BIOMETRICS 64, 1–7 DOI: 10.1111/j.1541-0420.2005.00454.x
December 2015

Web-Supplementary Material

Accounting for interactions and complex inter-subject dependency in estimating treatment effect in cluster randomized trials with missing outcomes

Melanie Prague[1,*], Rui Wang[1,2], Alisa Stephens[3], Eric Tchetgen Tchetgen[4] and Victor DeGruttola[1]

[1] Department of Biostatistics, Harvard T.H. Chan School of Public Health, Boston, MA, U.S.A.
[2] Division of Sleep and Circadian Disorders, Departments of Medicine and Neurology,
Brigham and Women's Hospital, Boston, MA, U.S.A.
[3] Center for Clinical Epidemiology and Biostatistics, Perelman School of Medicine,
University of Pennsylvania, Philadelphia, PA, U.S.A.
[4] Department of Epidemiology, Harvard T.H. Chan School of Public Health, Boston, MA, U.S.A.
**email:* mprague@hsph.harvard.edu

A - The causal definition of covariate interference

Another way to look at covariate interference is to consider the DAG and figure out what are the assumption needed to estimate the marginal effect of treatment (path $A \to Y$ in Figure 1).

[Figure 1 about here.]

We focus on cluster i composed of two subjects (1,2); X is an interfering covariate. The paths $X_2 \to Y_1$ and $X_1 \to Y_2$ represent the presence of covariate interference at the outcome level and $X_2 \to R_1$ and $X_1 \to R_2$ the presence of covariate interference at the level of missingness process. When we analyze the dataset with missing data, subjects with $R_1 = 0$ have Y_1 unobserved. In other words, values of Y_1 are conditioned by R_1. Thus, all paths flowing through R_1 are 'unblocked'. Thus, in the presence of covariate interference for the outcome and the missing data process, two paths are unblocked compared to an analysis without missing data: $A \to R_1 \leftarrow X_1 \to Y_1$ and $A \to R_1 \leftarrow X_2 \to Y_1$. This is due to collider stratification bias, also called bias due to conditioning on a collider (Pearl et al., 2009). In order to ensure unbiasedness of the estimation with GEE approaches, solution is to adjust on X_1 (individual covariates) and X_2 (interfering covariates). If covariate interference are only on the outcome generation process, no additional path may create a bias in estimation of the marginal effect of treatment. If covariate interference are only on the missing data generation process, $A \to R_1 \leftarrow X_1 \to Y_1$ is unblocked, solution is to adjust on X_1 (individual covariates) but there is no need to adjust for X_2 (interfering covariates).

B - Derivation of the estimating Equation (EE) for the DR estimator

In this section we present the derivation to obtain the EE for the DR estimator. The idea is to project the IPW EE $\psi_i(\boldsymbol{Y}_i, \boldsymbol{R}_i, A_i, \boldsymbol{\beta}, \boldsymbol{\eta}_W)$ onto the span of scores corresponding to all smooth parametric models for the missing data process and the treat- ment assignment mechanism given covariates:

$$
\begin{aligned}
0 \;=\; \sum_{i}^{M} \Bigg[& \boldsymbol{\psi}_i(\boldsymbol{Y}_i, \boldsymbol{R}_i, A_i, \boldsymbol{\beta}, \boldsymbol{\eta}_W) \\
& \underbrace{-E(\boldsymbol{\psi}_i(\boldsymbol{Y}_i, \boldsymbol{R}_i, A_i, \boldsymbol{\beta}, \boldsymbol{\eta}_W)|\boldsymbol{X}_i, \boldsymbol{R}_i, A_i) + E(\boldsymbol{\psi}_i(\boldsymbol{Y}_i, \boldsymbol{R}_i, A_i, \boldsymbol{\beta}, \boldsymbol{\eta}_W)|\boldsymbol{X}_i, A_i)}_{\text{for missing data mechanism}} \\
& \underbrace{-E(\boldsymbol{\psi}_i(\boldsymbol{Y}_i, \boldsymbol{R}_i, A_i, \boldsymbol{\beta}, \boldsymbol{\eta}_W)|\boldsymbol{X}_i, A_i) + E(\boldsymbol{\psi}_i(\boldsymbol{Y}_i, \boldsymbol{R}_i, A_i, \boldsymbol{\beta}, \boldsymbol{\eta}_W)|\boldsymbol{X}_i)}_{\text{for treatment assignment mechanism}} \Bigg].
\end{aligned}
$$

