Supporting Information for

Randomization-Based Inference for a Marginal Treatment Effect in Stepped Wedge Cluster Randomized Trials

Dustin J. Rabideau[*1,2] **and Rui Wang**[3,4]

[1]Biostatistics Center, Massachusetts General Hospital, Boston, Massachusetts
[2]Department of Medicine, Harvard Medical School, Boston, Massachusetts
[3]Department of Population Medicine, Harvard Pilgrim Health Care Institute and Harvard Medical School, Boston, Massachusetts
[4]Department of Biostatistics, Harvard T. H. Chan School of Public Health, Boston, Massachusetts
[*]*email:* drabideau@mgh.harvard.edu

A Details of Alternative Methods Using Cluster-Period Summary Statistics

We introduce some additional notation for the alternative randomization-based methods using cluster-period summary statistics we considered in our simulations and data example. Given N total clusters in the study, we use N_{1j} and N_{0j} to denote the number of clusters on treatment and control, respectively, in period j. Z_{ij} represents a single summary measure for the ijth cluster-period. For example, with continuous individual-level outcomes Y_{ijk}, $Z_{ij} = m_{ij}^{-1} \sum_{k=1}^{m_{ij}} Y_{ijk}$ is the sample average of the outcomes; with binary Y_{ijk}, $Z_{ij} = \log\{p_{ij}/(1-p_{ij})\}$ with $p_{ij} = m_{ij}^{-1} \sum_{k=1}^{m_{ij}} Y_{ijk}$ is the empirical log odds of the outcomes.

A.1 Nonparametric Within-Period Method, Thompson et al. (2018)

Thompson et al proposed a nonparametric within-period (NPWP) randomization-based method for analyzing SWTs. Their NPWP method compares weighted cluster-period means between treatment conditions within each period $1 < j < J$. The test statistic for their approach is

$$\hat{\psi} = \frac{1}{w} \sum_{1<j<J} w_j \hat{\psi}_j, \tag{S1}$$

where

$$\hat{\psi}_j = \frac{1}{N_{1j}} \sum_{i=1}^{N} X_{ij} Z_{ij} - \frac{1}{N_{0j}} \sum_{i=1}^{N} (1 - X_{ij}) Z_{ij}, \quad 1 < j < J,$$

$$w_j = \left[\left\{ \frac{(N_{1j} - 1)s_{1j}^2 + (N_{0j} - 1)s_{0j}^2}{N_{1j} + N_{0j} - 2} \right\} \times \left(\frac{1}{N_{1j}} + \frac{1}{N_{0j}} \right) \right]^{-1}, \quad 1 < j < J,$$

with $w = \sum_{1<j<J} w_j$, and where s_{1j}^2 and s_{0j}^2 are the empirical variances of the cluster-period summaries in the treatment and control groups, respectively. The randomization test is based

on permuting treatment assignments in accordance with the SWT design to recalculate the test statistic in (S1), i.e. using $X_{ij} = X_{ij}^{(1)} = x_{ij}$ (the observed treatment values) to calculate the observed test statistic $\hat{\psi}^{(1)}$, and $X_{ij} = X_{ij}^{(p)}$ to calculate the pth permuted test statistic $\hat{\psi}^{(p)}$. Randomization-based CIs can be calculated using the usual transformation approach for the continuous cluster-period summaries, i.e. to test $H_0 : \psi = \psi_0$, we would actually conduct a randomization test of $H_0^* : \psi = 0$ using transformed cluster-period summaries

$$Z_{ij}^* = (Z_{ij} - \psi_0)x_{ij} + Z_{ij}(1 - x_{ij}).$$

Note, we use ψ here to distinguish it from θ, as defined in (1) in the manuscript main text, as they are generally distinct parameters when there is a non-zero intervention effect.

A.2 Crossover Method, Kennedy-Shaffer et al. (2020)

Kennedy-Shaffer et al proposed a crossover randomization-based method for analyzing SWTs. Their crossover method compares weighted cluster-period mean *differences* between clusters that cross over to those that do not at each crossover point, i.e. time periods $j > 1$. The test statistic for this approach takes a form similar to (S1), specifically,

$$\hat{\psi} = \frac{1}{w} \sum_{j>1} w_j \hat{\psi}_j,$$

but now,

$$\hat{\psi}_j = \frac{1}{N_{1j}^*} \sum_{i=1}^{N} X_{ij} D_{ij} - \frac{1}{N_{0j}^*} \sum_{i=1}^{N} (1 - X_{ij}) D_{ij}, \quad j > 1,$$

where $D_{ij} = Z_{ij} - Z_{i(j-1)}$ represents the difference in cluster-period summaries between consecutive periods, and $N_{1j}^* = \sum_{i=1}^{N} (X_{ij} - X_{i(j-1)})$ and $N_{0j}^* = \sum_{i=1}^{N} \{1 - (X_{ij} - X_{i(j-1)})\}$ are the number of clusters that did or did not crossover to treatment in period j, respectively. Note, we included clusters on and off treatment in both consecutive periods in N_{0j}^* (i.e. $\hat{\psi}$ corresponds to $\tilde{\beta}$ from equation 12 in their paper). We applied the weights provided by Kennedy-Shaffer et al. (2020)

$$w_j = \left(\frac{1}{N_{1j}^*} + \frac{1}{N_{0j}^*} \right)^{-1},$$

which maximize efficiency when all variances of the cluster-period summaries are assumed equal. Randomization-based p-values and CIs are calculated in the same way as outlined above in Section A.1, but now using this alternative way of calculating $\hat{\psi}$.

A.3 Closed-Form Permutation-Based Approach, Hughes et al. (2020)

Hughes et al proposed closed-form randomization-based expressions to estimate a vertical intervention effect (similar to that targeted by the NPWP method) and its corresponding variance in SWTs. Their estimator was defined as

$$\hat{\psi} = \frac{\sum_{i=1}^{N}\sum_{j=1}^{J} Z_{ij}(x_{ij} - \bar{x}_j)}{N\sum_{j=1}^{J}\bar{x}_j(1-\bar{x}_j)}, \tag{S2}$$

where $\bar{x}_j = N^{-1}\sum_{i=1}^{N} x_{ij}$ is the sample proportion of clusters under intervention in period j (i.e. $\hat{\psi}$ corresponds to $\hat{\delta}$ from equation 12 in their paper). Note that like the NPWP method, only cluster-period summary statistics from periods $1 < j < J$ contribute to (S2), as the numerator term $(x_{ij} - \bar{x}_j) = 0$ for $j \in \{1, J\}$. We used one of the expressions they proposed to estimate the variance of $\hat{\psi}$, specifically

$$\hat{v} = \left(\frac{N}{N-1}\right)\left[\sum_{i=1}^{N}\left\{\sum_{j=1}^{J}(Z_{ij} - x_{ij}\hat{\psi})^2\bar{x}_j(1-\bar{x}_j) + 2\sum_{j<j'}(Z_{ij} - x_{ij}\hat{\psi})(Z_{ij'} - x_{ij'}\hat{\psi})\bar{x}_j(1-\bar{x}_{j'})\right\}\right.$$
$$\left. - \frac{2}{N-1}\sum_{i<i'}\sum_{j=1}^{J}\sum_{j'=1}^{J}(Z_{ij} - x_{ij}\hat{\psi})(Z_{i'j'} - x_{i'j'}\hat{\psi})\bar{x}_{\min(j,j')}(1-\bar{x}_{\max(j,j')})\right] \Bigg/$$
$$\{N\sum_{j=1}^{J}\bar{x}_j(1-\bar{x}_j)\}^2,$$

to carry out Wald tests and calculate CIs, which demonstrated adequate performance across their simulations (i.e. $\hat{v}$ corresponds to $\{N/(N-1)\} \times V^1_{\delta=\hat{\delta}}(\hat{\delta})$ from equation 15 in their paper). Note, their expressions of $\hat{\psi}$ and $\hat{v}$ were derived assuming equal cluster-period sizes.

B Details of Search Procedure Adapted from Garthwaite and Jones (2009)

For longer searches (e.g. $P \geq 200,000$), Garthwaite and Jones (2009) proposed an improvement on the algorithm in Garthwaite (1996) by modifying the step size during later phases of the search and averaging, rather than using only the final values, for CI estimation. For example, suppose we carry out a P-step search for each bound, broken into three separate phases, with $P = P_1 + P_2 + P_3$. Updates are made using the same general formulas

$$U^{(p+1)} = \begin{cases} U^{(p)} - a_p(\alpha/2), & \text{if } \hat{\tau}^{(p)}(\mathbf{X}^{(p)}) > \hat{\tau}^{(p)}(\mathbf{x}) \\ U^{(p)} + a_p(1-\alpha/2), & \text{if } \hat{\tau}^{(p)}(\mathbf{X}^{(p)}) \leq \hat{\tau}^{(p)}(\mathbf{x}) \end{cases}$$
$$L^{(p+1)} = \begin{cases} L^{(p)} + a_p(\alpha/2), & \text{if } \hat{\tau}^{(p)}(\mathbf{X}^{(p)}) < \hat{\tau}^{(p)}(\mathbf{x}) \\ L^{(p)} - a_p(1-\alpha/2), & \text{if } \hat{\tau}^{(p)}(\mathbf{X}^{(p)}) \geq \hat{\tau}^{(p)}(\mathbf{x}), \end{cases}$$

but using a different step length a_p in each phase:

$$a_p = \begin{cases} s/p, & \text{(Phase 1)} \quad p = m, \ldots, m + P_1 \\ s/(m+P_1), & \text{(Phase 2)} \quad p = m + P_1, \ldots, m + P_1 + P_2 \\ s/\{p(m+P_1)/(m+P_1+P_2)\}, & \text{(Phase 3)} \quad p = m + P_1 + P_2, \ldots, m + P\,. \end{cases}$$

1. Phase 1 begins with $p = m$ and $[L^{(m)}, U^{(m)}]$ as starting values. We carry out a P_1-step search with $a_p = s/p$. Garthwaite and Jones (2009) suggest keeping this phase short by choosing $P_1 = \min(5000, P/20)$. Phase 1 ends with estimates $[L^{(m+P_1)}, U^{(m+P_1)}]$.

