	0.05	0	-2.5	2.7	-0.4	-0.4	0.6	93.7	87.4	95.1	95.1	95.0	0.125	0.100	0.129	0.129	0.131	0.124	0.126	0.127	0.127	0.126
		0.1	-2.4	2.6	-0.5	-0.3	0.3	93.5	85.2	94.6	93.7	95.5	0.126	0.100	0.129	0.137	0.131	0.130	0.132	0.133	0.144	0.131
		0.3	-1.8	2.7	-0.3	0.2	0.4	93.0	83.0	94.3	91.9	95.9	0.127	0.100	0.129	0.164	0.133	0.133	0.136	0.136	0.180	0.135
		0.5	-1.1	2.8	-0.1	0.7	0.5	93.3	83.2	93.6	90.8	95.1	0.128	0.101	0.129	0.192	0.135	0.134	0.137	0.136	0.219	0.135
	0.1	0	-2.6	2.4	-0.7	-0.7	0.6	93.9	80.1	95.1	95.0	95.6	0.147	0.098	0.150	0.150	0.151	0.145	0.148	0.147	0.147	0.148
		0.1	-2.5	2.3	-0.7	-0.5	0.3	93.9	79.0	94.1	94.4	95.4	0.148	0.098	0.150	0.158	0.152	0.150	0.154	0.153	0.164	0.152
		0.3	-1.9	2.4	-0.6	0.1	0.4	93.9	77.2	94.4	92.7	96.4	0.149	0.098	0.151	0.190	0.153	0.154	0.158	0.156	0.204	0.156
		0.5	-1.3	2.5	-0.4	0.8	0.4	94.2	75.8	95.2	90.9	94.9	0.151	0.098	0.152	0.225	0.156	0.155	0.159	0.156	0.251	0.156
	0.2	0	-2.7	1.6	-0.9	-0.9	0.5	94.6	66.8	94.8	94.7	95.8	0.185	0.094	0.186	0.186	0.188	0.181	0.186	0.180	0.180	0.185
		0.1	-2.5	1.7	-0.8	-0.8	0.3	94.2	65.5	94.9	94.0	95.1	0.186	0.093	0.187	0.197	0.189	0.187	0.193	0.189	0.202	0.191
		0.3	-2.1	1.8	-0.9	0.1	0.3	94.6	63.8	94.6	93.0	94.8	0.188	0.094	0.188	0.235	0.191	0.191	0.197	0.192	0.255	0.194
		0.5	-1.4	1.9	-0.6	1.0	0.4	94.1	63.0	94.2	91.7	95.7	0.190	0.094	0.190	0.281	0.194	0.194	0.200	0.194	0.318	0.196

Notes: k= # clusters per arm; ρ_O is the ICC of the (conditional) outcome model (see Eqn. 6 in the manuscript text); ρ_M is the ICC of the probability of missing (POM) model (see Eqn. 7B in the manuscript text); MCSD = Monte Carlo standard deviation; Results based on 1000 simulated data sets per scenario using standard Wald Z-based confidence intervals for each modeling approach, except for MMI-GEE for which t-based confidence intervals are used based on 15 imputations (see Section 4.2 of main manuscript for details, including the t-distribution degrees of freedom). Acceptable coverage ranges from 93.6% to 96.4%; Note that all GEE models converged except for 5 W-GEE and approximately 100 for CW-GEE across all simulations, with almost all issues in the small sample case (i.e. k=10). Simulated data sets for which there was non-convergence were replaced to ensure a total of 1000 simulations per scenario.

Table S4: Coverage (%) with exchangeable working correlation matrix and robust standard errors, uncorrected and with three finite-sample corrections

			No finite-sample correction					KC correction					MD correction					FG correction				
k	ρ_O	ρ_M	CRA-GEE	A-CRA-GEE	W-GEE	CW-GEE	MMI-GEE	CRA-GEE	A-CRA-GEE	W-GEE	CW-GEE	MMI-GEE	CRA-GEE	A-CRA-GEE	W-GEE	CW-GEE	MMI-GEE	CRA-GEE	A-CRA-GEE	W-GEE	CW-GEE	MMI-GEE
10	0.05	0	89.9	83.4	91.4	91.6	97.0	92.0	84.6	92.8	92.9	97.7	93.8	86.1	95.0	94.9	98.2	92.4	84.6	98.6	98.6	97.9
		0.1	92.1	83.3	92.8	89.9	96.7	93.4	84.7	93.8	92.5	97.6	94.7	86.2	95.0	94.0	98.1	93.5	84.7	99.0	99.0	97.6
		0.3	90.7	82.7	91.5	83.9	96.2	92.8	84.2	92.9	87.7	96.9	93.8	85.4	93.9	91.5	97.4	93.1	84.2	97.8	96.4	97.0
		0.5	91.8	81.9	91.8	76.5	97.0	93.1	84.1	93.1	83.8	97.7	94.6	85.0	94.4	88.3	98.3	93.5	84.1	98.2	92.6	97.9
	0.1	0	91.5	75.6	91.4	91.6	95.8	93.0	77.5	93.3	93.2	96.3	94.3	79.3	94.6	94.7	96.9	93.8	77.4	98.5	98.5	96.6
		0.1	91.1	76.2	91.2	89.3	96.1	92.6	77.7	93.3	92.2	97.0	94.0	79.0	94.6	94.7	97.5	93.2	77.6	98.6	99.2	97.1
		0.3	91.1	74.8	91.1	85.4	95.8	92.3	76.2	92.7	89.1	96.4	94.2	78.1	94.0	92.7	97.1	92.7	76.2	98.1	98.3	96.5
		0.5	91.8	73.8	92.1	78.2	96.9	93.2	75.4	93.2	84.6	97.6	94.6	77.5	94.7	90.7	97.8	93.3	75.4	98.3	95.9	97.6
	0.2	0	91.0	63.1	91.5	91.4	94.2	92.5	64.4	93.5	93.6	95.2	93.8	65.9	94.4	94.4	96.2	93.0	64.4	98.4	98.4	95.5
		0.1	90.8	63.5	91.9	89.8	94.5	93.4	64.8	93.2	91.9	95.7	94.3	65.9	94.7	93.9	96.8	94.0	64.7	98.0	98.7	95.8
		0.3	91.6	63.3	91.6	85.1	94.5	92.8	65.1	93.1	89.7	95.5	94.0	66.5	94.3	91.9	96.4	93.2	65.1	97.9	98.0	96.2
		0.5	90.8	62.4	90.7	79.9	94.9	92.1	64.0	92.8	86.1	96.2	93.6	65.6	94.2	91.1	97.0	92.6	64.0	97.8	97.8	96.4
25	0.05	0	93.2	85.4	93.3	93.2	96.6	93.7	86.0	93.7	93.5	97.1	94.2	86.5	94.4	94.2	97.2	93.8	86.0	95.9	96.0	97.1
		0.1	94.0	86.0	94.3	93.8	96.5	94.4	86.7	95.0	93.9	96.7	95.0	87.5	95.8	94.5	97.2	94.5	86.7	96.9	98.5	96.8
		0.3	93.5	85.1	93.8	91.1	95.4	94.0	85.5	94.1	92.9	95.8	94.5	85.8	94.5	93.6	96.2	94.3	85.4	95.8	99.4	95.8
		0.5	92.7	84.1	93.8	87.5	95.9	93.3	84.5	94.4	89.5	96.0	94.0	85.2	94.6	92.0	96.4	93.3	84.5	95.7	99.1	96.1
	0.1	0	93.3	75.6	93.6	93.7	95.4	94.1	76.4	94.1	94.2	95.8	94.3	77.2	94.6	94.6	96.1	94.2	76.4	96.4	96.5	95.9
		0.1	94.1	78.8	94.0	94.4	96.1	94.5	79.6	94.7	94.8	96.5	94.9	79.9	95.3	95.6	97.0	94.7	79.6	96.9	99.0	96.6
		0.3	93.2	77.5	93.6	90.9	95.5	94.3	78.2	94.4	92.5	95.6	94.5	78.5	95.0	94.1	96.1	94.3	78.2	96.5	99.6	95.9
		0.5	92.8	75.8	93.6	88.3	95.2	93.8	76.5	93.8	90.8	95.9	94.4	77.2	94.4	93.4	96.1	94.2	76.5	96.0	99.4	96.0
	0.2	0	94.0	65.1	94.3	94.0	95.5	94.9	65.8	94.9	94.7	95.5	95.2	66.5	95.2	95.2	96.0	95.0	65.8	96.7	96.9	95.6
		0.1	93.4	62.2	93.8	93.6	94.8	94.0	63.3	94.6	94.4	95.1	94.3	63.7	95.2	94.9	95.5	94.2	63.3	96.3	98.6	95.3
		0.3	92.5	65.1	93.0	91.5	94.4	93.2	65.6	93.4	92.7	94.7	93.8	66.3	93.9	94.2	95.3	93.3	65.6	96.2	99.2	95.1
		0.5	92.4	63.6	93.1	89.2	94.8	93.1	63.8	93.5	91.4	95.0	93.4	64.4	93.9	93.2	95.5	93.2	63.8	95.2	99.4	95.2
50	0.05	0	93.7	87.4	95.1	95.1	95.0	94.0	87.7	95.3	95.3	95.3	94.1	87.8	95.6	95.5	95.6	94.0	87.7	96.4	96.4	95.3
		0.1	93.5	85.2	94.6	93.7	95.5	93.7	85.3	94.6	93.8	95.9	94.2	85.4	94.8	94.1	95.9	93.7	85.3	95.6	97.2	95.9
		0.3	93.0	83.0	94.3	91.9	95.9	93.5	83.8	94.6	93.1	96.3	93.5	84.1	94.7	93.8	96.4	93.5	83.8	95.2	99.6	96.3
		0.5	93.3	83.2	93.6	90.8	95.1	93.6	83.9	93.8	91.8	95.5	93.9	84.4	94.1	92.7	95.5	93.6	83.9	94.7	100.0	95.5
	0.1	0	93.9	80.1	95.1	95.0	95.6	94.1	80.1	95.2	95.3	95.9	94.3	80.3	95.5	95.4	96.1	94.2	80.1	95.9	95.9	96.0
		0.1	93.9	79.0	94.1	94.4	95.4	94.3	79.1	94.4	94.5	95.4	94.5	79.5	94.4	95.0	95.4	94.3	79.1	95.5	97.3	95.4
		0.3	93.9	77.2	94.4	92.7	96.4	94.6	77.3	94.7	93.4	96.4	94.8	77.8	94.9	94.1	96.7	94.7	77.3	95.9	99.6	96.4
		0.5	94.2	75.8	95.2	90.9	94.9	94.5	75.9	95.5	92.3	95.2	95.0	76.2	95.7	93.4	95.5	94.7	75.9	96.0	100.0	95.3
	0.2	0	94.6	66.8	94.8	94.7	95.8	94.6	67.1	95.3	95.4	96.1	94.8	67.7	95.5	95.6	96.2	94.6	67.1	96.0	96.0	96.1
		0.1	94.2	65.5	94.9	94.0	95.1	94.8	65.7	95.1	94.0	95.0	94.8	66.1	95.4	94.3	95.1	94.8	65.7	95.6	96.9	95.1
		0.3	94.6	63.8	94.6	93.0	94.8	94.9	64.0	94.8	93.6	95.0	95.1	64.4	95.0	94.0	95.2	95.0	64.0	95.5	99.5	95.0
		0.5	94.1	63.0	94.2	91.7	95.7	94.3	63.3	94.4	92.8	95.8	94.5	63.5	94.9	94.1	96.0	94.3	63.3	95.2	99.7	95.9