This gives the following derivations:

$$
\begin{aligned}
0 \;=\; \sum_{i=1}^{M} \Bigg[& \boldsymbol{D}_i^T \boldsymbol{V}_i^{-1} \boldsymbol{W}_i(\boldsymbol{X}_i, A_i, \boldsymbol{\eta}_W) \left(\boldsymbol{Y}_i - \boldsymbol{\mu}_i(\boldsymbol{\beta}, A_i)\right) \\
& + \Big\{ -\boldsymbol{D}_i^T \boldsymbol{V}_i^{-1} \boldsymbol{W}_i(\boldsymbol{X}_i, A_i, \boldsymbol{\eta}_W) \left(\boldsymbol{B}_i(\boldsymbol{X}_i, A_i, \boldsymbol{\eta}_B) - \boldsymbol{\mu}_i(\boldsymbol{\beta}, A_i)\right) + \boldsymbol{D}_i^T \boldsymbol{V}_i^{-1} \left(\boldsymbol{B}_i(\boldsymbol{X}_i, A_i, \boldsymbol{\eta}_B) - \boldsymbol{\mu}_i(\boldsymbol{\beta}, A_i)\right) \Big\} \\
& + \Bigg\{ -\boldsymbol{D}_i^T \boldsymbol{V}_i^{-1} \left(\boldsymbol{B}_i(\boldsymbol{X}_i, A_i, \boldsymbol{\eta}_B) - \boldsymbol{\mu}_i(\boldsymbol{\beta}, A_i)\right) + \sum_{a=0,1} p^a (1-p)^{1-a} \boldsymbol{D}_i^T \boldsymbol{V}_i^{-1} \Big(\boldsymbol{B}_i(\boldsymbol{X}_i, A_i = a, \boldsymbol{\eta}_B) - \boldsymbol{\mu}_i(\boldsymbol{\beta}, A_i = a) \Big) \Bigg\} \Bigg].
\end{aligned} \tag{1}
$$

We cancel the 3rd term with the 4th term and combine the 1st and 2nd terms to get the EE for the DR:

$$
\begin{aligned}
0 \;=\; & \sum_{i=1}^{M} \Bigg[\boldsymbol{D}_i^T \boldsymbol{V}_i^{-1} \boldsymbol{W}_i(\boldsymbol{X}_i, A_i, \boldsymbol{\eta}_W) \left(\boldsymbol{Y}_i - \boldsymbol{B}_i(\boldsymbol{X}_i, A_i, \boldsymbol{\eta}_B)\right) \\
& \quad + \sum_{a=0,1} p^a (1-p)^{1-a} \boldsymbol{D}_i^T \boldsymbol{V}_i^{-1} \Big(\boldsymbol{B}_i(\boldsymbol{X}_i, A_i = a, \boldsymbol{\eta}_B) - \boldsymbol{\mu}_i(\boldsymbol{\beta}, A_i = a) \Big) \Bigg], \qquad (2) \\
=\; & \sum_{i=1}^{M} \boldsymbol{\Phi}_i(\boldsymbol{Y}_i, \boldsymbol{R}_i, A_i, \boldsymbol{X}_i, \boldsymbol{\beta}, \boldsymbol{\eta}_W, \boldsymbol{\eta}_B).
\end{aligned}
$$

C - Proof of CAN for the DR estimator

1- Correct specification of the OM or the PS

The consistency can be considered by evaluating (a) defined in Equation 3 and checking that (a)=0 when the OM or the PS is correctly specified. For notation simplicity, we omit the nuisance parameters η_W and η_B while writing this demonstration.