2. Phase 2 begins with $p = m+P_1$ and $[L^{(m+P_1)}, U^{(m+P_1)}]$ as starting values. We carry out a P_2-step search with $a_p = s/(m+P_1)$. Garthwaite and Jones (2009) recommend $P_2 = 14 \times P_1$ as a reasonable choice. Phase 2 ends with estimates $[L^{(m+P_1+P_2)}, U^{(m+P_1+P_2)}]$.

3. Phase 3 begins with $p = m + P_1 + P_2$ and $[L^{(m+P_1+P_2)}, U^{(m+P_1+P_2)}]$ as starting values. We carry out a P_3-step search with $a_p = s/\{p(m+P_1)/(m+P_1+P_2)\}$. Garthwaite and Jones (2009) suggest keeping this phase long; they found particularly good efficiency for $P_3 \geq 200,000 - P_1 - P_2$, i.e. an overall $P \geq 200,000$. Phase 3 ends with estimates $[L^{(m+P)}, U^{(m+P)}]$.

Rather than choosing the final updated values $[L^{(m+P)}, U^{(m+P)}]$, the unweighted averages of the final n values are taken as the final CI, i.e. $[\bar{L}, \bar{U}]$ where

$$\bar{L} = \frac{1}{n} \sum_{p=m+P-n+1}^{m+P} L^{(p)} \quad \text{and} \quad \bar{U} = \frac{1}{n} \sum_{p=m+P-n+1}^{m+P} U^{(p)}.$$

Garthwaite and Jones (2009) suggest using $n = P - 2P_1$. Additional guidance about fine-tuning the algorithm, such as choosing an appropriate number of steps, starting values, or step length, can be found in Rabideau and Wang (2020), Garthwaite (1996), and Garthwaite and Jones (2009).

C Figures

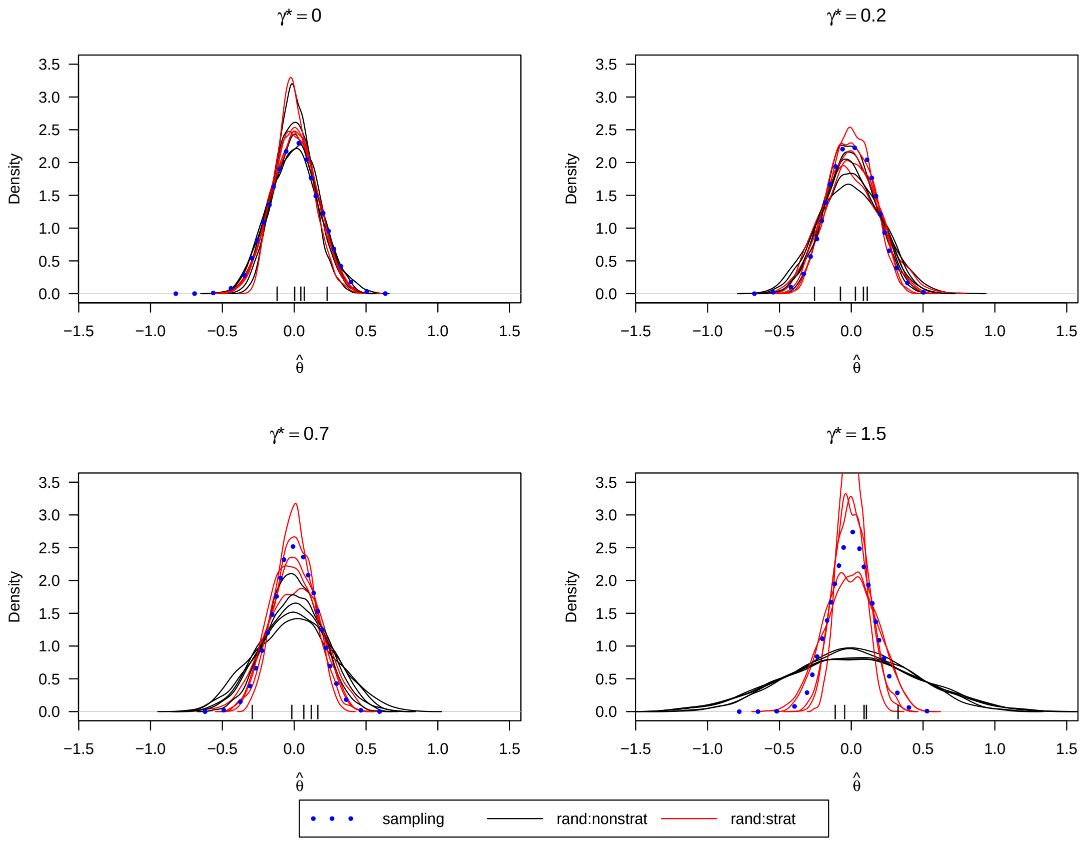

Figure S1. Kernel density plots of stratified (lighter red lines) and nonstratified (darker black lines) randomization distributions of $\hat{\theta}$ with $P = 5{,}000$ in 20 (5 at each level of γ^*) simulated stratified SWTs. Dotted blue lines correspond to the underlying sampling distributions of $\hat{\theta}$. Increasing values of γ^* correspond to increasing strengths of Y-Z association. Data sets were generated via model (7) as described in Section 3.2.1.

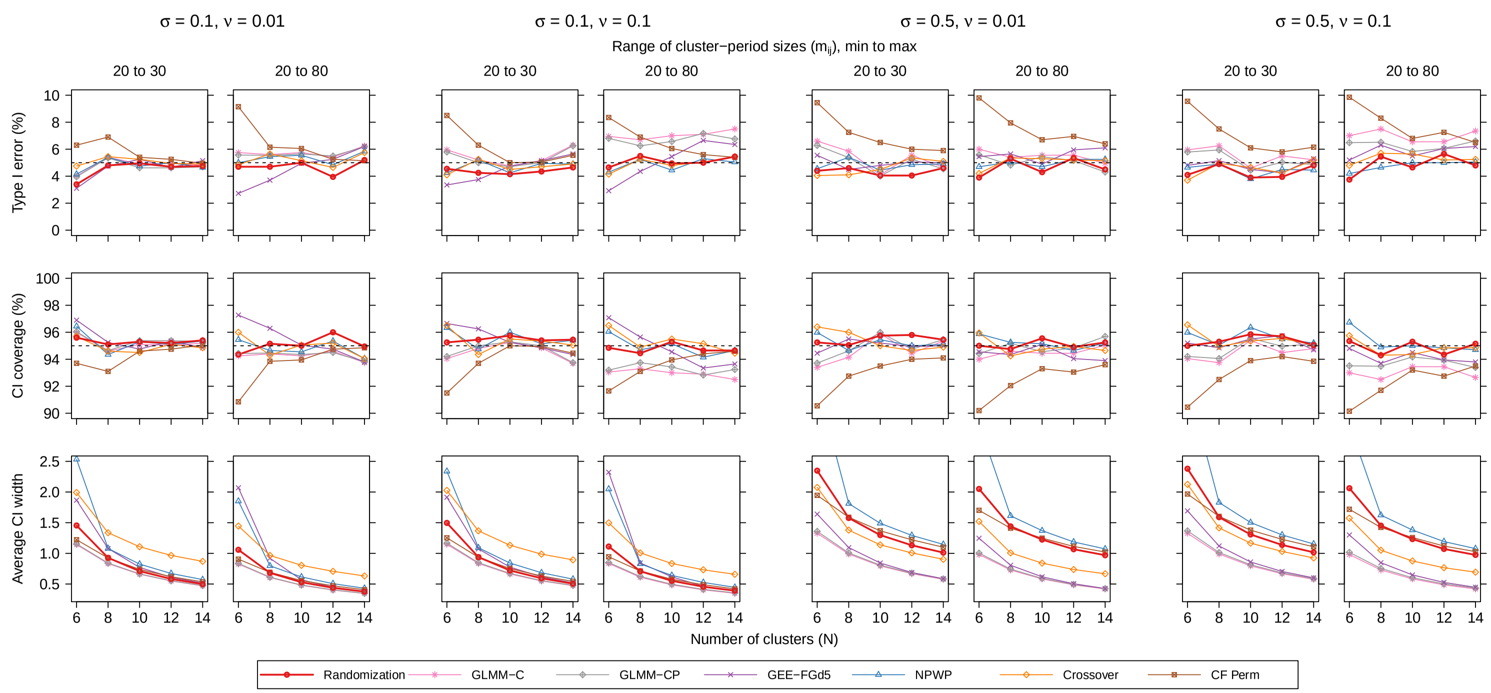

Figure S2. [Nonstratified randomization, **larger** (σ, ν)] Simulation results across 2,000 data sets simulated via model (5) as described in Section 3.1.1 with $\theta^* = 0$. Methods considered included our proposed individual-level randomization-based approach (Randomization), a logistic mixed model with random cluster effects (GLMM-C), a logistic mixed model with random cluster and cluster-period effects (GLMM-CP), a marginal model fit via a small-sample adjusted generalized estimating equation (GEE-FGd5), a nonparametric within-period approach (NPWP), a crossover method (Crossover), and a closed-form permutation-based estimator (CF-Perm).

