Research Article

Statistics in Medicine

Received XXXX

(www.interscience.wiley.com) DOI: 10.1002/sim.0000

Supplementary Material to Augmented GEE for improving efficiency and validity of estimation in cluster randomized trials by leveraging cluster- and individual-level covariates

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

Appendix A: $\hat{\beta}$ Solutions for Standard and Augmented Logistic GEE in cluster randomized designs

Let Y_{ij} denote the response (0 or 1) for the j_{th} individual in the i_{th} cluster. $\mathbf{Y_i} = (Y_{i1},\ Y_{i2}, ..., Y_{in_i})^{\mathrm{T}}$, where n_i is the number of subjects within the i_{th} cluster. A typical model for binary data is $E(Y_{ij}|A_i) = g(A_i;\beta) = g(\beta_0 + \beta_1 A_i)$, where g is the inverse logit link function. The standard GEE for the marginal treatment effect are given by

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

where bold $\mathbf{g}(A_i;\beta)$ denotes the n_i-dimensional link function for the outcome vector $\mathbf{Y_i}$, $\mathbf{D_i}$ is the $n_i \times p$ matrix defined by $\frac{\partial \mathbf{g}(A_i;\beta)}{\partial \beta^{\mathrm{T}}}$, and $\mathbf{V_i}$ is a $n_i \times n_i$ working covariance matrix for $\mathbf{Y_i}$.

$\mathbf{D_i}$ is composed of the n_i-dimensional columns $\vec{D}_{i \cdot 0} = \frac{\partial \mathbf{g}(A_i;\beta)}{\partial \beta_0}$ and $\vec{D}_{i \cdot 1} = \frac{\partial \mathbf{g}(A_i;\beta)}{\partial \beta_1}$. Because of the cluster-randomized design, $\vec{D}_{i \cdot 0}$ and $\vec{D}_{i \cdot 1}$ are vectors of the form $\vec{D}_{i \cdot p} = (D_{i_p}, D_{i_p}, \ldots, D_{i_p})^{\mathrm{T}}$ for $p = 0, 1$ (intercept and treatment effect), with

$$D_{i_0} = \frac{\partial g(A_i;\beta)}{\partial \beta_0} = \frac{\exp(\beta_0 + \beta_1 A_i)}{(1 + \exp(\beta_0 + \beta_1 A_i))} \left(1 - \frac{\exp(\beta_0 + \beta_1 A_i)}{(1 + \exp(\beta_0 + \beta_1 A_i))}\right) = \pi(A_i)\{1 - \pi(A_i)\}$$
$$D_{i_1} = \frac{\partial g(A_i;\beta)}{\partial \beta_1} = \frac{\exp(\beta_0 + \beta_1 A_i) A_i}{(1 + \exp(\beta_0 + \beta_1 A_i))} \left(1 - \frac{\exp(\beta_0 + \beta_1 A_i) A_i}{(1 + \exp(\beta_0 + \beta_1 A_i))}\right) = \pi(A_i)\{1 - \pi(A_i)\} A_i,$$

where $\pi(A_i) = E(Y_{ij}|A_i)$. We recall that $\mathbf{D_i}$ is evaluated using an initial estimator $\hat{\beta}_{init}$, usually obtained from standard logistic regression that does not account for clustering.

The inverse working covariance matrix $\mathbf{V_i}^{-1}$ can be broken down into its columns and scalar elements. Let $\mathbf{V_i}^{-1} = \begin{bmatrix} \vec{V}_{i \cdot 1}^{-1} & \vec{V}_{i \cdot 2}^{-1} & \cdots & \vec{V}_{i \cdot n_i}^{-1} \end{bmatrix}$, where $\vec{V}_{i \cdot j}^{-1}$ is the j_{th} column of $\mathbf{V_i}^{-1}$, and $V_{i_{q,j}}^{-1}$ represents the scalar element in the q_{th} row, j_{th} column. Using this construction, after some matrix algebra, a closed form solution for β under a cluster randomized