$$
\begin{aligned}
(a) \;=\; E\Bigg[& \boldsymbol{D}_i^T \boldsymbol{V}_i^{-1} \boldsymbol{W}_i(\boldsymbol{X}_i, A_i, \eta_W) \left(\boldsymbol{Y}_i - \boldsymbol{B}_i(\boldsymbol{X}_i, A_i, \boldsymbol{\eta}_B)\right) \\
& + \sum_{a=0,1} p^a (1-p)^{1-a} \boldsymbol{D}_i^T (A_i = a) \boldsymbol{V}_i^{-1} \Big(\boldsymbol{B}_i(\boldsymbol{X}_i, A_i = a, \boldsymbol{\eta}_B) - \boldsymbol{\mu}_i(\boldsymbol{\beta}, A_i = a) \Big) \Bigg].
\end{aligned} \tag{3}
$$

The OM is correctly specified

When OM is correctly specified we have $B_{ij}(\boldsymbol{X}_i, A_i, \boldsymbol{\eta}_B) = E(\boldsymbol{Y}_{ij}|A_i, \boldsymbol{X}_i)$. Let's denote (om1) and (om2) the first and second terms of (a) in Equation 3.

$$\begin{aligned}
(om1) &= E\left[E\left[\boldsymbol{D}_i^T\boldsymbol{V}_i^{-1}\boldsymbol{W}_i(\boldsymbol{X}_i, A_i, \eta_W)\left(\boldsymbol{Y}_i - \boldsymbol{B}_i(\boldsymbol{X}_i, A_i, \boldsymbol{\eta}_B)\right)|R_{ij}, \boldsymbol{X}_i, A_i\right]\right]\\
&= E\left[\boldsymbol{D}_i^T\boldsymbol{V}_i^{-1}\boldsymbol{W}_i(\boldsymbol{X}_i, A_i, \eta_W)E\left[\left(\boldsymbol{Y}_i - \boldsymbol{B}_i(\boldsymbol{X}_i, A_i, \boldsymbol{\eta}_B)\right)|R_{ij}, \boldsymbol{X}_i, A_i\right]\right]\\
&= E\left[\boldsymbol{D}_i^T\boldsymbol{V}_i^{-1}\boldsymbol{W}_i(\boldsymbol{X}_i, A_i, \eta_W\left(E\left[\boldsymbol{Y}_i|R_{ij}, \boldsymbol{X}_i, A_i\right] - \boldsymbol{B}_i(\boldsymbol{X}_i, A_i, \boldsymbol{\eta}_B)\right)\right]\\
&= E\left[\boldsymbol{D}_i^T\boldsymbol{V}_i^{-1}\boldsymbol{W}_i(\boldsymbol{X}_i, A_i, \eta_W)\left(E\left[\boldsymbol{Y}_i|\boldsymbol{X}_i, A_i\right] - \boldsymbol{B}_i(\boldsymbol{X}_i, A_i, \boldsymbol{\eta}_B)\right)\right]\\
&= 0
\end{aligned}$$

$$\begin{aligned}
(om2) &= E\left[\sum_{a=0,1} p^a(1-p)^{1-a}\boldsymbol{D}_i^T(A_i = a)\boldsymbol{V}_i^{-1}\left(\boldsymbol{B}_i(\boldsymbol{X}_i, A_i = a, \boldsymbol{\eta}_B) - \boldsymbol{\mu}_i(\boldsymbol{\beta}, A_i = a)\right)\right]\\
&= \sum_{a=0,1} p^a(1-p)^{1-a}\boldsymbol{D}_i^T(A_i = a)\boldsymbol{V}_i^{-1}\left(E\left[\boldsymbol{B}_i(\boldsymbol{X}_i, A_i = a, \boldsymbol{\eta}_B)\right] - \boldsymbol{\mu}_i(\boldsymbol{\beta}, A_i = a)\right)\\
&= \sum_{a=0,1} p^a(1-p)^{1-a}\boldsymbol{D}_i^T(A_i = a)\boldsymbol{V}_i^{-1}\left(E\left[\boldsymbol{Y}_i|A_i = a\right] - \boldsymbol{\mu}_i(\boldsymbol{\beta}, A_i = a)\right)\\
&= 0.
\end{aligned}$$