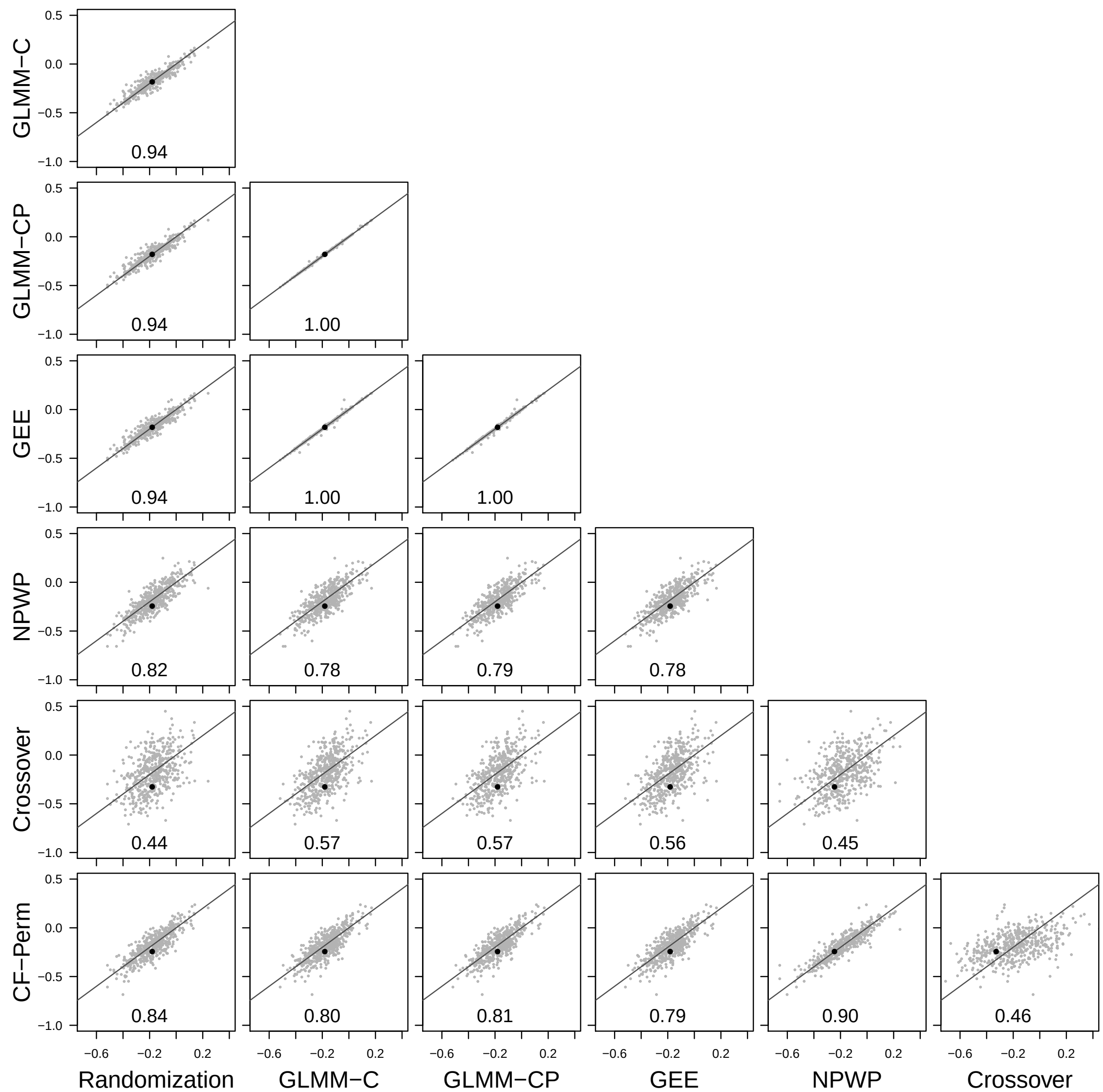

Figure S3. Simulated (smaller grey dots) and actual (larger black dots) log OR estimates of the intervention effect in the XpertMTB/RIF SWT. Data were generated via model (5), as described in Section 3.1.1, with parameters set at their estimated values from fitting GLMM-CP to the actual XpertMTB/RIF trial data. Lines indicate $y = x$. Numbers inside plot region indicate Pearson correlation coefficients. Methods considered included our proposed individual-level randomization-based approach (Randomization), a logistic mixed model with random cluster effects (GLMM-C), a logistic mixed model with random cluster and cluster-period effects (GLMM-CP), a marginal model fit via a generalized estimating equation (GEE), a small-sample adjusted GEE (GEE-FGd5), a nonparametric within-period approach (NPWP), a crossover method (Crossover), and a closed-form permutation-based estimator (CF-Perm).

D Tables

Table S1. [Nonstratified randomization, $(\sigma, \nu) = (0.1, 0.01)$] **Power** (%) across 2,000 data sets simulated via model (5) as described in Section 3.1.1. We considered different treatment effects (θ^*), cluster-period size ranges (m_{ij}), and numbers of clusters (N). Methods considered included our proposed individual-level randomization-based approach (Randomization), a logistic mixed model with random cluster effects (GLMM-C), a logistic mixed model with random cluster and cluster-period effects (GLMM-CP), a marginal model fit via a generalized estimating equation (GEE), a small-sample adjusted GEE (GEE-FGd5), a cluster-level nonparametric within-period approach (NPWP), a cluster-level crossover method (Crossover), and a closed-form permutation-based estimator (CF-Perm).

θ^*	Method	$m_{ij} \sim U\{20, 30\}$					$m_{ij} \sim U\{20, 80\}$				
		N									
		6	8	10	12	14	6	8	10	12	14
0	Randomization	3	5	5	5	5	5	5	5	4	5
	GLMM-C	4	5	5	5	5	6	6	6	5	6
	GLMM-CP	4	5	5	5	5	6	6	6	5	6
	GEE	24	18	15	13	9	25	18	14	12	12
	GEE-FGd5	3	5	5	5	5	3	4	5	5	6
	NPWP	4	5	5	5	5	5	5	6	5	6
	Crossover	5	5	5	5	5	5	6	5	5	6
	CF-Perm	6	7	5	5	5	9	6	6	5	5
0.25	Randomization	11	22	28	41	52	17	31	47	63	76
	GLMM-C	15	25	32	45	56	24	38	55	70	82
	GLMM-CP	15	25	32	45	56	23	37	54	70	82
	GEE	36	41	45	56	64	46	53	65	76	85
	GEE-FGd5	6	15	25	39	49	8	24	42	61	76
	NPWP	10	17	23	36	45	15	24	39	52	69
	Crossover	9	15	17	20	24	13	20	27	32	37
	CF-Perm	18	24	30	42	52	26	33	48	60	74
0.5	Randomization	30	63	83	93	99	52	86	97	100	100
	GLMM-C	43	70	86	95	99	70	91	99	100	100
	GLMM-CP	42	69	86	95	99	70	91	99	100	100
	GEE	63	80	90	96	99	82	95	99	100	100
	GEE-FGd5	18	51	75	91	98	31	74	95	99	100
	NPWP	26	55	74	89	96	44	77	93	98	100
	Crossover	20	40	51	64	72	35	62	77	85	93
	CF-Perm	42	68	83	94	98	65	86	96	99	100

Table S2. [Nonstratified randomization, $(\sigma, \nu) = (0.1, 0.01)$] **CI coverage** (%) across 2,000 data sets simulated via model (5) as described in Section 3.1.1. We considered different treatment effects (θ^*), cluster-period size ranges (m_{ij}), and numbers of clusters (N). Methods considered included our proposed individual-level randomization-based approach (Randomization), a logistic mixed model with random cluster effects (GLMM-C), a logistic mixed model with random cluster and cluster-period effects (GLMM-CP), a marginal model fit via a generalized estimating equation (GEE), a small-sample adjusted GEE (GEE-FGd5), a cluster-level nonparametric within-period approach (NPWP), a cluster-level crossover method (Crossover), and a closed-form permutation-based estimator (CF-Perm).