Department of Biostatistics, Harvard School of Public Health, Boston, MA, USA
** Correspondence to: Alisa Stephens, Department of Biostatisics, Harvard School of Public Health, Boston, MA, USA. Email: astephen@hsph.harvard.edu*

design is given by

$$\hat{\beta}_0 = logit\left(\left[\sum_{i=1}^{m}\left\{I(A_i=0)D_{i_0}\sum_{q,j\leq n_i}V_{i_{q,j}}^{-1}\right\}\right]^{-1}\left[\sum_{i=1}^{m}\left\{(\vec{D}_{i.0}-\vec{D}_{i.1})^{\mathrm{T}}\sum_{j=1}^{n_i}(\vec{V}_{i.j}^{-1}Y_{ij})\right\}\right]\right)$$
$$\hat{\beta}_1 = logit\left(\left[\sum_{i=1}^{m}\left\{I(A_i=1)D_{i_1}\sum_{q,j\leq n_i}V_{i_{q,j}}^{-1}\right\}\right]^{-1}\left[\sum_{i=1}^{m}\left\{\vec{D}_{i.1}^{\mathrm{T}}\sum_{j=1}^{n_i}(\vec{V}_{i.j}^{-1}Y_{ij})\right\}\right]\right)-\hat{\beta}_0$$

This solution can be simplified using the working covariance structure. Under exchangeable correlation, $V_{i_{q,q}} = \phi$ and $V_{i_{q,j}} = \rho$ for $q \neq j$. We note working independence as a special case with off-diagonal elements $\rho = 0$. Proceeding, let ϕ^{-1} and ρ^{-1} denote the diagonal and off-diagonal elements of $\mathbf{V_i}^{-1}$, respectively. The above simplifies to

$$\hat{\beta}_0 = logit\Bigg(\left[\sum_{i=1}^{m}D_{i_0}I(A_i=0)\left\{n_i\phi^{-1}+n_i(n_i-1)\rho^{-1}\right\}\right]^{-1}\times$$
$$\left[\sum_{i=1}^{m}\left\{(D_{i_0}-D_{i_1})\{(n_i-1)\rho_i^{-1}+\phi^{-1}\}\sum_{j=1}^{n_i}Y_{ij}\right\}\right]\Bigg)$$
$$\hat{\beta}_1 = logit\Bigg(\left[\sum_{i=1}^{m}I(A_i=1)D_{i_1}\{n_i\phi^{-1}+n_i(n_i-1)\rho_1^{-1}\}\right]^{-1}\times$$
$$\left[\sum_{i=1}^{m}\left\{D_{i_1}\{(n_i-1)\rho^{-1}+\phi^{-1}\}\sum_{j=1}^{n_i}Y_{ij}\right\}\right]\Bigg)-\hat{\beta}_0$$

In the case of the augmented GEE, we estimate $\hat{\beta}$ using the augmented estimating equations

$$\sum_{i=1}^{m}\left[\mathbf{D_i}^{\mathrm{T}}\mathbf{V_i}^{-1}\{\mathbf{Y_i}-\mathbf{g}(A_i;\beta)\}-(A_i-\pi)\hat{\gamma}(\mathbf{X_i})\right]=\mathbf{0},$$

where $\hat{\gamma}(X_i) = \left[\mathbf{D_i}(1)^{\mathrm{T}}\mathbf{V_i}(1)^{-1}\{f_1(\mathbf{X_i};\hat{\eta}_1)-\mathbf{g}(1;\beta)\}-\mathbf{D_i}(0)^{\mathrm{T}}\mathbf{V_i}(0)^{-1}\{f_0(\mathbf{X_i};\hat{\eta}_0)-\mathbf{g}(0;\beta)\}\right]$.

Above, we take $\mathbf{D_i}(k) = \frac{\partial\mathbf{g}(k;\beta)}{\partial\beta^{\mathrm{T}}}$, $\mathbf{V_i}(k) = \mathbf{V_i}$ evaluated under treatment k, and $f_k(\mathbf{X_i};\hat{\eta}_k) = \hat{E}[\mathbf{Y_i}|A_i=k,\mathbf{X_i}]$ for $k = 0, 1$. Vectors $\vec{D}_{i.p}(k)$, $\vec{V}_{i.j}^{-1}(k)$, and scalars $D_{i_p}(k)$, $V_{q,j}^{-1}(k)$ are defined similarly as above, evaluated under treatment k. For brevity, we write $\mathbf{F}_{k_i}$ below, where $\mathbf{F}_{k_i}$=$f_k(\mathbf{X_i};\hat{\eta}_k)$. Solutions for $\hat{\beta}_0$ and $\hat{\beta}_1$ are given by