The PS is correctly specified

When PS is correctly specified $\pi_{ij}(\boldsymbol{X}_i, A_i, \boldsymbol{\eta}_W) = P(R_{ij} = 1|\boldsymbol{X}_i, A_i)$. We use the extended form of (a) which is given in Equation 1 in this Web supplementary material Section B, we have :

$$\begin{aligned}
(a) &= E\Bigg[\boldsymbol{D}_i^T\boldsymbol{V}_i^{-1}\boldsymbol{W}_i(\boldsymbol{X}_i, A_i, \eta_W)\left(\boldsymbol{Y}_i - \boldsymbol{\mu}_i(\boldsymbol{\beta}, A_i)\right) - \boldsymbol{D}_i^T\boldsymbol{V}_i^{-1}\boldsymbol{W}_i(\boldsymbol{X}_i, A_i, \eta_W)\left(\boldsymbol{B}_i(\boldsymbol{X}_i, A_i, \boldsymbol{\eta}_B) - \boldsymbol{\mu}_i(\boldsymbol{\beta}, A_i)\right)\\
&\quad -\boldsymbol{D}_i^T\boldsymbol{V}_i^{-1}\left(\boldsymbol{B}_i(\boldsymbol{X}_i, A_i, \boldsymbol{\eta}_B) - \boldsymbol{\mu}_i(\boldsymbol{\beta}, A_i)\right) + \boldsymbol{D}_i^T\boldsymbol{V}_i^{-1}\left(\boldsymbol{B}_i(\boldsymbol{X}_i, A_i, \boldsymbol{\eta}_B) - \boldsymbol{\mu}_i(\boldsymbol{\beta}, A_i)\right)\\
&\quad + \sum_{a=0,1} p^a(1-p)^{1-a}\boldsymbol{D}_i^T(A_i = a)\boldsymbol{V}_i^{-1}\Big(\boldsymbol{B}_i(\boldsymbol{X}_i, A_i = a, \boldsymbol{\eta}_B) - \boldsymbol{\mu}_i(\boldsymbol{\beta}, A_i = a)\Big)\Bigg],\\
&= E\left[\boldsymbol{D}_i^T\boldsymbol{V}_i^{-1}\boldsymbol{W}_i(\boldsymbol{X}_i, A_i, \eta_W)\left(\boldsymbol{Y}_i - \boldsymbol{\mu}_i(\boldsymbol{\beta}, A_i)\right)\right] + E\left[\boldsymbol{D}_i^T\left(\boldsymbol{V}_i^{-1} - \boldsymbol{V}_i^{-1}\boldsymbol{W}_i(\boldsymbol{X}_i, A_i, \eta_W)\right)\left(\boldsymbol{B}_i(\boldsymbol{X}_i, A_i, \boldsymbol{\eta}_B) - \boldsymbol{\mu}_i(\boldsymbol{\beta}, A_i)\right)\right]\\
&\quad +E\left[\sum_{a=0,1} p^a(1-p)^{1-a}\boldsymbol{D}_i^T(A_i = a)\boldsymbol{V}_i^{-1}\Big(\boldsymbol{B}_i(\boldsymbol{X}_i, A_i = a, \boldsymbol{\eta}_B) - \boldsymbol{\mu}_i(\boldsymbol{\beta}, A_i = a)\Big) - \boldsymbol{D}_i^T\boldsymbol{V}_i^{-1}\left(\boldsymbol{B}_i(\boldsymbol{X}_i, A_i, \boldsymbol{\eta}_B) - \boldsymbol{\mu}_i(\boldsymbol{\beta}, A_i)\right)\right].
\end{aligned}$$

Let's denote (ps1), (ps2) and (ps3) the first, second and third terms of equation (a). The term (ps1) is the traditional IPW estimating equation, then (ps1)=0 if the PS is correctly specified. Then,

$$
\begin{aligned}
(ps2) &= E\Big[\boldsymbol{D}_i^T E\Big[(\boldsymbol{V}_i^{-1} - [\boldsymbol{V}_i^{-1}\boldsymbol{W}_i(\boldsymbol{X}_i, A_i, \eta_W)])\,(\boldsymbol{B}_i(\boldsymbol{X}_i, A_i, \boldsymbol{\eta}_B) - \boldsymbol{\mu}_i(\boldsymbol{\beta}, A_i))\,|\boldsymbol{X}_i, A_i\Big]\Big], \\
&= E\Big[\boldsymbol{D}_i^T E\Big[\underbrace{(\boldsymbol{V}_i^{-1} - [\boldsymbol{V}_i^{-1}\boldsymbol{W}_i(\boldsymbol{X}_i, A_i, \eta_W)])}_{\boldsymbol{K}}|\boldsymbol{X}_i, A_i\Big]\Big[(\boldsymbol{B}_i(\boldsymbol{X}_i, A_i, \boldsymbol{\eta}_B) - \boldsymbol{\mu}_i(\boldsymbol{\beta}, A_i))\Big]\Big].
\end{aligned}
$$