θ^*	Method	$m_{ij} \sim U\{20, 30\}$					$m_{ij} \sim U\{20, 80\}$				
		N									
		6	8	10	12	14	6	8	10	12	14
0	Randomization	96	95	95	95	95	94	95	95	96	95
	GLMM-C	96	95	95	95	95	94	94	94	95	94
	GLMM-CP	96	95	95	95	95	94	94	94	95	94
	GEE	76	82	85	87	91	75	82	86	88	88
	GEE-FGd5	97	95	95	95	95	97	96	95	95	94
	NPWP	96	94	95	95	95	95	95	95	95	94
	Crossover	96	95	94	95	95	96	94	95	95	94
	CF-Perm	94	93	95	95	95	91	94	94	95	95
0.25	Randomization	95	96	96	96	96	95	95	95	95	95
	GLMM-C	94	95	95	95	95	94	95	94	93	94
	GLMM-CP	94	95	95	95	95	94	95	94	94	94
	GEE	76	83	86	89	90	76	82	85	86	88
	GEE-FGd5	97	96	95	95	95	97	96	95	93	94
	NPWP	96	95	95	95	96	96	94	95	95	94
	Crossover	96	94	94	96	95	96	95	95	95	95
	CF-Perm	92	94	94	95	96	92	93	94	94	94
0.5	Randomization	95	95	95	95	95	94	95	95	95	95
	GLMM-C	95	95	94	94	95	94	94	94	94	94
	GLMM-CP	95	95	94	94	95	94	94	95	94	94
	GEE	77	84	84	87	89	76	81	86	88	88
	GEE-FGd5	97	97	95	94	95	97	96	96	94	94
	NPWP	96	96	95	95	95	97	95	95	94	94
	Crossover	97	95	95	95	94	96	95	95	95	95
	CF-Perm	91	94	93	94	95	92	93	94	94	94

Table S3. [Nonstratified randomization, $(\sigma, \nu) = (0.1, 0.01)$] **Average CI width** across 2,000 data sets simulated via model (5) as described in Section 3.1.1. We considered different treatment effects (θ^*), cluster-period size ranges (m_{ij}), and numbers of clusters (N). Methods considered included our proposed individual-level randomization-based approach (Randomization), a logistic mixed model with random cluster effects (GLMM-C), a logistic mixed model with random cluster and cluster-period effects (GLMM-CP), a marginal model fit via a generalized estimating equation (GEE), a small-sample adjusted GEE (GEE-FGd5), a cluster-level nonparametric within-period approach (NPWP), a cluster-level crossover method (Crossover), and a closed-form permutation-based estimator (CF-Perm).

θ^*	Method	$m_{ij} \sim U\{20, 30\}$					$m_{ij} \sim U\{20, 80\}$				
		N									
		6	8	10	12	14	6	8	10	12	14
0	Randomization	1.45	0.93	0.71	0.58	0.50	1.06	0.68	0.53	0.44	0.38
	GLMM-C	1.14	0.83	0.66	0.55	0.47	0.83	0.61	0.48	0.40	0.35
	GLMM-CP	1.15	0.84	0.67	0.55	0.47	0.84	0.61	0.48	0.41	0.35
	GEE	0.81	0.66	0.55	0.47	0.42	0.61	0.48	0.40	0.35	0.31
	GEE-FGd5	1.87	1.08	0.77	0.61	0.51	2.07	0.92	0.57	0.45	0.38
	NPWP	2.53	1.08	0.82	0.67	0.57	1.85	0.80	0.62	0.51	0.43
	Crossover	1.99	1.34	1.11	0.97	0.87	1.45	0.97	0.81	0.71	0.63
	CF-Perm	1.22	0.92	0.74	0.62	0.54	0.91	0.69	0.56	0.47	0.41
0.25	Randomization	1.43	0.91	0.70	0.57	0.48	1.03	0.67	0.52	0.43	0.37
	GLMM-C	1.11	0.81	0.65	0.54	0.46	0.80	0.59	0.47	0.39	0.34
	GLMM-CP	1.12	0.82	0.65	0.54	0.46	0.81	0.60	0.47	0.40	0.34
	GEE	0.80	0.65	0.54	0.47	0.41	0.59	0.47	0.40	0.34	0.30
	GEE-FGd5	1.92	1.15	0.75	0.60	0.50	2.12	0.95	0.56	0.44	0.37
	NPWP	2.38	1.05	0.80	0.65	0.55	2.03	0.78	0.60	0.49	0.42
	Crossover	1.93	1.29	1.07	0.94	0.84	1.39	0.95	0.78	0.68	0.62
	CF-Perm	1.19	0.89	0.72	0.60	0.51	0.88	0.67	0.54	0.45	0.40
0.5	Randomization	1.39	0.88	0.68	0.56	0.48	1.01	0.65	0.51	0.42	0.36
	GLMM-C	1.09	0.80	0.63	0.53	0.45	0.79	0.58	0.46	0.39	0.33
	GLMM-CP	1.11	0.81	0.64	0.53	0.46	0.80	0.59	0.47	0.39	0.34
	GEE	0.79	0.63	0.53	0.45	0.40	0.58	0.46	0.39	0.34	0.30
	GEE-FGd5	2.00	1.04	0.73	0.58	0.49	2.12	0.85	0.55	0.43	0.36
	NPWP	2.52	1.01	0.78	0.64	0.54	2.46	0.76	0.59	0.48	0.41
	Crossover	1.89	1.25	1.04	0.91	0.82	1.39	0.93	0.77	0.67	0.60
	CF-Perm	1.16	0.87	0.70	0.59	0.51	0.87	0.66	0.53	0.45	0.39

Table S4. [Nonstratified randomization, $(\sigma, \nu) = (0.1, 0.01)$] **Average CI width** across 2,000 data sets simulated via model (5) as described in Section 3.1.1 with $\theta^* = 0$. We considered different **fixed and variable** cluster-period size ranges (m_{ij}) and numbers of clusters (N). Methods considered included our proposed individual-level randomization-based approach (Randomization), a logistic mixed model with random cluster effects (GLMM-C), a logistic mixed model with random cluster and cluster-period effects (GLMM-CP), a marginal model fit via a generalized estimating equation (GEE), a small-sample adjusted GEE (GEE-FGd5), a cluster-level nonparametric within-period approach (NPWP), a cluster-level crossover method (Crossover), and a closed-form permutation-based estimator (CF-Perm).

Method	Fixed $m_{ij} = 25$					Variable $m_{ij} \sim U\{20, 30\}$				
	N									
	6	8	10	12	14	6	8	10	12	14
Randomization	1.44	0.93	0.71	0.58	0.50	1.45	0.93	0.71	0.58	0.50
GLMM-C	1.14	0.83	0.66	0.55	0.47	1.14	0.83	0.66	0.55	0.47
GLMM-CP	1.15	0.84	0.67	0.55	0.47	1.15	0.84	0.67	0.55	0.47
GEE	0.82	0.66	0.55	0.47	0.42	0.81	0.66	0.55	0.47	0.42
GEE-FGd5	1.76	1.05	0.77	0.61	0.51	1.87	1.08	0.77	0.61	0.51
NPWP	2.59	1.07	0.82	0.67	0.57	2.53	1.08	0.82	0.67	0.57
Crossover	1.98	1.33	1.09	0.96	0.86	1.99	1.34	1.11	0.97	0.87
CF-Perm	1.20	0.92	0.73	0.61	0.53	1.22	0.92	0.74	0.62	0.54
	$m_{ij} = 50$					$m_{ij} = 20$ to 80				
Randomization	1.01	0.67	0.53	0.43	0.37	1.06	0.68	0.53	0.44	0.38
GLMM-C	0.81	0.60	0.48	0.40	0.34	0.83	0.61	0.48	0.40	0.35
GLMM-CP	0.82	0.60	0.48	0.40	0.35	0.84	0.61	0.48	0.41	0.35
GEE	0.59	0.48	0.41	0.35	0.31	0.61	0.48	0.40	0.35	0.31
GEE-FGd5	1.23	0.74	0.55	0.44	0.37	2.07	0.92	0.57	0.45	0.38
NPWP	1.85	0.75	0.59	0.48	0.41	1.85	0.80	0.62	0.51	0.43
Crossover	1.32	0.89	0.73	0.64	0.58	1.45	0.97	0.81	0.71	0.63
CF-Perm	0.84	0.64	0.52	0.44	0.38	0.91	0.69	0.56	0.47	0.41

Table S5. [Nonstratified randomization, **larger** (σ, ν)] **Type I error** (%) across 2,000 data sets simulated via model (5) as described in Section 3.1.1 with $\theta^* = 0$. We considered different cluster-period size ranges (m_{ij}) and numbers of clusters (N). Methods considered included our proposed individual-level randomization-based approach (Randomization), a logistic mixed model with random cluster effects (GLMM-C), a logistic mixed model with random cluster and cluster-period effects (GLMM-CP), a marginal model fit via a generalized estimating equation (GEE), a small-sample adjusted GEE (GEE-FGd5), a cluster-level non-parametric within-period approach (NPWP), a cluster-level crossover method (Crossover), and a closed-form permutation-based estimator (CF-Perm).