Notes: k= # clusters per arm; ρ_O is the ICC of the (conditional) outcome model (see Eqn. 6 in the manuscript text); ρ_M is the ICC of the probability of missing (POM) model (see Eqn. 7B in the manuscript text; Results based on 1000 simulated data sets per scenario using standard Wald Z-based confidence intervals for each modeling approach, except for MMI-GEE for which t-based confidence intervals are used based on 15 imputations (see Section 4.2 of main manuscript for details, including the t-distribution degrees of freedom). Acceptable coverage ranges from 93.6% to 96.4%; Note that the finite-sample corrections are due to: Kauermann G & Carroll R (2001, JASA 96(456):1387-1396); Mancl LA & DeRouen TA (2001, Biometrics 57(1):126-134); and Fay & Graubard (2001, Biometrics 57(4):1198-1206), with all details described in Section 2 above.

Table S5: Mean (robust) SE using exchangeable working correlation matrix, uncorrected and with three finite-sample corrections

			No finite-sample correction					KC correction					MD correction					FG correction				
k	ρ_O	ρ_M	CRA-GEE	A-CRA-GEE	W-GEE	CW-GEE	MMI-GEE	CRA-GEE	A-CRA-GEE	W-GEE	CW-GEE	MMI-GEE	CRA-GEE	A-CRA-GEE	W-GEE	CW-GEE	MMI-GEE	CRA-GEE	A-CRA-GEE	W-GEE	CW-GEE	MMI-GEE
10	0.05	0	0.268	0.217	0.276	0.276	0.294	0.282	0.224	0.292	0.292	0.307	0.298	0.232	0.308	0.308	0.321	0.286	0.224	0.401	0.403	0.311
		0.1	0.266	0.214	0.274	0.280	0.294	0.281	0.222	0.289	0.301	0.307	0.297	0.230	0.306	0.324	0.321	0.285	0.221	0.394	0.482	0.311
		0.3	0.268	0.216	0.274	0.297	0.301	0.283	0.224	0.290	0.339	0.314	0.300	0.232	0.306	0.393	0.327	0.287	0.223	0.393	0.589	0.317
		0.5	0.269	0.216	0.274	0.315	0.309	0.286	0.224	0.290	0.382	0.321	0.303	0.233	0.308	0.478	0.335	0.290	0.224	0.403	0.664	0.325
	0.1	0	0.313	0.212	0.320	0.320	0.333	0.330	0.219	0.338	0.338	0.348	0.348	0.227	0.356	0.357	0.364	0.335	0.219	0.471	0.473	0.352
		0.1	0.313	0.211	0.320	0.327	0.334	0.331	0.218	0.337	0.350	0.349	0.349	0.226	0.356	0.376	0.366	0.335	0.218	0.463	0.561	0.353
		0.3	0.316	0.212	0.321	0.350	0.341	0.334	0.220	0.339	0.396	0.356	0.353	0.228	0.358	0.454	0.372	0.338	0.220	0.461	0.690	0.360
		0.5	0.318	0.212	0.322	0.371	0.348	0.337	0.220	0.341	0.446	0.363	0.357	0.229	0.361	0.553	0.379	0.342	0.220	0.472	0.781	0.367
	0.2	0	0.390	0.203	0.394	0.394	0.402	0.411	0.210	0.415	0.415	0.421	0.434	0.218	0.438	0.438	0.442	0.417	0.210	0.586	0.588	0.427
		0.1	0.395	0.205	0.399	0.408	0.408	0.417	0.212	0.420	0.436	0.427	0.440	0.219	0.443	0.466	0.449	0.423	0.212	0.583	0.699	0.433
		0.3	0.399	0.205	0.402	0.438	0.414	0.421	0.213	0.424	0.492	0.433	0.444	0.221	0.447	0.559	0.454	0.427	0.213	0.580	0.863	0.439
		0.5	0.402	0.205	0.404	0.464	0.421	0.424	0.212	0.427	0.554	0.440	0.448	0.221	0.451	0.678	0.461	0.431	0.212	0.592	0.977	0.446
25	0.05	0	0.174	0.141	0.180	0.180	0.185	0.178	0.143	0.184	0.184	0.188	0.182	0.145	0.188	0.188	0.191	0.179	0.143	0.202	0.202	0.189
		0.1	0.174	0.140	0.179	0.189	0.187	0.178	0.142	0.183	0.194	0.190	0.182	0.144	0.187	0.200	0.193	0.179	0.142	0.200	0.265	0.190
		0.3	0.177	0.141	0.180	0.217	0.190	0.180	0.143	0.184	0.230	0.193	0.184	0.145	0.188	0.244	0.196	0.181	0.143	0.200	0.412	0.194
		0.5	0.178	0.141	0.180	0.245	0.193	0.182	0.144	0.184	0.268	0.196	0.186	0.146	0.188	0.295	0.199	0.183	0.144	0.201	0.512	0.197
	0.1	0	0.205	0.137	0.210	0.210	0.213	0.210	0.139	0.214	0.214	0.216	0.214	0.141	0.218	0.218	0.220	0.211	0.139	0.235	0.236	0.217
		0.1	0.206	0.138	0.210	0.220	0.214	0.211	0.140	0.214	0.226	0.217	0.215	0.141	0.219	0.233	0.221	0.212	0.139	0.235	0.308	0.218
		0.3	0.208	0.139	0.211	0.254	0.218	0.213	0.141	0.216	0.268	0.221	0.217	0.143	0.220	0.284	0.225	0.214	0.141	0.235	0.479	0.222
		0.5	0.211	0.139	0.213	0.288	0.222	0.215	0.141	0.217	0.314	0.225	0.220	0.143	0.222	0.344	0.228	0.217	0.141	0.237	0.601	0.226
	0.2	0	0.259	0.131	0.261	0.261	0.265	0.264	0.133	0.266	0.266	0.268	0.270	0.135	0.272	0.272	0.274	0.266	0.133	0.294	0.294	0.270
		0.1	0.259	0.131	0.260	0.273	0.265	0.264	0.133	0.266	0.280	0.268	0.270	0.135	0.271	0.288	0.273	0.266	0.133	0.292	0.382	0.270
		0.3	0.262	0.133	0.263	0.315	0.268	0.267	0.134	0.268	0.332	0.272	0.273	0.136	0.274	0.351	0.277	0.269	0.134	0.293	0.596	0.273
		0.5	0.264	0.133	0.265	0.361	0.271	0.270	0.135	0.270	0.392	0.275	0.275	0.137	0.276	0.428	0.280	0.271	0.135	0.295	0.754	0.276
50	0.05	0	0.125	0.100	0.129	0.129	0.131	0.126	0.101	0.130	0.130	0.132	0.128	0.102	0.132	0.132	0.133	0.127	0.101	0.136	0.136	0.132
		0.1	0.126	0.100	0.129	0.137	0.131	0.127	0.101	0.130	0.139	0.132	0.128	0.101	0.132	0.141	0.133	0.127	0.101	0.136	0.164	0.132
		0.3	0.127	0.100	0.129	0.164	0.133	0.128	0.101	0.130	0.169	0.134	0.129	0.102	0.132	0.174	0.135	0.128	0.101	0.135	0.295	0.134
		0.5	0.128	0.101	0.129	0.192	0.135	0.129	0.101	0.131	0.202	0.136	0.131	0.102	0.132	0.211	0.137	0.130	0.101	0.136	0.396	0.136
	0.1	0	0.147	0.098	0.150	0.150	0.151	0.149	0.099	0.152	0.152	0.152	0.150	0.099	0.153	0.153	0.154	0.149	0.099	0.158	0.158	0.153
		0.1	0.148	0.098	0.150	0.158	0.152	0.149	0.098	0.152	0.161	0.153	0.151	0.099	0.153	0.163	0.154	0.150	0.098	0.158	0.189	0.153
		0.3	0.149	0.098	0.151	0.190	0.153	0.151	0.099	0.152	0.195	0.154	0.152	0.099	0.154	0.201	0.156	0.151	0.099	0.158	0.340	0.155
		0.5	0.151	0.098	0.152	0.225	0.156	0.152	0.099	0.154	0.236	0.157	0.154	0.100	0.155	0.247	0.158	0.153	0.099	0.159	0.463	0.157
	0.2	0	0.185	0.094	0.186	0.186	0.188	0.187	0.095	0.187	0.187	0.189	0.188	0.095	0.189	0.189	0.191	0.187	0.095	0.196	0.196	0.189
		0.1	0.186	0.093	0.187	0.197	0.189	0.188	0.094	0.189	0.199	0.190	0.190	0.095	0.191	0.202	0.192	0.188	0.094	0.197	0.233	0.191
		0.3	0.188	0.094	0.188	0.235	0.191	0.190	0.094	0.190	0.242	0.192	0.192	0.095	0.192	0.249	0.194	0.190	0.094	0.198	0.420	0.192
		0.5	0.190	0.094	0.190	0.281	0.194	0.192	0.095	0.192	0.293	0.194	0.194	0.095	0.194	0.307	0.196	0.193	0.095	0.200	0.579	0.195

Notes: k= # clusters per arm; ρ_O is the ICC of the (conditional) outcome model (see Eqn. 6 in the manuscript text); ρ_M is the ICC of the probability of missing (POM) model (see Eqn. 7B in the manuscript text); Results based on 1000 simulated data sets per scenario; Note that the finite-sample corrections are due to: Kauermann G & Carroll R (2001, JASA 96(456):1387-1396); Mancl LA & DeRouen TA (2001, Biometrics 57(1):126-134); and Fay & Graubard (2001, Biometrics 57(4):1198-1206), with all details described in Section 2 above.

Table S6: Simulation study results summary with independent working correlation matrix