The matrix $\boldsymbol{K}$ is equal to $\boldsymbol{V}_i^{-1}[\boldsymbol{I}-\boldsymbol{W}_i(\boldsymbol{X}_i, A_i, \eta_W)]$. The diagonal terms of $[\boldsymbol{I}-\boldsymbol{W}_i(\boldsymbol{X}_i, A_i, \eta_W)]$ are given by $\left(\frac{\pi_{ij}(\boldsymbol{X}_i,A_i)-R_{ij}}{\pi_{ij}(\boldsymbol{X}_i,A_i)}\right)$. When taking the expectation with respect to $(\boldsymbol{X}_i, A_i)$ we have $E(\frac{\pi_{ij}(\boldsymbol{X}_i,A_i)-R_{ij}}{\pi_{ij}(\boldsymbol{X}_i,A_i)}|\boldsymbol{X}_i, A_i) = \frac{\pi_{ij}(\boldsymbol{X}_i,A_i)-E(R_{ij}|\boldsymbol{X}_i,A_i)}{\pi_{ij}(\boldsymbol{X}_i,A_i)} = 0$. Follows (ps2)=0. Finally,

$$
\begin{aligned}
(ps3) &= E\Bigg[\sum_{a=0,1} p^a(1-p)^{1-a}\boldsymbol{D}_i^T(A_i = a)\boldsymbol{V}_i^{-1}\Big(\boldsymbol{B}_i(\boldsymbol{X}_i, A_i = a, \boldsymbol{\eta}_B) \\
&\qquad -\boldsymbol{\mu}_i(\boldsymbol{\beta}, A_i = a)\Big) - \boldsymbol{D}_i^T\boldsymbol{V}_i^{-1}\,(\boldsymbol{B}_i(\boldsymbol{X}_i, A_i, \boldsymbol{\eta}_B) - \boldsymbol{\mu}_i(\boldsymbol{\beta}, A_i))\,|\boldsymbol{X}_i\Bigg] \\
&= \sum_{a=0,1} p^a(1-p)^{1-a}\boldsymbol{D}_i^T(A_i = a)\boldsymbol{V}_i^{-1}\Big(\boldsymbol{B}_i(\boldsymbol{X}_i, A_i = a, \boldsymbol{\eta}_B) - \boldsymbol{\mu}_i(\boldsymbol{\beta}, A_i = a)\Big) \\
&\qquad -\boldsymbol{D}_i^T\boldsymbol{V}_i^{-1}E\,[(\boldsymbol{B}_i(\boldsymbol{X}_i, A_i, \boldsymbol{\eta}_B) - \boldsymbol{\mu}_i(\boldsymbol{\beta}, A_i))\,|\boldsymbol{X}_i] \\
&= \sum_{a=0,1} p^a(1-p)^{1-a}\boldsymbol{D}_i^T(A_i = a)\boldsymbol{V}_i^{-1}\Big(\boldsymbol{B}_i(\boldsymbol{X}_i, A_i = a, \boldsymbol{\eta}_B) - \boldsymbol{\mu}_i(\boldsymbol{\beta}, A_i = a)\Big) \\
&\qquad - \sum_{a=0,1} p^a(1-p)^{1-a}\boldsymbol{D}_i^T(A_i = a)\boldsymbol{V}_i^{-1}\Big(\boldsymbol{B}_i(\boldsymbol{X}_i, A_i = a, \boldsymbol{\eta}_B) - \boldsymbol{\mu}_i(\boldsymbol{\beta}, A_i = a)\Big) \\
&= 0.
\end{aligned}
$$

Finally, under certain regularity assumption defined in Van der Vaart (2000), we can demonstrate with the Slutsky's theorem and the central limit theorem that any estimator solving this Doubly Robust estimating equation is CAN.