σ	ν	Method	$m_{ij} \sim U\{20, 30\}$					$m_{ij} \sim U\{20, 80\}$				
			N									
			6	8	10	12	14	6	8	10	12	14
0.1	0.01	Randomization	3	5	5	5	5	5	5	5	4	5
		GLMM-C	4	5	5	5	5	6	6	6	5	6
		GLMM-CP	4	5	5	5	5	6	6	6	5	6
		GEE	24	18	15	13	9	25	18	14	12	12
		GEE-FGd5	3	5	5	5	5	3	4	5	5	6
		NPWP	4	5	5	5	5	5	5	6	5	6
		Crossover	5	5	5	5	5	5	6	5	5	6
		CF-Perm	6	7	5	5	5	9	6	6	5	5
0.1	0.1	Randomization	5	4	4	4	5	5	6	5	5	5
		GLMM-C	6	5	5	5	6	7	7	7	7	8
		GLMM-CP	6	5	5	5	6	7	6	7	7	7
		GEE	24	17	14	12	11	25	17	14	13	12
		GEE-FGd5	3	4	5	5	6	3	4	5	7	6
		NPWP	4	5	4	5	5	4	5	4	5	5
		Crossover	4	5	4	5	5	4	5	5	5	5
		CF-Perm	8	6	5	5	6	8	7	6	6	5
0.5	0.01	Randomization	4	5	4	4	5	4	5	4	5	4
		GLMM-C	7	6	4	6	5	6	5	6	6	5
		GLMM-CP	6	5	4	5	5	6	5	5	5	4
		GEE	20	14	11	10	9	20	15	12	11	11
		GEE-FGd5	6	4	5	5	5	5	6	5	6	6
		NPWP	5	5	4	5	5	5	5	5	5	5
		Crossover	4	4	4	5	5	4	5	5	5	5
		CF-Perm	9	7	6	6	6	10	8	7	7	6
0.5	0.1	Randomization	4	5	4	4	5	4	5	5	6	5
		GLMM-C	6	6	5	6	5	7	8	7	7	7
		GLMM-CP	6	6	4	5	5	6	7	6	6	7
		GEE	19	15	11	10	10	20	16	12	11	11
		GEE-FGd5	5	5	4	4	5	5	6	6	6	6
		NPWP	5	5	4	4	4	4	5	5	5	5
		Crossover	4	5	5	4	5	5	6	6	5	5
		CF-Perm	10	8	6	6	6	10	8	7	7	6

Table S6. [Nonstratified randomization, **larger** (σ, ν)] **CI coverage** (%) across 2,000 data sets simulated via model (5) as described in Section 3.1.1 with $\theta^* = 0$. We considered different cluster-period size ranges (m_{ij}) and numbers of clusters (N). Methods considered included our proposed individual-level randomization-based approach (Randomization), a logistic mixed model with random cluster effects (GLMM-C), a logistic mixed model with random cluster and cluster-period effects (GLMM-CP), a marginal model fit via a generalized estimating equation (GEE), a small-sample adjusted GEE (GEE-FGd5), a cluster-level nonparametric within-period approach (NPWP), a cluster-level crossover method (Crossover), and a closed-form permutation-based estimator (CF-Perm).

			$m_{ij} \sim U\{20, 30\}$					$m_{ij} \sim U\{20, 80\}$				
σ	ν	Method	N									
			6	8	10	12	14	6	8	10	12	14
0.1	0.01	Randomization	96	95	95	95	95	94	95	95	96	95
		GLMM-C	96	95	95	95	95	94	94	94	95	94
		GLMM-CP	96	95	95	95	95	94	94	94	95	94
		GEE	76	82	85	87	91	75	82	86	88	88
		GEE-FGd5	97	95	95	95	95	97	96	95	95	94
		NPWP	96	94	95	95	95	95	95	95	95	94
		Crossover	96	95	94	95	95	96	94	95	95	94
		CF-Perm	94	93	95	95	95	91	94	94	95	95
0.1	0.1	Randomization	95	95	96	95	95	95	94	95	95	95
		GLMM-C	94	95	95	95	94	93	93	93	93	92
		GLMM-CP	94	95	95	95	94	93	94	93	93	93
		GEE	76	83	86	88	89	75	83	86	87	88
		GEE-FGd5	97	96	95	95	94	97	96	95	93	94
		NPWP	96	95	96	95	95	96	95	95	94	95
		Crossover	96	94	96	95	95	96	95	96	95	94
		CF-Perm	92	94	95	95	94	92	93	94	94	95
0.5	0.01	Randomization	95	95	96	96	95	95	95	96	95	95
		GLMM-C	93	94	96	94	95	94	95	94	94	95
		GLMM-CP	94	95	96	95	95	94	95	95	95	96
		GEE	80	86	89	90	91	80	85	88	89	89
		GEE-FGd5	94	96	95	95	95	95	94	95	94	94
		NPWP	96	95	95	95	95	96	95	95	95	95
		Crossover	96	96	95	95	95	96	94	95	95	95
		CF-Perm	91	93	94	94	94	90	92	93	93	94
0.5	0.1	Randomization	95	95	96	96	95	95	94	95	94	95
		GLMM-C	94	94	95	94	95	93	92	93	93	93
		GLMM-CP	94	94	96	95	95	94	93	94	94	93
		GEE	81	85	89	90	90	80	84	88	89	89
		GEE-FGd5	95	95	96	96	95	95	94	94	94	94
		NPWP	96	95	96	96	95	97	95	95	95	95
		Crossover	97	95	95	96	95	96	94	94	95	95
		CF-Perm	90	92	94	94	94	90	92	93	93	94

Table S7. [Nonstratified randomization, **larger** (σ, ν)] **Average CI width** across 2,000 data sets simulated via model (5) as described in Section 3.1.1 with $\theta^* = 0$. We considered different cluster-period size ranges (m_{ij}) and numbers of clusters (N). Methods considered included our proposed individual-level randomization-based approach (Randomization), a logistic mixed model with random cluster effects (GLMM-C), a logistic mixed model with random cluster and cluster-period effects (GLMM-CP), a marginal model fit via a generalized estimating equation (GEE), a small-sample adjusted GEE (GEE-FGd5), a cluster-level non-parametric within-period approach (NPWP), a cluster-level crossover method (Crossover), and a closed-form permutation-based estimator (CF-Perm).

σ	ν	Method	$m_{ij} \sim U\{20, 30\}$					$m_{ij} \sim U\{20, 80\}$				
			N									
			6	8	10	12	14	6	8	10	12	14
0.1	0.01	Randomization	1.45	0.93	0.71	0.58	0.50	1.06	0.68	0.53	0.44	0.38
		GLMM-C	1.14	0.83	0.66	0.55	0.47	0.83	0.61	0.48	0.40	0.35
		GLMM-CP	1.15	0.84	0.67	0.55	0.47	0.84	0.61	0.48	0.41	0.35
		GEE	0.81	0.66	0.55	0.47	0.42	0.61	0.48	0.40	0.35	0.31
		GEE-FGd5	1.87	1.08	0.77	0.61	0.51	2.07	0.92	0.57	0.45	0.38
		NPWP	2.53	1.08	0.82	0.67	0.57	1.85	0.80	0.62	0.51	0.43
		Crossover	1.99	1.34	1.11	0.97	0.87	1.45	0.97	0.81	0.71	0.63
		CF-Perm	1.22	0.92	0.74	0.62	0.54	0.91	0.69	0.56	0.47	0.41
0.1	0.1	Randomization	1.50	0.94	0.72	0.59	0.51	1.11	0.71	0.55	0.46	0.39
		GLMM-C	1.15	0.84	0.66	0.55	0.47	0.84	0.61	0.49	0.41	0.35
		GLMM-CP	1.17	0.85	0.67	0.56	0.48	0.85	0.62	0.50	0.41	0.36
		GEE	0.83	0.66	0.56	0.48	0.42	0.62	0.50	0.42	0.36	0.32
		GEE-FGd5	1.91	1.07	0.78	0.62	0.52	2.32	0.84	0.62	0.46	0.39
		NPWP	2.34	1.09	0.84	0.68	0.58	2.05	0.83	0.64	0.53	0.45
		Crossover	2.03	1.37	1.13	0.99	0.89	1.50	1.01	0.83	0.73	0.66
		CF-Perm	1.25	0.94	0.75	0.63	0.54	0.94	0.71	0.58	0.49	0.42
0.5	0.01	Randomization	2.35	1.58	1.30	1.13	1.01	2.05	1.44	1.22	1.07	0.97
		GLMM-C	1.33	0.99	0.80	0.67	0.58	0.98	0.73	0.58	0.49	0.42
		GLMM-CP	1.36	1.01	0.81	0.68	0.58	1.00	0.74	0.59	0.49	0.43
		GEE	0.95	0.78	0.66	0.57	0.50	0.71	0.56	0.48	0.41	0.36
		GEE-FGd5	1.64	1.09	0.84	0.69	0.59	1.25	0.81	0.62	0.51	0.43
		NPWP	3.76	1.81	1.49	1.29	1.15	3.11	1.61	1.37	1.19	1.07
		Crossover	2.08	1.38	1.14	1.01	0.90	1.52	1.01	0.84	0.74	0.67
		CF-Perm	1.95	1.59	1.37	1.22	1.10	1.70	1.41	1.25	1.11	1.02
0.5	0.1	Randomization	2.38	1.59	1.31	1.14	1.02	2.06	1.45	1.23	1.07	0.97
		GLMM-C	1.33	0.99	0.80	0.67	0.58	0.98	0.73	0.58	0.49	0.42
		GLMM-CP	1.37	1.01	0.81	0.68	0.59	1.02	0.75	0.60	0.50	0.43
		GEE	0.98	0.80	0.67	0.58	0.51	0.73	0.59	0.50	0.43	0.38
		GEE-FGd5	1.69	1.12	0.86	0.71	0.60	1.30	0.85	0.65	0.53	0.45
		NPWP	3.77	1.83	1.50	1.30	1.15	3.10	1.62	1.38	1.19	1.08
		Crossover	2.12	1.42	1.17	1.03	0.93	1.57	1.05	0.88	0.77	0.69
		CF-Perm	1.97	1.60	1.38	1.23	1.11	1.72	1.42	1.26	1.12	1.03

Table S8. [Nonstratified randomization, $(\sigma, \nu) = (0.1, 0.1)$] **Type I error** (%) across 2,000 data sets simulated via **model (6) (random cluster-intervention effects)** as described in Section 3.1.1 with $\theta^* = 0$. We considered different cluster-period size ranges (m_{ij}) and numbers of clusters (N). Methods considered included our proposed individual-level randomization-based approach (Randomization), a logistic mixed model with random cluster effects (GLMM-C), a logistic mixed model with random cluster and cluster-period effects (GLMM-CP), a logistic mixed model with random cluster and cluster-period and cluster-intervention effects (GLMM-CPI), a marginal model fit via a generalized estimating equation (GEE), a small-sample adjusted GEE (GEE-FGd5), a cluster-level nonparametric within-period approach (NPWP), a cluster-level crossover method (Crossover), and a closed-form permutation-based estimator (CF-Perm).