			Mean relative bias (%)					Coverage (%)					Mean SE					MCSD				
k	ρ_O	ρ_M	CRA-GEE	A-CRA-GEE	W-GEE	CW-GEE	MMI-GEE	CRA-GEE	A-CRA-GEE	W-GEE	CW-GEE	MMI-GEE	CRA-GEE	A-CRA-GEE	W-GEE	CW-GEE	MMI-GEE	CRA-GEE	A-CRA-GEE	W-GEE	CW-GEE	MMI-GEE
10	0.05	0	-2.4	3.2	0.0	0.0	0.6	89.8	82.6	91.5	91.5	96.9	0.268	0.218	0.276	0.276	0.294	0.291	0.298	0.299	0.299	0.294
		0.1	-2.2	3.0	0.0	0.7	0.3	91.8	84.4	92.8	89.8	96.8	0.267	0.215	0.274	0.283	0.295	0.283	0.291	0.289	0.316	0.287
		0.3	-1.3	3.2	0.4	1.8	0.5	91.0	83.1	91.6	85.7	96.3	0.268	0.217	0.274	0.301	0.301	0.292	0.301	0.296	0.388	0.297
		0.5	-1.1	3.0	0.1	2.8	0.3	91.7	82.8	91.5	79.2	96.8	0.270	0.217	0.274	0.314	0.309	0.291	0.299	0.295	0.456	0.296
	0.1	0	-2.6	2.9	-0.3	-0.3	0.5	91.0	76.1	91.3	91.1	95.7	0.314	0.215	0.320	0.321	0.333	0.344	0.353	0.348	0.348	0.348
		0.1	-2.3	2.8	-0.2	0.5	0.3	90.6	75.5	91.5	88.7	96.4	0.315	0.214	0.320	0.333	0.335	0.337	0.345	0.343	0.376	0.337
		0.3	-1.2	3.2	0.4	2.1	0.6	91.4	74.8	91.3	83.9	95.8	0.318	0.216	0.322	0.357	0.342	0.345	0.354	0.348	0.465	0.347
		0.5	-0.9	3.0	0.2	3.5	0.3	91.4	73.8	91.5	78.4	96.9	0.321	0.216	0.324	0.373	0.349	0.349	0.358	0.352	0.548	0.349
	0.2	0	-2.7	2.5	-0.4	-0.5	0.5	91.3	64.8	91.6	91.7	94.2	0.392	0.211	0.395	0.395	0.402	0.431	0.441	0.432	0.433	0.435
		0.1	-2.1	2.8	-0.2	0.8	0.3	90.4	64.7	90.9	88.5	94.6	0.398	0.213	0.401	0.417	0.408	0.436	0.447	0.438	0.482	0.435
		0.3	-1.1	3.1	0.4	3.4	0.4	90.3	64.2	91.1	84.1	94.7	0.403	0.214	0.405	0.451	0.414	0.448	0.459	0.446	0.599	0.440
		0.5	-0.7	3.1	0.2	5.8	0.1	90.2	62.5	90.6	77.5	94.8	0.408	0.214	0.409	0.473	0.421	0.455	0.466	0.454	0.705	0.442
25	0.05	0	-2.7	2.9	-0.3	-0.3	0.6	93.4	85.5	93.7	93.8	96.8	0.175	0.142	0.180	0.180	0.185	0.178	0.182	0.186	0.186	0.179
		0.1	-2.6	2.5	-0.7	-0.8	0.2	93.9	85.9	94.4	93.2	96.3	0.175	0.141	0.180	0.193	0.187	0.177	0.182	0.181	0.200	0.180
		0.3	-1.9	2.6	-0.4	-0.0	0.1	93.9	85.3	94.0	90.5	95.2	0.177	0.142	0.181	0.224	0.190	0.183	0.187	0.186	0.252	0.185
		0.5	-1.3	2.7	-0.4	0.8	0.0	92.9	85.3	92.9	87.2	95.8	0.179	0.142	0.181	0.250	0.193	0.186	0.189	0.187	0.305	0.189
	0.1	0	-2.8	2.7	-0.6	-0.6	0.5	93.2	76.3	93.6	93.7	95.3	0.206	0.140	0.210	0.210	0.213	0.212	0.216	0.218	0.218	0.214
		0.1	-2.6	2.5	-0.7	-0.8	0.2	94.1	78.9	94.8	93.8	96.0	0.208	0.140	0.211	0.227	0.214	0.212	0.217	0.215	0.238	0.214
		0.3	-1.9	2.6	-0.4	0.2	0.1	93.5	78.5	93.7	89.8	95.5	0.211	0.141	0.213	0.265	0.218	0.218	0.223	0.220	0.302	0.218
		0.5	-1.3	2.6	-0.3	1.3	0.2	93.5	76.7	94.1	87.1	95.3	0.214	0.142	0.215	0.297	0.222	0.222	0.226	0.222	0.367	0.221
	0.2	0	-3.0	2.2	-1.0	-1.0	0.3	94.5	66.6	94.3	94.2	95.5	0.260	0.138	0.262	0.262	0.265	0.264	0.269	0.267	0.268	0.265
		0.1	-2.8	2.1	-1.1	-1.2	0.2	93.4	64.6	93.6	93.1	94.6	0.263	0.137	0.263	0.285	0.265	0.275	0.282	0.275	0.304	0.276
		0.3	-2.1	2.1	-0.8	-0.2	-0.0	92.6	65.4	92.6	91.0	94.4	0.267	0.139	0.267	0.335	0.268	0.281	0.288	0.280	0.383	0.278
		0.5	-1.4	2.3	-0.5	1.6	0.0	93.1	63.7	92.7	87.7	94.5	0.271	0.139	0.271	0.376	0.271	0.287	0.293	0.285	0.466	0.285
50	0.05	0	-2.6	2.8	-0.4	-0.4	0.6	93.4	87.2	94.9	94.9	95.1	0.125	0.101	0.129	0.129	0.131	0.125	0.127	0.128	0.128	0.126
		0.1	-2.3	2.7	-0.3	-0.2	0.4	93.5	84.4	94.4	93.0	95.3	0.126	0.101	0.129	0.141	0.131	0.131	0.134	0.135	0.150	0.131
		0.3	-1.6	2.8	-0.1	0.3	0.5	93.0	83.2	94.4	91.5	95.8	0.127	0.101	0.129	0.171	0.133	0.135	0.137	0.137	0.190	0.135
		0.5	-0.9	3.0	0.1	0.9	0.5	93.5	83.7	94.1	90.1	95.5	0.129	0.101	0.130	0.198	0.135	0.136	0.138	0.138	0.227	0.135
	0.1	0	-2.7	2.6	-0.7	-0.7	0.5	93.8	80.7	94.4	94.5	95.8	0.148	0.100	0.151	0.151	0.151	0.147	0.150	0.149	0.149	0.148
		0.1	-2.4	2.6	-0.5	-0.4	0.3	93.5	78.6	93.7	94.2	95.3	0.149	0.099	0.151	0.165	0.152	0.153	0.156	0.156	0.173	0.152
		0.3	-1.6	2.7	-0.2	0.4	0.4	93.2	77.2	95.0	91.7	96.4	0.151	0.100	0.152	0.202	0.154	0.157	0.161	0.159	0.221	0.156
		0.5	-1.0	2.8	-0.1	1.1	0.4	93.4	76.6	94.5	90.6	94.9	0.153	0.100	0.154	0.235	0.156	0.159	0.162	0.160	0.266	0.156
	0.2	0	-2.8	2.2	-0.9	-0.9	0.4	94.3	69.8	94.7	94.7	96.0	0.186	0.098	0.187	0.187	0.188	0.184	0.187	0.183	0.183	0.185
		0.1	-2.4	2.3	-0.7	-0.6	0.4	93.4	66.3	94.4	93.6	95.0	0.188	0.098	0.189	0.207	0.189	0.193	0.197	0.193	0.216	0.191
		0.3	-1.7	2.4	-0.4	0.6	0.3	94.2	63.5	94.8	92.2	94.9	0.192	0.098	0.192	0.254	0.191	0.198	0.203	0.198	0.281	0.194
		0.5	-1.1	2.5	-0.2	1.4	0.4	94.2	63.1	94.1	90.5	95.7	0.196	0.099	0.195	0.297	0.194	0.203	0.207	0.202	0.342	0.196

Notes: k= # clusters per arm; ρ_O is the ICC of the (conditional) outcome model (see Eqn. 6 in the manuscript text); ρ_M is the ICC of the probability of missing (POM) model (see Eqn. 7B in the manuscript text); MCSD = Monte Carlo standard deviation; Results based on 1000 simulated data sets per scenario using standard Wald Z-based confidence intervals for each modeling approach, except for MMI-GEE for which t-based confidence intervals are used based on 15 imputations (see Section 4.2 of main manuscript for details, including the t-distribution degrees of freedom). Acceptable coverage ranges from 93.6% to 96.4%; Note that all GEE models converged except for 5 W-GEE and approximately 100 for CW-GEE across all simulations, with almost all issues in the small sample case (i.e. k=10). Simulated data sets for which there was non-convergence were replaced to ensure a total of 1000 simulations per scenario.

Table S7: Coverage (%) with independent working correlation matrix and robust standard errors, uncorrected and with three finite-sample corrections