$$\hat{\beta}_0 = logit\Bigg(\left[(1-\pi)\sum_{i=1}^{m}\left\{D_{i_0}(0)\sum_{q,j\leq n_i}V_{i_{q,j}}^{-1}(0)\right\}\right]^{-1}\times$$
$$\sum_{i=1}^{m}\left[\sum_{j=1}^{n_i}\left\{(\vec{D}_{i.0}-\vec{D}_{i.1})^{\mathrm{T}}V_{i.j}^{-1}Y_{ij}-(A_i-\pi)\left(-\vec{D}_{i.0}(0)\vec{V}_{i.j}^{-1}(0)F_{0_{ij}}\right)\right\}\right]\Bigg)$$
$$\hat{\beta}_1 = logit\Bigg(\left[\pi\sum_{i=1}^{m}\left\{D_{i_1}(1)\sum_{q,j\leq n_i}V_{i_{q,j}}^{-1}(1)\right\}\right]^{-1}\times$$
$$\sum_{i=1}^{m}\left[\sum_{j=1}^{n_i}\left\{(\vec{D}_{i.1}^{\mathrm{T}}V_{i.j}^{-1}Y_{ij})-(A_i-\pi)\left(\vec{D}_{i.1}(1)^{\mathrm{T}}\vec{V}_{i.j}^{-1}(1)F_{1_{ij}}\right)\right\}\right]\Bigg)-\hat{\beta}_0,$$

The simplified expression in case of exchangeable structure is

$$\hat{\beta}_0 = logit\left(\left[(1-\pi)\sum_{i=1}^{m} D_{i_0}(0)\{n_i\phi^{-1} + n_i(n_i-1)\rho_0^{-1}\}\right]^{-1} \times \right.$$
$$\left.\sum_{i=1}^{m}\left[(D_{i_0} - D_{i_1})\{(n_i-1)\rho_i^{-1} + \phi_i^{-1}\}\sum_{j=1}^{n_i} Y_{ij} - (A_i - \pi)\left\{-D_{i_1}(0)\{(n_i-1)\rho_0^{-1} + \phi_0^{-1}\}\sum_{j=1}^{n_i} F_{0_{ij}}\right\}\right]\right)$$
$$\hat{\beta}_1 = logit\left(\left[\pi\sum_{i=1}^{m}\left\{D_{i_1}(1)\{n_i\phi_1^{-1} + n_i(n_i-1)\rho_1^{-1}\}\right\}\right]^{-1} \times \right.$$
$$\left.\sum_{i=1}^{m}\left[D_{i_1}\{(n_i-1)\rho_i^{-1} + \phi_i^{-1}\}\sum_{j=1}^{n_i} Y_{ij} - (A_i - \pi)\left\{D_{i_1}(1)\{(n_i-1)\rho_1^{-1} + \phi_1^{-1}\}\sum_{j=1}^{n_i} F_{1_{ij}}\right\}\right]\right) - \hat{\beta}_0,$$

where we maintian the index i in $D_{i_p}(k)$ to be consistent with the unsimplified expressions above, in which the index i on $\vec{D}_{i.p}(k)$ is retained to be mindful of varying cluster size. The quantity $D_{i_p}(k)$, however, is a fixed function of $E(Y_{ij}|A_i = k)$.