2- When weights are implemented $\boldsymbol{W}_i^{1/2}(\boldsymbol{X}_i, A_i, \eta_W)\boldsymbol{V}_i^{-1}\boldsymbol{W}_i^{1/2}(\boldsymbol{X}_i, A_i, \eta_W)$ *instead of* $\boldsymbol{V}_i^{-1}\boldsymbol{W}_i(\boldsymbol{X}_i, A_i, \eta_W)$

The demonstration for (ps2)=0 does not hold anymore because $E\Big[\boldsymbol{K}|\boldsymbol{X}_i, A_i\Big] = E\Big[\boldsymbol{V}_i^{-1} - \boldsymbol{W}_i^{1/2}(\boldsymbol{X}_i, A_i, \eta_W)\boldsymbol{V}_i^{-1}\boldsymbol{W}_i^{1/2}(\boldsymbol{X}_i, A_i, \eta_W)|\boldsymbol{X}_i, A_i\Big] \neq 0$. Let's illustrate this in a cluster i with 2 individuals, when weights are implemented such as in the GENMOD procedure in SAS, we have:

$$
\Big[\boldsymbol{W}_i^{1/2}(\boldsymbol{X}_i, A_i, \eta_W)\boldsymbol{V}_i^{-1}\boldsymbol{W}_i^{1/2}(\boldsymbol{X}_i, A_i, \eta_W)\Big] = \begin{pmatrix} \alpha w_{i1} & \gamma\sqrt{w_{i1}}\sqrt{w_{i2}} \\ \gamma\sqrt{w_{i1}}\sqrt{w_{i2}} & \alpha w_{i2} \end{pmatrix},
$$

with $\boldsymbol{V}_i^{-1} = \begin{pmatrix} \alpha & \gamma \\ \gamma & \alpha \end{pmatrix}$ and therefore:

$$\boldsymbol{V}_i^{-1} - \left[\boldsymbol{W}_i^{1/2}(\boldsymbol{X}_i, A_i, \eta_W)\boldsymbol{V}_i^{-1}\boldsymbol{W}_i^{1/2}(\boldsymbol{X}_i, A_i, \eta_W)\right] = \begin{pmatrix} \alpha(1-w_{i1}) & \gamma(1-\sqrt{w_{i1}}\sqrt{w_{i2}}) \\ \gamma(1-\sqrt{w_{i1}}\sqrt{w_{i2}}) & \alpha(1-w_{i2}) \end{pmatrix}.$$

In order to ensure (ps2)=0, γ must be set to 0, which correspond to an independence working correlation structure to ensure CAN of the DR with $\boldsymbol{W}_i^{1/2}(\boldsymbol{X}_i, A_i, \eta_W)\boldsymbol{V}_i^{-1}\boldsymbol{W}_i^{1/2}(\boldsymbol{X}_i, A_i, \eta_W)$ implementation of weights. Thus, the OM has to be correctly specified or an independence correlation structure has to be used. In other words, the DR is doubly robust if and only if the independence working correlation structure is used. Another alternative is to have $w_{i1} = w_{i2}$, i.e. all the weights equal in the same cluster and non individual-specific, which is very unlikely for CRTs.

3- Checking asymptotic normality with simulations

Figure 2 displays approximate asymptotic normality of GEE, IPW, AUG, and DR on simulations described Section 5.2 of the main manuscript.

[Figure 2 about here.]

D - Use of DR in practice: The *geeDoublyRobust* R package

Implementation of this method in *R* is available on the CRAN in the function *drGeeEstimation* of the package *CRTgeeDR*. Parts of this package had been based on the *geeM* package which allows sparse matrix representations, avoiding loops in *R* and improving computation times (McDaniel and Henderson, 2014). In particular, estimation of the working correlation structure and the scale parameters are exactly the same as in *geeM*, which is derived from the procedure GENMOD in SAS as well as the *geeglm* packages in *R* (Halekoh et al., 2006).