λ	Method	$m_{ij} \sim U\{20, 30\}$					$m_{ij} \sim U\{20, 80\}$				
		N									
		6	8	10	12	14	6	8	10	12	14
0.1	Randomization	5	5	5	5	5	4	6	6	5	6
	GLMM-C	6	5	6	6	6	6	9	9	7	8
	GLMM-CP	6	5	6	6	6	6	8	9	6	8
	GLMM-CPI	5	5	5	6	6	6	8	8	6	7
	GEE	23	18	14	13	10	26	20	17	12	12
	GEE-FGd5	3	4	5	6	6	4	5	6	6	7
	NPWP	5	5	6	5	5	5	6	5	5	6
	Crossover	5	4	5	5	5	4	6	6	5	6
	CF-Perm	8	7	6	6	6	9	7	6	6	8
0.5	Randomization	5	8	8	9	9	7	8	9	10	10
	GLMM-C	11	13	15	16	16	17	20	24	23	26
	GLMM-CP	9	11	12	13	14	13	15	17	16	18
	GLMM-CPI	7	8	8	6	7	10	9	9	8	9
	GEE	22	17	15	13	11	24	19	16	14	13
	GEE-FGd5	4	6	7	7	7	6	7	8	7	8
	NPWP	5	7	8	10	9	7	9	10	12	13
	Crossover	6	6	8	8	9	6	9	10	10	12
	CF-Perm	10	10	9	10	9	12	12	10	11	12
1	Randomization	8	10	10	11	13	9	12	14	12	11
	GLMM-C	21	26	30	34	40	33	40	42	47	51
	GLMM-CP	13	16	17	21	24	15	19	21	23	26
	GLMM-CPI	9	9	8	7	8	9	9	8	8	8
	GEE	22	19	15	14	15	25	20	16	16	16
	GEE-FGd5	6	8	7	7	9	7	8	8	10	10
	NPWP	8	13	15	17	18	9	18	20	21	22
	Crossover	7	11	13	14	16	10	14	17	20	22
	CF-Perm	13	13	12	11	12	15	14	14	12	12

Table S9. [Nonstratified randomization, $(\sigma, \nu) = (0.1, 0.1)$] **CI coverage** (%) across 2,000 data sets simulated via **model (6) (random cluster-intervention effects)** as described in Section 3.1.1 with $\theta^* = 0$. We considered different cluster-period size ranges (m_{ij}) and numbers of clusters (N). Methods considered included our proposed individual-level randomization-based approach (Randomization), a logistic mixed model with random cluster effects (GLMM-C), a logistic mixed model with random cluster and cluster-period effects (GLMM-CP), a logistic mixed model with random cluster and cluster-period and cluster-intervention effects (GLMM-CPI), a marginal model fit via a generalized estimating equation (GEE), a small-sample adjusted GEE (GEE-FGd5), a cluster-level nonparametric within-period approach (NPWP), a cluster-level crossover method (Crossover), and a closed-form permutation-based estimator (CF-Perm).

λ	Method	$m_{ij} \sim U\{20,30\}$					$m_{ij} \sim U\{20,80\}$				
		N									
		6	8	10	12	14	6	8	10	12	14
0.1	Randomization	95	95	95	95	95	95	94	94	95	94
	GLMM-C	94	95	94	94	94	94	91	91	93	92
	GLMM-CP	94	95	94	94	94	94	92	91	94	92
	GLMM-CPI	95	95	95	94	94	94	92	92	94	93
	GEE	77	82	86	87	90	74	80	83	88	88
	GEE-FGd5	97	96	95	94	94	96	95	94	94	93
	NPWP	96	95	94	95	95	96	95	95	95	94
	Crossover	96	96	95	95	95	97	94	94	95	94
	CF-Perm	92	93	94	94	94	91	93	94	94	92
0.5	Randomization	94	92	92	91	92	93	91	91	91	90
	GLMM-C	89	87	85	84	84	83	80	76	77	74
	GLMM-CP	91	89	88	87	86	87	85	83	84	82
	GLMM-CPI	93	92	92	94	93	90	91	91	92	91
	GEE	78	83	85	87	89	76	81	84	86	87
	GEE-FGd5	96	94	93	93	93	94	93	92	93	92
	NPWP	95	93	92	90	91	94	91	90	89	88
	Crossover	96	94	92	92	92	95	91	91	90	89
	CF-Perm	90	90	91	90	91	88	88	90	89	88
1	Randomization	91	90	90	89	87	90	88	86	88	88
	GLMM-C	79	74	70	66	60	67	60	58	53	49
	GLMM-CP	87	84	83	79	76	85	81	79	77	74
	GLMM-CPI	91	91	92	93	92	91	91	92	92	92
	GEE	78	81	85	86	85	75	80	84	84	84
	GEE-FGd5	94	92	93	93	91	93	92	92	90	90
	NPWP	93	88	87	85	83	92	85	82	81	79
	Crossover	94	90	88	87	84	92	88	85	82	80
	CF-Perm	87	87	88	89	88	85	86	86	88	88

Table S10. [Nonstratified randomization, $(\sigma, \nu) = (0.1, 0.1)$] **Average CI width** across 2,000 data sets simulated via **model (6) (random cluster-intervention effects)** as described in Section 3.1.1 with $\theta^* = 0$. We considered different cluster-period size ranges (m_{ij}) and numbers of clusters (N). Methods considered included our proposed individual-level randomization-based approach (Randomization), a logistic mixed model with random cluster effects (GLMM-C), a logistic mixed model with random cluster and cluster-period effects (GLMM-CP), a logistic mixed model with random cluster and cluster-period and cluster-intervention effects (GLMM-CPI), a marginal model fit via a generalized estimating equation (GEE), a small-sample adjusted GEE (GEE-FGd5), a cluster-level nonparametric within-period approach (NPWP), a cluster-level crossover method (Crossover), and a closed-form permutation-based estimator (CF-Perm).

λ	Method	$m_{ij} \sim U\{20, 30\}$					$m_{ij} \sim U\{20, 80\}$				
		N									
		6	8	10	12	14	6	8	10	12	14
0.1	Randomization	1.50	0.95	0.74	0.61	0.52	1.13	0.74	0.57	0.47	0.41
	GLMM-C	1.15	0.84	0.67	0.56	0.48	0.83	0.61	0.49	0.41	0.35
	GLMM-CP	1.17	0.85	0.68	0.56	0.48	0.85	0.63	0.50	0.42	0.36
	GLMM-CPI	1.19	0.87	0.69	0.57	0.49	0.87	0.64	0.52	0.43	0.38
	GEE	0.85	0.68	0.58	0.50	0.44	0.65	0.51	0.44	0.38	0.34
	GEE-FGd5	2.01	1.14	0.80	0.64	0.54	3.58	0.91	0.61	0.48	0.41
	NPWP	2.55	1.11	0.86	0.70	0.60	1.98	0.85	0.66	0.54	0.46
	Crossover	2.05	1.37	1.14	0.99	0.90	1.53	1.02	0.85	0.74	0.66
	CF-Perm	1.26	0.95	0.77	0.65	0.56	0.96	0.74	0.60	0.50	0.44
0.5	Randomization	1.92	1.24	1.00	0.86	0.76	1.57	1.09	0.90	0.78	0.69
	GLMM-C	1.23	0.91	0.74	0.62	0.54	0.92	0.69	0.55	0.46	0.40
	GLMM-CP	1.33	0.98	0.80	0.67	0.58	1.06	0.80	0.65	0.55	0.48
	GLMM-CPI	1.42	1.09	0.92	0.81	0.72	1.14	0.92	0.79	0.70	0.63
	GEE	1.06	0.88	0.79	0.71	0.65	0.85	0.75	0.68	0.62	0.57
	GEE-FGd5	2.02	1.29	1.03	0.88	0.77	1.67	1.09	0.88	0.76	0.68
	NPWP	3.20	1.42	1.14	0.97	0.84	2.71	1.20	0.98	0.83	0.73
	Crossover	2.25	1.47	1.21	1.06	0.94	1.75	1.14	0.93	0.81	0.71
	CF-Perm	1.62	1.25	1.06	0.93	0.82	1.32	1.08	0.92	0.81	0.73
1	Randomization	2.76	1.81	1.49	1.32	1.17	2.52	1.73	1.43	1.27	1.13
	GLMM-C	1.33	0.99	0.80	0.67	0.58	0.98	0.73	0.58	0.49	0.42
	GLMM-CP	1.71	1.28	1.04	0.87	0.75	1.54	1.15	0.93	0.79	0.67
	GLMM-CPI	1.94	1.59	1.40	1.26	1.13	1.72	1.45	1.29	1.17	1.07
	GEE	1.38	1.21	1.11	1.03	0.95	1.23	1.11	1.03	0.97	0.90
	GEE-FGd5	2.34	1.69	1.42	1.25	1.12	2.14	1.58	1.34	1.20	1.07
	NPWP	3.62	2.01	1.66	1.43	1.27	3.60	1.82	1.50	1.32	1.17
	Crossover	2.68	1.72	1.39	1.19	1.05	2.24	1.45	1.15	0.98	0.86
	CF-Perm	2.31	1.86	1.62	1.46	1.32	2.09	1.75	1.53	1.39	1.26

Table S11. [Stratified randomization, $m_{ij} \sim U\{20, 30\}$] **Type I error** (%) across 2,000 data sets simulated via model (7) as described in Section 3.2.1 with $\theta^* = 0$. Methods considered included our proposed individual-level randomization-based approach (Randomization), a logistic mixed model with random cluster effects (GLMM-C), a logistic mixed model with random cluster and cluster-period effects (GLMM-CP), a marginal model fit via a generalized estimating equation (GEE), and a small-sample adjusted GEE (GEE-FGd5). For each method, rows correspond to increasing Y-Z association, controlled by increasing data generation parameter γ^*. Both nonstratified and stratified analyses were carried out.