Appendix B: Variance Estimators

Let $\mathbf{Y_i}$ be the n_i-dimensional response vector, A_i the scalar treatment variable, and $\mathbf{X_i}$ a collection of baseline covariates potentially at the cluster and individual level. The model $E(\mathbf{Y_i}|A_i) = \mathbf{g}(A_i;\beta)$ is assumed, and the estimator $\hat{\beta}$ is obtained by solving the augmented estimating equations detailed in Section 2. Recall that $\mathbf{V_i}$ is a working covariance matrix as typically used in GEE for estimating coefficients in restricted moment models and π=$P(A_i = 1)$. Formulas for the variance estimators discussed in section 3 are presented below. The asymptotic variability of $\hat{\beta}_{aug}$ is shown to be $var(\hat{\beta}_{aug}) = \Gamma^{-1}\Delta\Gamma^{-1^{\mathrm{T}}}$, where $\Gamma = E\left[\frac{\partial\psi_{i_{opt}}(\mathbf{Y},A,\mathbf{X};\beta)}{\partial\beta^{\mathrm{T}}}\right]$, and $\Delta = E\left[\psi_{i_{opt}}(\mathbf{Y},A,\mathbf{X};\beta)^{\bigotimes 2}\right]$, with $U^{\bigotimes 2} = UU^{\mathrm{T}}$. In each of the below, $\hat{\Gamma} = m^{-1}\sum_i \mathbf{D_i}^{\mathrm{T}}\mathbf{V_i}^{-1}\mathbf{D_i}$. The four variance estimators considered are:

1. $v\hat{a}r_1(\hat{\beta}_{aug}) = \hat{\Gamma}^{-1}\hat{\Delta}\hat{\Gamma}^{-1^{\mathrm{T}}}$, where $\hat{\Delta} = m^{-1}\sum_{i=1}^{m}\hat{\psi}_{i_{opt}}^{\bigotimes 2}$, and

$$\hat{\psi}_{i_{opt}}(\mathbf{Y},A,\mathbf{X};\beta) = \mathbf{D_i}^{\mathrm{T}}\mathbf{V_i}^{-1}\{\mathbf{Y_i} - \mathbf{g}(A_i;\hat{\beta}_{aug})\}$$
$$-(A_i-\pi)\big[\mathbf{D_i}(1)^{\mathrm{T}}\mathbf{V_i}(1)^{-1}\{f_1(\mathbf{X_i};\hat{\eta}_1) - \mathbf{g}(1;\hat{\beta}_{aug})\} - \mathbf{D_i}(0)^{\mathrm{T}}\mathbf{V_i}(0)^{-1}\{f_0(\mathbf{X_i};\hat{\eta}_0) - \mathbf{g}(0;\hat{\beta}_{aug})\}\big]$$

2. $v\hat{a}r_2(\hat{\beta}_{aug}) = \hat{\Gamma}^{-1}\hat{\Delta}^*\hat{\Gamma}^{-1^{\mathrm{T}}}$, where $\hat{\Delta}^* = m^{-1}\sum_{i=1}^{m}(\mathbf{H_i}\hat{\psi}_i)^{\bigotimes 2}$, and $\mathbf{H_i}$ is a diagonal matrix with $H_{i_{jj}} = \left[1 - min\{q, (\frac{\partial\psi_i(\mathbf{Y_i},A,\mathbf{X_i};\beta)}{\partial\beta^{\mathrm{T}}} \times \hat{\Gamma})_{jj}\}\right]^{-1/2}$ [1], and $\hat{\psi}_i = \hat{\psi}_i(\mathbf{Y},A,\mathbf{X};\beta)$ is as defined in 1).
3. $v\hat{a}r_3(\hat{\beta}_{aug}) = \hat{\Gamma}^{-1}\tilde{\Delta}\hat{\Gamma}^{-1^{\mathrm{T}}}$, where $\tilde{\Delta} = m^{-1}\sum_{i=1}^{m}\tilde{\psi}_{i_{opt}}^{\bigotimes 2}$, and