			No finite-sample correction					KC correction					MD correction					FG correction				
k	ρ_O	ρ_M	CRA-GEE	A-CRA-GEE	W-GEE	CW-GEE	MMI-GEE	CRA-GEE	A-CRA-GEE	W-GEE	CW-GEE	MMI-GEE	CRA-GEE	A-CRA-GEE	W-GEE	CW-GEE	MMI-GEE	CRA-GEE	A-CRA-GEE	W-GEE	CW-GEE	MMI-GEE
10	0.05	0	89.8	82.6	91.5	91.5	96.9	91.8	84.2	92.6	92.7	97.8	93.6	86.0	94.6	94.7	98.4	92.0	84.2	98.6	98.6	98.1
		0.1	91.8	84.4	92.8	89.8	96.8	93.1	85.4	93.9	92.5	97.5	94.5	87.2	95.1	94.5	98.0	93.4	85.5	99.2	99.6	97.6
		0.3	91.0	83.1	91.6	85.7	96.3	92.2	84.0	92.8	89.9	96.8	93.7	86.0	94.3	93.3	97.6	92.8	83.9	97.8	98.6	97.0
		0.5	91.7	82.8	91.5	79.2	96.8	93.3	84.3	93.0	86.9	97.7	94.3	85.7	94.5	91.7	98.2	93.6	84.2	98.4	97.2	97.9
	0.1	0	91.0	76.1	91.3	91.1	95.7	93.3	78.1	93.4	93.5	96.0	94.4	79.8	94.5	94.5	96.9	93.7	78.0	98.5	98.5	96.4
		0.1	90.6	75.5	91.5	88.7	96.4	92.4	76.7	92.6	91.7	97.1	94.4	78.7	94.9	95.0	97.6	93.0	76.5	98.8	99.3	97.3
		0.3	91.4	74.8	91.3	83.9	95.8	92.7	76.6	93.1	89.1	96.4	93.9	78.8	94.4	92.7	97.2	93.0	76.4	98.4	98.5	96.6
		0.5	91.4	73.8	91.5	78.4	96.9	92.9	76.1	92.9	84.9	97.4	94.2	78.0	94.9	91.7	97.8	93.4	75.9	98.5	96.9	97.5
	0.2	0	91.3	64.8	91.6	91.7	94.2	92.9	67.3	93.2	93.3	95.1	94.0	69.1	94.0	94.0	96.2	93.0	67.1	98.7	98.8	95.6
		0.1	90.4	64.7	90.9	88.5	94.6	92.8	66.5	92.7	91.3	95.8	94.1	68.0	95.0	93.9	96.6	93.2	66.5	98.5	98.8	96.1
		0.3	90.3	64.2	91.1	84.1	94.7	92.4	65.6	92.7	89.1	95.4	94.1	67.0	94.3	91.8	96.6	92.8	65.5	98.1	98.3	95.7
		0.5	90.2	62.5	90.6	77.5	94.8	92.0	64.3	92.3	85.1	96.2	94.0	65.7	94.3	91.0	97.1	92.8	64.1	98.5	97.6	96.5
25	0.05	0	93.4	85.5	93.7	93.8	96.8	94.0	86.3	94.4	94.4	96.9	94.5	87.1	95.1	95.0	97.4	94.2	86.3	96.4	96.5	97.0
		0.1	93.9	85.9	94.4	93.2	96.3	94.0	86.6	94.8	94.0	96.6	95.0	86.9	95.7	94.3	96.9	94.1	86.6	96.9	98.5	96.9
		0.3	93.9	85.3	94.0	90.5	95.2	94.3	86.0	94.3	92.1	95.5	94.5	86.6	94.8	93.8	96.1	94.4	85.9	95.9	99.4	95.7
		0.5	92.9	85.3	92.9	87.2	95.8	93.5	85.7	93.6	89.7	96.0	93.9	86.6	94.1	92.6	96.5	93.5	85.7	95.9	99.4	96.2
	0.1	0	93.2	76.3	93.6	93.7	95.3	93.6	77.1	94.3	94.4	96.1	94.3	77.9	95.1	95.2	96.3	93.6	77.1	96.4	96.4	96.2
		0.1	94.1	78.9	94.8	93.8	96.0	94.6	79.5	95.2	94.7	96.2	94.8	80.1	95.3	95.0	96.6	94.7	79.5	97.1	99.1	96.5
		0.3	93.5	78.5	93.7	89.8	95.5	94.0	79.2	94.4	92.2	95.7	94.6	79.9	94.9	93.6	96.2	94.0	79.2	96.4	99.6	95.7
		0.5	93.5	76.7	94.1	87.1	95.3	93.9	77.4	94.7	91.0	95.8	94.3	77.9	95.3	92.9	96.2	94.0	77.4	96.3	99.3	96.1
	0.2	0	94.5	66.6	94.3	94.2	95.5	95.0	67.3	94.8	94.9	95.8	95.5	67.9	95.0	95.0	96.1	95.0	67.2	97.1	97.1	95.9