Method	γ^*	Nonstratified analysis					Stratified analysis				
		N									
		6	8	10	12	14	6	8	10	12	14
Randomization	0	5	5	6	5	4	3	4	5	5	4
	0.2	4	4	4	4	4	3	5	5	5	5
	0.7	1	0	0	0	0	3	5	5	6	5
	1.5	0	0	0	0	0	3	5	5	5	4
GLMM-C	0	6	6	7	5	5	6	6	7	5	5
	0.2	5	5	5	6	5	5	6	6	6	6
	0.7	4	3	4	4	4	6	5	6	6	6
	1.5	4	4	5	4	5	5	6	5	6	6
GLMM-CP	0	6	5	6	5	5	6	6	6	5	5
	0.2	4	5	5	6	5	5	6	5	7	6
	0.7	4	3	3	4	4	5	5	5	6	6
	1.5	3	4	5	4	4	5	6	5	6	5
GEE	0	25	18	15	11	11	30	21	17	13	12
	0.2	21	16	14	11	9	28	21	17	15	13
	0.7	15	12	10	8	8	29	20	16	15	12
	1.5	16	13	10	10	8	27	21	17	14	11
GEE-FGd5	0	3	4	5	5	5	2	4	6	5	5
	0.2	3	4	5	6	4	2	4	5	6	6
	0.7	3	4	5	4	4	2	4	5	6	6
	1.5	4	4	5	5	4	2	4	5	5	5

Table S12. [Stratified randomization, $m_{ij} \sim U\{20,30\}$] **CI coverage** (%) across 2,000 data sets simulated via model (7) as described in Section 3.2.1 with $\theta^* = 0$. Methods considered included our proposed individual-level randomization-based approach (Randomization), a logistic mixed model with random cluster effects (GLMM-C), a logistic mixed model with random cluster and cluster-period effects (GLMM-CP), a marginal model fit via a generalized estimating equation (GEE), and a small-sample adjusted GEE (GEE-FGd5). For each method, rows correspond to increasing Y-Z association, controlled by increasing data generation parameter γ^*. Both nonstratified and stratified analyses were carried out.

Method	γ^*	Nonstratified analysis					Stratified analysis				
		N									
		6	8	10	12	14	6	8	10	12	14
Randomization	0	94	95	95	95	95	94	95	95	95	96
	0.2	96	96	96	96	96	96	95	95	95	95
	0.7	99	100	100	100	100	95	95	95	95	94
	1.5	100	100	100	100	100	95	95	95	95	96
GLMM-C	0	94	94	93	95	95	94	94	93	95	95
	0.2	95	95	95	94	95	95	94	94	94	94
	0.7	96	97	96	96	96	94	95	94	94	94
	1.5	96	96	95	96	95	95	94	95	94	94
GLMM-CP	0	94	95	94	95	95	94	94	94	95	95
	0.2	96	95	95	94	95	95	94	95	93	94
	0.7	96	97	97	96	96	95	95	95	94	94
	1.5	97	96	95	96	96	95	94	95	94	95
GEE	0	75	82	85	89	89	70	79	83	87	88
	0.2	79	84	86	89	91	72	79	83	85	87
	0.7	85	88	90	92	92	71	80	84	85	88
	1.5	84	87	90	90	92	73	79	83	86	89
GEE-FGd5	0	97	96	95	95	95	98	96	94	95	95
	0.2	97	96	95	94	96	98	96	95	94	94
	0.7	97	96	95	96	96	98	96	95	94	94
	1.5	96	96	95	95	96	98	96	95	95	95

Table S13. [Stratified randomization, $m_{ij} \sim U\{20, 30\}$] **Average CI width** across 2,000 data sets simulated via model (7) as described in Section 3.2.1 with $\theta^* = 0$. Methods considered included our proposed individual-level randomization-based approach (Randomization), a logistic mixed model with random cluster effects (GLMM-C), a logistic mixed model with random cluster and cluster-period effects (GLMM-CP), a marginal model fit via a generalized estimating equation (GEE), and a small-sample adjusted GEE (GEE-FGd5). For each method, rows correspond to increasing Y-Z association, controlled by increasing data generation parameter γ^*. Both nonstratified and stratified analyses were carried out.

Method	γ^*	Nonstratified analysis					Stratified analysis				
		N									
		6	8	10	12	14	6	8	10	12	14
Randomization	0	1.42	0.93	0.71	0.58	0.50	1.67	0.96	0.72	0.58	0.50
	0.2	1.48	0.95	0.74	0.61	0.53	1.69	0.93	0.71	0.58	0.49
	0.7	2.08	1.39	1.12	0.95	0.84	1.58	0.88	0.67	0.54	0.46
	1.5	4.73	2.69	2.10	1.79	1.58	1.46	0.81	0.61	0.51	0.43
GLMM-C	0	1.14	0.84	0.66	0.55	0.47	1.12	0.83	0.65	0.54	0.47
	0.2	1.14	0.83	0.66	0.56	0.48	1.11	0.81	0.64	0.53	0.46
	0.7	1.21	0.90	0.73	0.61	0.53	1.06	0.78	0.62	0.52	0.44
	1.5	1.29	0.96	0.77	0.65	0.56	1.08	0.79	0.63	0.52	0.45
GLMM-CP	0	1.15	0.85	0.67	0.55	0.47	1.13	0.83	0.66	0.55	0.47
	0.2	1.15	0.84	0.67	0.56	0.48	1.11	0.81	0.65	0.54	0.46
	0.7	1.24	0.92	0.74	0.62	0.53	1.07	0.79	0.62	0.52	0.45
	1.5	1.32	0.98	0.78	0.65	0.56	1.09	0.80	0.63	0.53	0.45
GEE	0	0.80	0.65	0.55	0.48	0.42	0.73	0.62	0.53	0.46	0.41
	0.2	0.83	0.66	0.57	0.49	0.44	0.73	0.60	0.52	0.46	0.40
	0.7	0.93	0.74	0.62	0.54	0.47	0.72	0.59	0.51	0.44	0.39
	1.5	0.89	0.71	0.59	0.50	0.45	0.73	0.60	0.51	0.45	0.40
GEE-FGd5	0	1.97	1.05	0.77	0.61	0.52	2.11	1.09	0.77	0.61	0.51
	0.2	1.80	1.05	0.77	0.62	0.53	2.02	1.06	0.75	0.60	0.50
	0.7	1.57	1.04	0.79	0.65	0.55	2.20	1.05	0.73	0.58	0.48
	1.5	1.49	1.00	0.76	0.61	0.52	2.13	1.08	0.74	0.58	0.49

Table S14. [Stratified randomization, $m_{ij} \sim U\{20, 80\}$] **Type I error** (%) across 2,000 data sets simulated via model (7) as described in Section 3.2.1 with $\theta^* = 0$. Methods considered included our proposed individual-level randomization-based approach (Randomization), a logistic mixed model with random cluster effects (GLMM-C), a logistic mixed model with random cluster and cluster-period effects (GLMM-CP), a marginal model fit via a generalized estimating equation (GEE), and a small-sample adjusted GEE (GEE-FGd5). For each method, rows correspond to increasing Y-Z association, controlled by increasing data generation parameter γ^*. Both nonstratified and stratified analyses were carried out.