$$\tilde{\psi}_{i_{opt}} = \mathbf{D_i}^{\mathrm{T}}\mathbf{V_i}^{-1}\{\mathbf{Y_i} - \mathbf{g}(A_i;\hat{\beta}_{aug})\}$$
$$-(A_i-\pi)\left[\mathbf{D_i}(1)^{\mathrm{T}}\mathbf{V_i}(1)^{-1}\{f_1(\mathbf{X_i};\hat{\eta}_1) - \mathbf{g}(1;\hat{\beta}_{aug})\} - \mathbf{D_i}(0)^{\mathrm{T}}\mathbf{V_i}(0)^{-1}\{f_0(\mathbf{X_i};\hat{\eta}_0) - \mathbf{g}(0;\hat{\beta}_{aug})\}\right]$$
$$-(A_i-\pi)\left[\mathbf{D_i}(1)^{\mathrm{T}}\mathbf{V_i}(1)^{-1}\{f_1'(\mathbf{X_i};\hat{\eta}_1)\}\hat{\zeta}_1(\mathbf{Y_i},\mathbf{X_i};\hat{\eta}_1) - \mathbf{D_i}(0)^{\mathrm{T}}\mathbf{V_i}(0)^{-1}\{f_0'(\mathbf{X_i};\hat{\eta}_0)\}\hat{\zeta}_0(\mathbf{Y_i},\mathbf{X_i};\hat{\eta}_0)\right].$$

$\hat{\zeta}_k(\mathbf{Y_i},\mathbf{X_i};\hat{\eta})$ is the first order approximation of the term $(\hat{\eta}_k - \eta^*)$ that results from estimation of η_k in $E(\mathbf{Y_i}|\mathbf{X_i},A_i)$. If $E(\mathbf{Y_i}|\mathbf{X_i},A_i)$ is estimated by linear regression, $\hat{\zeta}_k(\mathbf{Y_i},\mathbf{X_i};\hat{\eta}) = \left(\sum_{i=1}^{m}\mathbf{X_i}^{\mathrm{T}}\mathbf{X_i}\right)^{-1}\sum_{i=1}^{m}\mathbf{X_i}^{\mathrm{T}}(\mathbf{Y_i} -$

$\mathbf{X_i}\hat{\eta}_k$). For nonlinear models, in which $E(\mathbf{Y_i}|\mathbf{X_i}, A_i) = \mu(\mathbf{X_i}; \eta_k)$, $(\hat{\eta}_k - \eta_k^*)$ may be approximated by $\left(\sum_{i=1}^{m} \mathbf{F_i}^{\mathrm{T}} \mathbf{W_i}^{-1} \mathbf{F_i}\right)^{-1} \sum_{i=1}^{m} \mathbf{F_i}^{\mathrm{T}} \mathbf{W_i}^{-1} (\mathbf{Y_i} - \mu(\mathbf{X_i}; \hat{\eta}_k))$, where $\mathbf{F_i} = \frac{\partial \mu_k(\mathbf{X_i};\eta)}{\partial \eta}\Big|_{\eta_k = \hat{\eta}_k}$, and $\mathbf{W_i}$ is a diagonal matrix with $W_{i_{jj}}$=$\phi_\mu \nu(\mu)$, following from generalized linear model theory. The function $\nu(\mu)$ denotes the variance function and ϕ_μ the dispersion parameter. We include the subscript μ in ϕ_μ to distinguish from ϕ involved in characterizing $V(Y_{ij}|A_i)$ in the main text.

4. $v\hat{a}r_4(\hat{\beta}_{aug}) = \hat{\Gamma}^{-1} \tilde{\Delta}^* \hat{\Gamma}^{-1^{\mathrm{T}}}$, where $\tilde{\Delta}^* = m^{-1} \sum_i (\mathbf{H_i} \tilde{\psi}_i)^{\bigotimes 2}$, with $\tilde{\psi}$ and $\mathbf{H_i}$ are as defined above.
In 2) and 4), the lower bound q is typically set to 0.75 to prevent gross inflation [1].

Appendix C: Additional Simulations

Table 1. Standard vs. Augmented GEE, Binary Outcome: 250 clusters, low and high association, $\rho = 0.05$. *Std: unaugmented. Correlation is exchangeable for all estimators. C,F,O,W: augmentation with 'Correct','Forward' selected, 'One-variable', or 'Wrong' model. ML, OLS: augmentation fit with maximum likelihood or ordinary least squares. SE: average unadjusted sandwich. MC RE: square of the Monte Carlo SE of the Std(Exch) estimator divided by the Monte Carlo SE for the indicated estimator. Coverage: coverage based on unadjusted sandwich SE.*