		0.1	93.4	64.6	93.6	93.1	94.6	94.2	65.4	94.3	93.8	95.1	94.8	66.1	95.1	94.7	95.4	94.4	65.4	96.4	98.4	95.2
		0.3	92.6	65.4	92.6	91.0	94.4	93.2	66.5	93.8	92.9	94.9	93.9	67.1	94.2	94.2	95.3	93.3	66.3	95.7	99.4	95.0
		0.5	93.1	63.7	92.7	87.7	94.5	93.5	64.1	93.1	90.7	95.1	93.9	64.7	94.0	93.0	95.5	93.5	63.9	96.1	99.3	95.2
50	0.05	0	93.4	87.2	94.9	94.9	95.1	93.6	87.7	95.3	95.2	95.2	93.9	87.8	95.5	95.5	95.2	93.7	87.6	96.3	96.3	95.2
		0.1	93.5	84.4	94.4	93.0	95.3	94.0	84.7	94.8	93.8	95.4	94.4	85.1	95.2	94.1	95.6	94.0	84.7	95.9	97.5	95.5
		0.3	93.0	83.2	94.4	91.5	95.8	93.1	83.6	94.8	92.6	96.0	93.4	83.8	95.0	93.3	96.4	93.1	83.5	95.7	99.6	96.0
		0.5	93.5	83.7	94.1	90.1	95.5	94.1	83.9	94.3	91.6	95.6	94.3	84.0	94.4	93.0	95.9	94.3	83.9	94.7	99.9	95.7
	0.1	0	93.8	80.7	94.4	94.5	95.8	94.0	81.3	94.8	94.8	96.1	94.3	81.6	95.2	95.2	96.3	94.0	81.3	95.8	95.8	96.1
		0.1	93.5	78.6	93.7	94.2	95.3	93.8	78.7	93.8	94.4	95.3	93.9	79.0	94.6	94.6	95.4	93.8	78.7	95.8	97.5	95.3
		0.3	93.2	77.2	95.0	91.7	96.4	93.4	77.4	95.2	92.5	96.6	93.9	77.6	95.5	93.6	96.7	93.6	77.4	96.0	99.7	96.7
		0.5	93.4	76.6	94.5	90.6	94.9	93.7	76.8	94.9	91.9	95.2	94.1	77.2	95.0	93.3	95.4	93.8	76.8	95.3	99.9	95.3
	0.2	0	94.3	69.8	94.7	94.7	96.0	94.7	70.0	94.9	95.0	96.3	94.7	70.5	95.2	95.2	96.3	94.7	69.8	96.3	96.2	96.3
		0.1	93.4	66.3	94.4	93.6	95.0	93.8	66.6	94.5	94.1	95.1	94.0	66.8	94.6	94.6	95.4	93.9	66.5	95.3	97.7	95.2
		0.3	94.2	63.5	94.8	92.2	94.9	94.4	63.8	95.0	93.1	95.2	94.7	64.0	95.2	94.4	95.4	94.5	63.8	95.8	99.7	95.3
		0.5	94.2	63.1	94.1	90.5	95.7	94.6	63.5	94.3	92.2	95.8	95.0	64.3	94.5	94.0	96.1	94.6	63.5	95.1	99.7	95.9

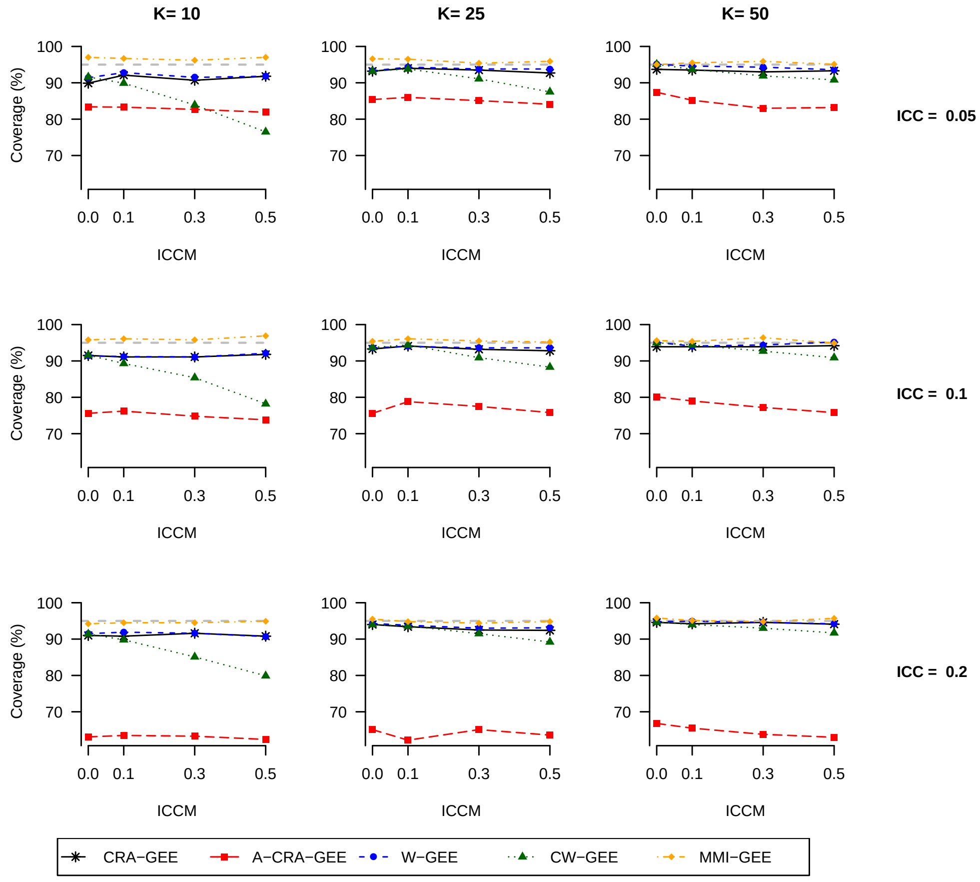
Notes: k= # clusters per arm; ρ_O is the ICC of the (conditional) outcome model (see Eqn. 6 in the manuscript text); ρ_M is the ICC of the probability of missing (POM) model (see Eqn. 7B in the manuscript text; Results based on 1000 simulated data sets per scenario using standard Wald Z-based confidence intervals for each modeling approach, except for MMI-GEE for which t-based confidence intervals are used based on 15 imputations (see Section 4.2 of main manuscript for details, including the t-distribution degrees of freedom). Acceptable coverage ranges from 93.6% to 96.4%;. Note that the finite-sample corrections are due to: Kauermann G & Carroll R (2001, JASA 96(456):1387-1396); Mancl LA & DeRouen TA (2001, Biometrics 57(1):126-134); and Fay & Graubard (2001, Biometrics 57(4):1198-1206), with all details described in Section 2 above.