Method	γ^*	Nonstratified analysis					Stratified analysis				
		N									
		6	8	10	12	14	6	8	10	12	14
Randomization	0	4	5	5	5	5	3	5	5	5	5
	0.2	3	4	4	3	3	3	5	5	4	5
	0.7	0	0	0	0	0	3	4	5	5	5
	1.5	0	0	0	0	0	2	5	5	5	5
GLMM-C	0	6	6	6	6	6	6	6	6	6	6
	0.2	5	5	5	4	5	6	7	7	5	7
	0.7	4	4	5	5	4	6	6	6	5	5
	1.5	4	5	5	5	5	5	7	6	7	7
GLMM-CP	0	5	6	6	5	6	6	6	6	6	6
	0.2	5	5	5	4	5	5	6	6	5	7
	0.7	4	3	5	5	4	6	6	6	5	5
	1.5	4	5	5	4	5	5	7	6	7	7
GEE	0	24	19	16	12	12	28	21	18	14	13
	0.2	22	15	13	9	8	28	22	18	12	12
	0.7	17	12	11	9	8	29	21	17	13	12
	1.5	18	14	12	10	10	30	21	17	15	14
GEE-FGd5	0	3	4	5	6	6	2	4	5	6	6
	0.2	3	4	5	4	4	2	5	6	5	6
	0.7	3	4	4	5	4	2	4	5	6	6
	1.5	3	4	5	5	5	2	4	6	6	6

Table S15. [Stratified randomization, $m_{ij} \sim U\{20, 80\}$] **CI coverage** (%) across 2,000 data sets simulated via model (7) as described in Section 3.2.1 with $\theta^* = 0$. Methods considered included our proposed individual-level randomization-based approach (Randomization), a logistic mixed model with random cluster effects (GLMM-C), a logistic mixed model with random cluster and cluster-period effects (GLMM-CP), a marginal model fit via a generalized estimating equation (GEE), and a small-sample adjusted GEE (GEE-FGd5). For each method, rows correspond to increasing Y-Z association, controlled by increasing data generation parameter γ^*. Both nonstratified and stratified analyses were carried out.

Method	γ^*	Nonstratified analysis					Stratified analysis				
		N									
		6	8	10	12	14	6	8	10	12	14
Randomization	0	96	94	95	95	95	95	95	95	95	95
	0.2	97	96	97	97	97	96	94	94	96	95
	0.7	100	100	100	100	100	95	96	95	95	95
	1.5	100	100	100	100	100	95	95	95	95	94
GLMM-C	0	94	94	94	94	94	94	94	94	94	94
	0.2	95	95	95	96	95	94	93	93	95	93
	0.7	96	96	95	95	96	94	94	94	95	95
	1.5	96	95	95	95	95	95	93	94	93	93
GLMM-CP	0	95	94	94	95	94	94	94	94	94	94
	0.2	95	95	95	96	95	95	94	94	95	93
	0.7	96	97	95	95	96	94	94	94	95	95
	1.5	96	95	95	96	95	95	93	94	93	93
GEE	0	76	81	84	88	88	72	79	82	86	87
	0.2	78	85	87	91	92	72	78	82	88	88
	0.7	83	88	89	91	92	71	79	83	87	88
	1.5	82	86	88	90	90	70	79	83	85	86
GEE-FGd5	0	97	96	95	94	94	98	96	95	94	94
	0.2	97	96	95	96	96	98	95	94	95	94
	0.7	97	96	96	95	96	98	96	95	94	94
	1.5	97	96	95	95	95	98	96	94	94	94

Table S16. [Stratified randomization, $m_{ij} \sim U\{20, 80\}$] **Average CI width** across 2,000 data sets simulated via model (7) as described in Section 3.2.1 with $\theta^* = 0$. Methods considered included our proposed individual-level randomization-based approach (Randomization), a logistic mixed model with random cluster effects (GLMM-C), a logistic mixed model with random cluster and cluster-period effects (GLMM-CP), a marginal model fit via a generalized estimating equation (GEE), and a small-sample adjusted GEE (GEE-FGd5). For each method, rows correspond to increasing Y-Z association, controlled by increasing data generation parameter γ^*. Both nonstratified and stratified analyses were carried out.

Method	γ^*	Nonstratified analysis					Stratified analysis				
		N									
		6	8	10	12	14	6	8	10	12	14
Randomization	0	1.06	0.68	0.53	0.44	0.38	1.24	0.70	0.54	0.44	0.38
	0.2	1.10	0.74	0.58	0.49	0.42	1.21	0.69	0.53	0.43	0.37
	0.7	1.84	1.28	1.04	0.90	0.79	1.22	0.70	0.53	0.44	0.37
	1.5	4.49	2.66	2.09	1.78	1.58	1.49	0.79	0.59	0.48	0.41
GLMM-C	0	0.82	0.60	0.48	0.40	0.35	0.81	0.60	0.48	0.40	0.34
	0.2	0.83	0.61	0.49	0.41	0.36	0.80	0.58	0.47	0.39	0.34
	0.7	0.90	0.67	0.54	0.45	0.39	0.77	0.57	0.45	0.38	0.33
	1.5	0.94	0.70	0.56	0.47	0.40	0.78	0.57	0.46	0.38	0.33
GLMM-CP	0	0.83	0.61	0.49	0.41	0.35	0.82	0.60	0.48	0.40	0.34
	0.2	0.84	0.62	0.50	0.42	0.36	0.80	0.59	0.47	0.39	0.34
	0.7	0.92	0.68	0.55	0.46	0.39	0.77	0.57	0.45	0.38	0.33
	1.5	0.96	0.71	0.57	0.47	0.40	0.78	0.57	0.46	0.39	0.33
GEE	0	0.61	0.48	0.41	0.35	0.31	0.60	0.46	0.39	0.34	0.31
	0.2	0.61	0.50	0.42	0.37	0.33	0.65	0.47	0.39	0.33	0.30
	0.7	0.67	0.54	0.46	0.39	0.34	0.54	0.44	0.38	0.33	0.29
	1.5	0.64	0.50	0.42	0.36	0.32	0.57	0.45	0.38	0.33	0.29
GEE-FGd5	0	2.24	0.89	0.58	0.46	0.38	2.87	0.85	0.58	0.45	0.37
	0.2	1.66	0.79	0.57	0.46	0.39	4.17	1.21	0.56	0.44	0.37
	0.7	1.16	0.77	0.59	0.48	0.40	2.96	0.87	0.63	0.43	0.35
	1.5	1.11	0.72	0.55	0.45	0.37	2.79	1.09	0.55	0.43	0.36

Table S17. Within-period correlation (WPC) and inter-period correlation (IPC) on the log-odds scale for nonstratified simulation scenarios in Section 3.1. WPC and IPC were calculated as described in Martin et al. (2016) using logistic approximation approach.

				Control		Intervention	
Model	σ	ν	λ	WPC	IPC	WPC	IPC
(5)	0.1	0.01	·	0.003	0.003	0.003	0.003
	0.1	0.1	·	0.006	0.003	0.006	0.003
	0.5	0.01	·	0.071	0.071	0.071	0.071
	0.5	0.1	·	0.073	0.070	0.073	0.070
(6)	0.1	0.1	0.1	0.006	0.003	0.009	0.006
	0.1	0.1	0.5	0.006	0.003	0.076	0.073
	0.1	0.1	1.0	0.006	0.003	0.237	0.234

Table S18. Within-period correlation (WPC) and inter-period correlation (IPC) on the proportion scale for nonstratified simulation scenarios in Section 3.1. WPC and IPC were calculated as described in Martin et al. (2016) using simulation approach adapted from Goldstein et al. (2002).

						Intervention					
				Control		$\theta^* = 0$		$\theta^* = 0.25$		$\theta^* = 0.5$	
Model	σ	ν	λ	WPC	IPC	WPC	IPC	WPC	IPC	WPC	IPC
(5)	0.1	0.01	·	0.002	0.002	0.002	0.002	0.002	0.002	0.002	0.002
	0.1	0.1	·	0.004	0.002	0.004	0.002	0.004	0.002	0.004	0.002
	0.5	0.01	·	0.046	0.046	0.046	0.046	0.047	0.046	0.052	0.052
	0.5	0.1	·	0.047	0.045	0.047	0.045	0.052	0.050	0.055	0.053
(6)	0.1	0.1	0.1	0.004	0.002	0.005	0.003	0.006	0.004	0.007	0.005
	0.1	0.1	0.5	0.004	0.002	0.047	0.045	0.052	0.050	0.060	0.058
	0.1	0.1	1.0	0.004	0.002	0.150	0.149	0.169	0.168	0.167	0.166

References

Garthwaite, P. H. (1996). Confidence intervals from randomization tests. *Biometrics* **52,** 1387–1393.

Garthwaite, P. H. and Jones, M. C. (2009). A stochastic approximation method and its application to confidence intervals. *Journal of Computational and Graphical Statistics* **18,** 184–200.

Goldstein, H., Browne, W., and Rasbash, J. (2002). Partitioning variation in multilevel models. *Understanding Statistics* **1,** 223–231.

Hughes, J. P., Heagerty, P. J., Xia, F., and Ren, Y. (2020). Robust inference for the stepped wedge design. *Biometrics* **76,** 119–130.

Kennedy-Shaffer, L., de Gruttola, V., and Lipsitch, M. (2020). Novel methods for the analysis of stepped wedge cluster randomized trials. *Statistics in Medicine* **39,** 815–844.

Martin, J., Girling, A., Nirantharakumar, K., Ryan, R., Marshall, T., and Hemming, K. (2016). Intra-cluster and inter-period correlation coefficients for cross-sectional cluster randomised controlled trials for type-2 diabetes in uk primary care. *Trials* **17,** 402.

Rabideau, D. J. and Wang, R. (2020). Randomization-based confidence intervals for cluster randomized trials. *Biostatistics* kxaa007.

Thompson, J. A., Davey, C., Fielding, K., Hargreaves, J. R., and Hayes, R. J. (2018). Robust analysis of stepped wedge trials using cluster-level summaries within periods. *Statistics in Medicine* **37,** 2487–2500.