	Estimator	$\hat{\beta}_1$	Bias	SE	MC SE	MC RE	Coverage
m=250, low	Std	-0.3036	0.0077	0.0739	0.0778	1.0000	0.936
	C - ML	-0.3029	0.0069	0.0705	0.0744	1.0951	0.935
	C - OLS	-0.3032	0.0072	0.0710	0.0750	1.0778	0.937
	F - ML	-0.3023	0.0064	0.0701	0.0753	1.0683	0.935
	F - OLS	-0.3026	0.0066	0.0703	0.0757	1.0571	0.937
	O - ML	-0.3033	0.0073	0.0717	0.0752	1.0704	0.937
	O - OLS	-0.3033	0.0073	0.0717	0.0754	1.0658	0.935
	W - ML	-0.3034	0.0075	0.0728	0.0768	1.0271	0.938
	W - OLS	-0.3035	0.0075	0.0728	0.0768	1.0258	0.938
m=250, high	Std	1.1310	0.0052	0.0567	0.0576	1.0000	0.943
	C - ML	1.1314	0.0047	0.0489	0.0497	1.3429	0.940
	C - OLS	1.1313	0.0049	0.0496	0.0505	1.2989	0.943
	F - ML	1.1314	0.0047	0.0485	0.0502	1.3156	0.937
	F - OLS	1.1314	0.0048	0.0491	0.0510	1.2719	0.941
	O - ML	1.1312	0.0050	0.0501	0.0509	1.2799	0.934
	O - OLS	1.1313	0.0048	0.0506	0.0512	1.2651	0.941
	W - ML	1.1310	0.0051	0.0531	0.0542	1.1301	0.938
	W - OLS	1.1310	0.0051	0.0533	0.0542	1.1266	0.937

Table 2. Standard vs. Augmented GEE, Binary Outcome: 250 clusters, low and high association, $\rho = 0.05$. *Std: unaugmented. Correlation is exchangeable for all estimators. C,F,O,W: augmentation with 'Correct','Forward' selected, 'One-variable', or 'Wrong' model. ML, OLS: augmentation fit with maximum likelihood or ordinary least squares. SE: average unadjusted sandwich. MC RE: square of the Monte Carlo SE of the Std(Exch) estimator divided by the Monte Carlo SE for the indicated estimator. Coverage: coverage based on unadjusted sandwich SE.*

	Estimator	$\hat{\beta}_1$	Bias	SE	MC SE	MC RE	Coverage
m=250, low	Std	-0.2299	0.0135	0.1164	0.1190	1.0000	0.938
	C - ML	-0.2290	0.0126	0.1144	0.1173	1.0293	0.936
	C - OLS	-0.2293	0.0129	0.1146	0.1175	1.0256	0.935
	F - ML	-0.2276	0.0112	0.1135	0.1185	1.0077	0.932
	F - OLS	-0.2280	0.0116	0.1135	0.1187	1.0041	0.933
	O - ML	-0.2294	0.0130	0.1150	0.1175	1.0253	0.935
	O - OLS	-0.2295	0.0130	0.1150	0.1176	1.0234	0.935
	W - ML	-0.2296	0.0131	0.1155	0.1188	1.0020	0.931
	W - OLS	-0.2297	0.0132	0.1155	0.1188	1.0021	0.932
m=250, high	Std	1.0429	0.0072	0.0883	0.0887	1.0000	0.944
	C - ML	1.0436	0.0065	0.0835	0.0848	1.0936	0.949
	C - OLS	1.0435	0.0066	0.0839	0.0851	1.0871	0.948
	F - ML	1.0442	0.0059	0.0828	0.0858	1.0694	0.938
	F - OLS	1.0444	0.0058	0.0831	0.0860	1.0643	0.941
	O - ML	1.0433	0.0068	0.0842	0.0851	1.0863	0.951
	O - OLS	1.0435	0.0067	0.0844	0.0851	1.0861	0.949
	W - ML	1.0431	0.0070	0.0859	0.0869	1.0409	0.951
	W - OLS	1.0431	0.0070	0.0860	0.0869	1.0406	0.950

References

1. Fay MP, Graubard BI. Small-sample adjustments for wald-type tests using sandwich estimators. *Biometrics* 2001; **57**:1198–1206. DOI: 10.1111/j.0006-341X.2001.01198.x.