Table S8: Mean (robust) SE using independent working correlation matrix, uncorrected and with three finite-sample corrections

			No finite-sample correction					KC correction					MD correction					FG correction				
k	ρ_O	ρ_M	CRA-GEE	A-CRA-GEE	W-GEE	CW-GEE	MMI-GEE	CRA-GEE	A-CRA-GEE	W-GEE	CW-GEE	MMI-GEE	CRA-GEE	A-CRA-GEE	W-GEE	CW-GEE	MMI-GEE	CRA-GEE	A-CRA-GEE	W-GEE	CW-GEE	MMI-GEE
10	0.05	0	0.268	0.218	0.276	0.276	0.294	0.283	0.226	0.292	0.292	0.308	0.299	0.235	0.308	0.309	0.322	0.286	0.226	0.403	0.406	0.311
		0.1	0.267	0.215	0.274	0.283	0.295	0.282	0.223	0.290	0.305	0.308	0.298	0.232	0.307	0.331	0.322	0.286	0.223	0.399	0.494	0.312
		0.3	0.268	0.217	0.274	0.301	0.301	0.285	0.225	0.290	0.346	0.314	0.303	0.234	0.309	0.404	0.329	0.289	0.225	0.401	0.598	0.318
		0.5	0.270	0.217	0.274	0.314	0.309	0.288	0.226	0.292	0.383	0.322	0.308	0.235	0.312	0.485	0.336	0.293	0.225	0.415	0.662	0.326
	0.1	0	0.314	0.215	0.320	0.321	0.333	0.331	0.223	0.339	0.339	0.349	0.350	0.232	0.358	0.358	0.366	0.335	0.223	0.475	0.478	0.353
		0.1	0.315	0.214	0.320	0.333	0.335	0.333	0.222	0.339	0.359	0.350	0.353	0.231	0.359	0.389	0.367	0.337	0.222	0.473	0.586	0.355
		0.3	0.318	0.216	0.322	0.357	0.342	0.338	0.224	0.342	0.410	0.357	0.359	0.233	0.363	0.480	0.374	0.342	0.223	0.478	0.712	0.362
		0.5	0.321	0.216	0.324	0.373	0.349	0.343	0.225	0.346	0.456	0.364	0.366	0.234	0.369	0.578	0.381	0.348	0.224	0.496	0.789	0.369
	0.2	0	0.392	0.211	0.395	0.395	0.402	0.414	0.220	0.417	0.418	0.423	0.437	0.229	0.441	0.442	0.445	0.419	0.218	0.594	0.597	0.429
		0.1	0.398	0.213	0.401	0.417	0.408	0.421	0.222	0.424	0.451	0.429	0.446	0.231	0.449	0.488	0.451	0.427	0.220	0.599	0.739	0.435
		0.3	0.403	0.214	0.405	0.451	0.414	0.429	0.224	0.430	0.518	0.435	0.456	0.234	0.458	0.607	0.457	0.435	0.222	0.608	0.904	0.441
		0.5	0.408	0.214	0.409	0.473	0.421	0.436	0.223	0.436	0.577	0.442	0.466	0.234	0.466	0.730	0.463	0.443	0.222	0.633	1.003	0.448
25	0.05	0	0.175	0.142	0.180	0.180	0.185	0.178	0.144	0.184	0.184	0.189	0.182	0.146	0.188	0.188	0.192	0.179	0.143	0.203	0.203	0.190
		0.1	0.175	0.141	0.180	0.193	0.187	0.179	0.143	0.184	0.199	0.190	0.183	0.145	0.188	0.206	0.193	0.180	0.143	0.202	0.282	0.191
		0.3	0.177	0.142	0.181	0.224	0.190	0.182	0.144	0.185	0.239	0.193	0.186	0.146	0.189	0.256	0.197	0.183	0.144	0.203	0.431	0.194
		0.5	0.179	0.142	0.181	0.250	0.193	0.183	0.145	0.185	0.275	0.196	0.188	0.147	0.190	0.305	0.200	0.184	0.144	0.204	0.523	0.197
	0.1	0	0.206	0.140	0.210	0.210	0.213	0.210	0.142	0.215	0.215	0.217	0.215	0.144	0.219	0.220	0.221	0.211	0.142	0.237	0.238	0.218
		0.1	0.208	0.140	0.211	0.227	0.214	0.213	0.142	0.216	0.235	0.218	0.217	0.144	0.221	0.243	0.222	0.214	0.142	0.238	0.336	0.219
		0.3	0.211	0.141	0.213	0.265	0.218	0.216	0.143	0.218	0.283	0.222	0.221	0.146	0.223	0.304	0.226	0.217	0.143	0.240	0.512	0.223
		0.5	0.214	0.142	0.215	0.297	0.222	0.219	0.144	0.220	0.327	0.226	0.225	0.146	0.226	0.362	0.230	0.221	0.144	0.243	0.623	0.227
	0.2	0	0.260	0.138	0.262	0.262	0.265	0.266	0.140	0.268	0.268	0.270	0.272	0.142	0.274	0.274	0.275	0.267	0.139	0.297	0.297	0.271
		0.1	0.263	0.137	0.263	0.285	0.265	0.269	0.140	0.269	0.294	0.270	0.275	0.142	0.275	0.304	0.275	0.270	0.139	0.298	0.424	0.272
		0.3	0.267	0.139	0.267	0.335	0.268	0.274	0.141	0.273	0.357	0.273	0.280	0.144	0.280	0.383	0.279	0.275	0.141	0.302	0.649	0.275
		0.5	0.271	0.139	0.271	0.376	0.271	0.278	0.142	0.278	0.414	0.276	0.286	0.144	0.285	0.458	0.282	0.280	0.141	0.308	0.791	0.278
50	0.05	0	0.125	0.101	0.129	0.129	0.131	0.127	0.102	0.131	0.131	0.132	0.128	0.103	0.132	0.132	0.133	0.127	0.102	0.136	0.136	0.132
		0.1	0.126	0.101	0.129	0.141	0.131	0.127	0.101	0.131	0.143	0.132	0.129	0.102	0.132	0.145	0.133	0.128	0.101	0.136	0.175	0.133
		0.3	0.127	0.101	0.129	0.171	0.133	0.129	0.102	0.131	0.177	0.134	0.130	0.102	0.132	0.184	0.135	0.129	0.102	0.136	0.315	0.134
		0.5	0.129	0.101	0.130	0.198	0.135	0.130	0.102	0.132	0.209	0.137	0.132	0.103	0.133	0.220	0.138	0.131	0.102	0.137	0.410	0.137
	0.1	0	0.148	0.100	0.151	0.151	0.151	0.149	0.101	0.152	0.152	0.153	0.151	0.101	0.154	0.154	0.154	0.150	0.101	0.159	0.159	0.153
		0.1	0.149	0.099	0.151	0.165	0.152	0.150	0.100	0.153	0.168	0.153	0.152	0.101	0.154	0.170	0.155	0.151	0.100	0.159	0.207	0.154
		0.3	0.151	0.100	0.152	0.202	0.154	0.153	0.101	0.154	0.209	0.155	0.154	0.101	0.156	0.217	0.156	0.153	0.101	0.160	0.373	0.155
		0.5	0.153	0.100	0.154	0.235	0.156	0.155	0.101	0.156	0.248	0.158	0.157	0.102	0.158	0.261	0.159	0.156	0.101	0.162	0.487	0.158
	0.2	0	0.186	0.098	0.187	0.187	0.188	0.188	0.099	0.189	0.189	0.190	0.190	0.100	0.191	0.191	0.192	0.189	0.099	0.198	0.198	0.191
		0.1	0.188	0.098	0.189	0.207	0.189	0.191	0.099	0.191	0.210	0.191	0.193	0.100	0.193	0.214	0.193	0.191	0.099	0.200	0.261	0.192
		0.3	0.192	0.098	0.192	0.254	0.191	0.194	0.099	0.194	0.263	0.193	0.197	0.100	0.196	0.273	0.195	0.195	0.099	0.203	0.470	0.193
		0.5	0.196	0.099	0.195	0.297	0.194	0.198	0.100	0.198	0.312	0.196	0.201	0.101	0.200	0.329	0.197	0.199	0.100	0.207	0.615	0.196

Notes: k= # clusters per arm; ρ_O is the ICC of the (conditional) outcome model (see Eqn. 6 in the manuscript text); ρ_M is the ICC of the probability of missing (POM) model (see Eqn. 7B in the manuscript text; Results based on 1000 simulated data sets per scenario; Note that the finite-sample corrections are due to: Kauermann G & Carroll R (2001, JASA 96(456):1387-1396); Mancl LA & DeRouen TA (2001, Biometrics 57(1):126-134); and Fay & Graubard (2001, Biometrics 57(4):1198-1206), with all details described in Section 2 above.

Figure S2: Coverage (%) of five GEE methods to handle missing outcomes in CRTs



Footnote: Results based on 1000 simulated data sets per scenario using standard Wald Z-based confidence intervals for each modeling approach, except for MMI-GEE for which t-based confidence intervals are used based on 15 imputations (see Section 4.2 of main manuscript for details, including the t-distribution degrees of freedom). Acceptable coverage ranges from 93.6% to 96.4%.

5 Supplemental materials for the Discussion section

5.1 Brief summary of references that consider weighted analyses of clustered outcome data

For weighted GEE, there is limited literature on whether the clustered data structure should be accounted for in the PS model (Eqn. 3) and, what does exist, is primarily for the propensity to be treated rather than propensity to be observed, where the latter is the situation considered in the current manuscript [6, 7, 8]. In the context of selection bias in CRTs in which participants are recruited after clusters have been randomised, Leyrat et al. [6, 7] did not account for clustering when calculating the propensity to be selected to a cluster because the CRT design meant that there was no variability in treatment indicator within a cluster, which is a different context to that considered in the current manuscript in which missingness is at the individual level. (Of note, if independent of each other, inverse probability weights for selection to treatment and those for missingness can be multiplied together to simultaneously account for both features.) In a different context with multilevel observational data and individual-level treatment assignment, Li et el. [8] concluded that the SE can be overestimated if clustering is ignored when calculating the propensity to be treated but that no bias appears when there is no unobserved cluster-level confounding (where we note that the settings considered in the current manuscript also exhibit no unobserved cluster-level confounding). Further, in the longitudinal data setting but with missing outcomes and not selection effects, most inverse probability weight calculations do not account for the clustering under MAR [9]. This may, however, be primarily due to the fact that the weights are cluster-specific (where the cluster is the individual) and thus different to those that we have considered in the current manuscript for outcome missingness at the individual-level in CRTs.

What all of these manuscripts indicate is that it is important to understand the role of clustering in the process for which weighting is being used. When used to account for missing outcomes under CDM with clustering that is unrelated to the outcome, the random effects in the true missingness mechanism induce correlated nonresponse but essentially have no impact on the observed or missing outcomes. With unmeasured cluster-level confounding, the random cluster effects in the missingness model and the outcome model are often correlated and the missingness is non-ignorable. For clustered survey nonresponse, Skinner and D'Arrigo [10] demonstrated substantial bias of the weighting estimator under both mechanisms when the PS was estimated by a random-effects logistic model, especially when the cluster sizes were not large. By contrast, they found that, under MAR, the weighting estimator remained unbiased when the PS weights were estimated ignoring the random effects. In the current manuscript, we have specifically considered the CDM mechanism with fully observed baseline covariates (a special case of MAR) and confirmed their findings in the context of CRTs with missing outcomes. However, future work is warranted to examine the implications of non-ignorable missing outcomes on the weighted GEE analyses of CRTs.

5.2 Supporting evidence for findings regarding generation of the weights for weighted GEE

In the current manuscript, we have examined the performance of two weighted-GEE approaches (W-GEE and CW-GEE) to account for informative missing outcome data in cohort CRTs with a single follow-up time point. Importantly, we assumed that the missing outcome data process was dependent on fully observed baseline covariates, was across a range of degrees of clustering of the missing outcomes (from no clustering to extreme clustering of missingness), and that the clustering in the missingness model was independent of the clustering in the outcome model. As noted in the Results and Discussion sections of the main manuscript, in that context, we demonstrated that clustering should not be accounted for when estimating the weights, even when there is clustering in the missingness mechanism (i.e. even when ICCM, $\rho_M > 0$). That is, W-GEE is preferred to CW-GEE.

Further explanation of these results concerning W-GEE and CW-GEE can be obtained from the perspective of balancing of covariates. From that perspective, there are many commonalities between addressing selection bias from missing data and addressing confounding bias in non-randomized observational studies [8, 11]. With missing outcomes, the ultimate goal of propensity score weighting is to make sure that the weighted covariate distributions among the observed group (i.e. that with complete data) mimics the unweighted covariate distribution of the full data [12]. For example, in our simulation study with a single individual-level binary covariate X_{ij}, we could quantify the effectiveness of weighting in finite samples by the following bias measures:

$$\text{bias}_X = \left| \frac{\sum_i \sum_j R_{ij} X_{ij} w_{ij}}{\sum_i \sum_j R_{ij} w_{ij}} - \frac{\sum_i \sum_j X_{ij}}{\sum_i n_i} \right|$$

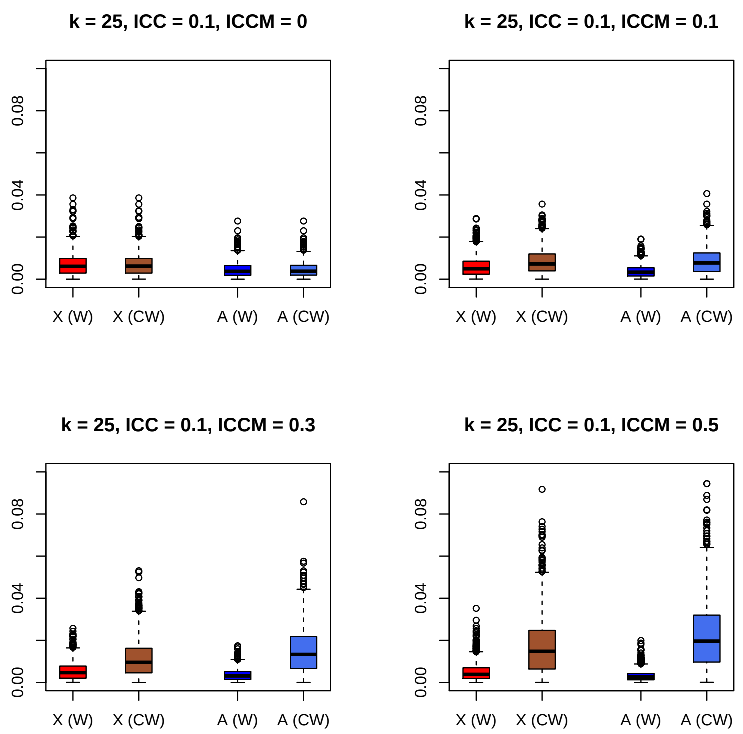
$$\text{bias}_A = \left| \frac{\sum_i \sum_j R_{ij} A_i w_{ij}}{\sum_i \sum_j R_{ij} w_{ij}} - \frac{\sum_i n_i A_i}{\sum_i n_i} \right|$$

where w_{ij} is the inverse probability weight estimated by the logistic model without clustering (W, see Eqn. 4A of the main manuscript) or with clustering (CW, see Eqn. 4B of the main manuscript) for the jth individual ($j = 1, \ldots, n_i$) in the ith cluster ($i = 1, \ldots, M$), and where there are k clusters in each arm so that $M = 2k$. As in the main manuscript, R_{ij} is an indicator of whether the individual-level outcome is observed (i.e. it equals 1 in this case, or 0 if the outcome is missing) and A_i is the treatment arm indicator of the ith cluster (i.e. it equals 1 for the intervention, or 0 for the control arm).

Regardless of how the weights are estimated, as long as bias_X and bias_A are small, the intervention effect estimate from the resulting weighted GEE should have small bias, since weighting has recovered the covariate distributions of the full data from the observed portion. Therefore, the key thing that drives the bias results is the quality of weights, as exemplified by the above bias measures. Even if the true $\rho_M > 0$ and the logistic model with clustering (CW i.e. Eqn. 4B) is the correct model for the weights, it does not necessarily guarantee small bias_X and bias_A. To show this, Figure S3 presents boxplots of these bias measures for 1000 simulations across a range of values of ρ_M, for a subset of the simulations performed in the current manuscript Specifically, we display the results at the four values of ρ_M considered in our simulation study (i.e. $0, 0.1, 0.3$ and 0.5) with the other two parameters fixed at typical values, namely with $k = 25$ and $\rho_O = 0.1$.

Figure S3 shows that for ICCM = 0, i.e. $\rho_M = 0$, the non-clustered logistic model and the clustered logistic model provide weights that achieved similar levels of balance, and so the bias of intervention effect is similar between the corresponding W-GEE and CW-GEE analysis approaches (see this bias in Figure 1 of the main manuscript text). In contrast, when ICCM > 0, i.e. $\rho_M > 0$, the non-clustered logistic model provides weights that are clearly better balanced for the covariates and the treatment indicator across the range of values of ICCM (ρ_M) and therefore it is not surprising that W-GEE is less biased than CW-GEE.

Figure S3: Boxplots of bias_X (X) and bias_A (A) when the weights are estimated by the logistic model without clustering (W) or by the logistic model with clustering (CW) across the 4 values of ρ_M for the setting of $k = 25$ and ICC (ρ_O) = 0.1



Footnote: More specifically, the logistic model without clustering is given by Eqn. (4A) of the main manuscript text and that for the logistic model with clustering is given by Eqn. (4B) of the main manuscript text.

5.3 Additional simulation results for smaller outcome ICC

The results in this section explore the performance of the five GEE methods for moderately-sized CRTs with small outcome clustering, specifically for $\rho_O = 0.01$ and $k = 25$. Both exchangeable (Table S9) and independent (Table S10) working correlation matrices are assumed. The results from these tables can be compared to those for $k = 25$ and $\rho_O = 0.05, 0.1, 0.2$. Specifically, to compare to the setting with $k = 25$ under $\rho_O = 0.05, 0.1, 0.2$, Table S9 is related to Table S3 (for exchangeable working correlation matrix) and Table S10 is related to Table S6 (for independent working correlation matrix).

Table S9 Additional simulation study results summary with exchangeable working correlation matrix for $\rho_O = 0.01$ and $k = 25$.

	Mean relative bias (%)					Coverage (%)					Mean SE					MCSD				
ρ_M	CRA-GEE	A-CRA-GEE	W-GEE	CW-GEE	MMI-GEE	CRA-GEE	A-CRA-GEE	W-GEE	CW-GEE	MMI-GEE	CRA-GEE	A-CRA-GEE	W-GEE	CW-GEE	MMI-GEE	CRA-GEE	A-CRA-GEE	W-GEE	CW-GEE	MMI-GEE
0	-2.5	3.2	0.0	0.0	0.8	92.1	91.4	92.8	93.0	96.5	0.146	0.143	0.153	0.153	0.163	0.151	0.154	0.160	0.160	0.154
0.1	-2.5	2.7	-0.5	-0.4	0.2	93.4	93.0	93.9	94.1	97.5	0.144	0.143	0.151	0.160	0.163	0.144	0.147	0.148	0.160	0.146
0.3	-1.8	2.9	-0.3	0.1	0.3	92.9	92.2	94.7	90.2	97.2	0.146	0.143	0.150	0.184	0.167	0.148	0.150	0.151	0.203	0.150
0.5	-1.2	3.0	-0.1	0.9	0.4	93.5	91.5	94.0	85.5	97.5	0.146	0.143	0.149	0.206	0.170	0.149	0.151	0.150	0.250	0.151

Table S10 Additional simulation study results summary with independent working correlation matrix for $\rho_O = 0.01$ and $k = 25$.

	Mean relative bias (%)					Coverage (%)					Mean SE					MCSD				
ρ_M	CRA-GEE	A-CRA-GEE	W-GEE	CW-GEE	MMI-GEE	CRA-GEE	A-CRA-GEE	W-GEE	CW-GEE	MMI-GEE	CRA-GEE	A-CRA-GEE	W-GEE	CW-GEE	MMI-GEE	CRA-GEE	A-CRA-GEE	W-GEE	CW-GEE	MMI-GEE
0	-2.5	3.2	0.0	0.0	0.8	92.1	91.9	93.0	93.0	96.5	0.146	0.143	0.153	0.153	0.163	0.151	0.154	0.159	0.159	0.153
0.1	-2.5	2.8	-0.4	-0.4	0.2	93.3	93.1	93.8	94.0	97.4	0.145	0.143	0.151	0.160	0.163	0.144	0.147	0.148	0.162	0.146
0.3	-1.7	2.9	-0.2	0.2	0.3	92.8	91.9	93.9	91.6	96.9	0.146	0.143	0.150	0.185	0.167	0.149	0.151	0.151	0.204	0.150
0.5	-1.1	3.0	-0.1	1.0	0.4	93.2	91.3	93.8	87.8	97.5	0.146	0.143	0.149	0.206	0.170	0.150	0.152	0.150	0.249	0.151

Footnote to Tables S9 and S10: Based on 1000 simulated data sets per scenario (with 15 imputed data sets for MMI-GEE). ICC and ICCM, on the logistic scale, correspond to the logistic outcome model ICC (ρ_O) and the missing outcome ICC (ρ_M), respectively (see equations 6 and 7A-B, respectively). Acronyms: CRA-GEE: complete records GEE; A-CRA-GEE: adjusted CRA-GEE; W-GEE: weighted GEE (no adjustment for clustering when estimating the weights); CW-GEE: weighted GEE accounting for clustering when estimating the weights; MMI-GEE: multilevel multiple imputation GEE.

References

[1] Geert Molenberghs, Garrett Fitzmaurice, Michael G Kenward, Anastasios Tsiatis, and Geert Verbeke. *Handbook of missing data methodology*. Chapman and Hall/CRC, 2014.

[2] Roderick JA Little. Modeling the drop-out mechanism in repeated-measures studies. *Journal of the american statistical association*, 90(431):1112–1121, 1995.

[3] Göran Kauermann and Raymond J Carroll. A note on the efficiency of sandwich covariance matrix estimation. *Journal of the American Statistical Association*, 96(456):1387–1396, 2001.

[4] Lloyd A Mancl and Timothy A DeRouen. A covariance estimator for gee with improved small-sample properties. *Biometrics*, 57(1):126–134, 2001.

[5] Michael P Fay and Barry I Graubard. Small-sample adjustments for wald-type tests using sandwich estimators. *Biometrics*, 57(4):1198–1206, 2001.

[6] C Leyrat, A Caille, A Donner, and B Giraudeau. Propensity scores used for analysis of cluster randomized trials with selection bias: a simulation study. *Statistics in medicine*, 32(19):3357–3372, 2013.

[7] C Leyrat, A Caille, A Donner, and B Giraudeau. Propensity score methods for estimating relative risks in cluster randomized trials with low-incidence binary outcomes and selection bias. *Statistics in medicine*, 33(20):3556–3575, 2014.

[8] Fan Li, Alan M Zaslavsky, and Mary Beth Landrum. Propensity score weighting with multilevel data. *Statistics in medicine*, 32(19):3373–3387, 2013.

[9] John S Preisser, Kurt K Lohman, and Paul J Rathouz. Performance of weighted estimating equations for longitudinal binary data with drop-outs missing at random. *Statistics in medicine*, 21(20):3035–3054, 2002.

[10] Chris J Skinner. Inverse probability weighting for clustered nonresponse. *Biometrika*, 98(4):953–966, 2011.

[11] Fan Li, Laine E Thomas, and Fan Li. Addressing extreme propensity scores via the overlap weights. *American journal of epidemiology*, 188(1):250–257, 2018.

[12] José R Zubizarreta. Stable weights that balance covariates for estimation with incomplete outcome data. *Journal of the American Statistical Association*, 110(511):910–922, 2015.