E - Result of simulation 5.1 with $\boldsymbol{W}_i^{1/2}(\boldsymbol{X}_i, A_i, \boldsymbol{\eta}_W)\boldsymbol{V}_i^{-1}\boldsymbol{W}_i^{1/2}(\boldsymbol{X}_i, A_i, \boldsymbol{\eta}_W)$ implementation of weights

In CRTs, weights are always defined at an individual level and thus differ for each individual in the same cluster. Thus, as demonstrated in Web-supplementary Material Section C2, implementation of the Estimating Equation should use $\boldsymbol{V}_i^{-1}\boldsymbol{W}_i(\boldsymbol{X}_i, A_i, \boldsymbol{\eta}_W)$ rather than $\boldsymbol{W}_i^{1/2}(\boldsymbol{X}_i, A_i, \boldsymbol{\eta}_W)\boldsymbol{V}_i^{-1}\boldsymbol{W}_i^{1/2}(\boldsymbol{X}_i, A_i, \boldsymbol{\eta}_W)$ if a non-independence working correlation matrix is used. To ensure invertibility in the sandwich estimator of the variance, some software packages use the latter (SASInc, 2015). In our package we made available, weights are correctly implemented.

Table 1 presents the same analysis as in Table 3 in the revised manuscript but with $\boldsymbol{W}_i^{1/2}(\boldsymbol{X}_i, A_i, \boldsymbol{\eta}_W)\boldsymbol{V}_i^{-1}\boldsymbol{W}_i^{1/2}(\boldsymbol{X}_i, A_i, \boldsymbol{\eta}_W)$ implementation of weights. There is a bias associated with the use of an exchangeable correlation structure (true correlation in this case) if the OM is not correctly specified. If it is, with the $\boldsymbol{W}_i^{1/2}(\boldsymbol{X}_i, A_i, \boldsymbol{\eta}_W)\boldsymbol{V}_i^{-1}\boldsymbol{W}_i^{1/2}(\boldsymbol{X}_i, A_i, \boldsymbol{\eta}_W)$ implementation of weights, even if the PS is also correct, the estimation with exchangeable correlation structure is seen to be consistent but inefficient (large SE).

[Table 1 about here.]

F - Complementary analysis for SAM study

In this section we provide a list of variable selected for the PS and OM (treated and control) for the SAM study for each outcome of the main manuscript. First regarding the PS in Table 2, we notice that there is always an effect of the treatment on the missingness, being treated increases the probability of being missing. There is no general pattern except for the Score for knowledge of HIV/STI which is negatively associated with missingness for every outcome. The direction of a covariate effect on the PS is always the same for all outcomes for which it has been selected. Regarding the OM, Table 3 describes the selection of variables in the treated arm and Table 4 in the control arm. Different variables are selected showing both treatment-covariates interactions and imbalance of baseline covariates. Here again, if selected the association of a covariate with the outcome is in the same direction for all outcomes.

[Table 2 about here.]

[Table 3 about here.]

[Table 4 about here.]

References

Halekoh, U., Højsgaard, S., and Yan, J. (2006). The R package geepack for generalized estimating equations. *Journal of Statistical Software* **15,** 1–11.

McDaniel, L. S. and Henderson, N. (2014). *geeM: Fit Generalized Estimating Equations*. R package version 0.7.2.

Pearl, J. et al. (2009). Causal inference in statistics: An overview. *Statistics Surveys* **3,** 96–146.

SASInc (2015). The genmod procedure SAS 12.3. *http://support.sas.com/documentation/* .

Van der Vaart, A. W. (2000). *Asymptotic statistics*, volume 3. Cambridge university press.

Received February 2015. *Revised ???* ???. *Accepted ???* ???.

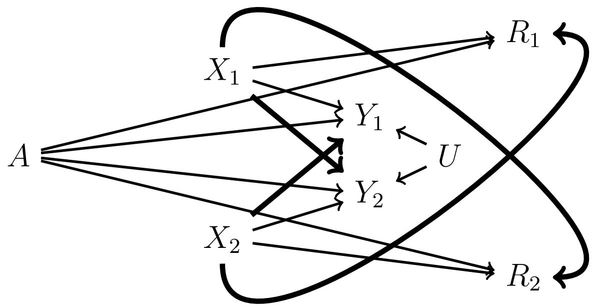

Figure 1: Directed Acyclic graph (DAG) for CRTs data with rMAR for two subjects in a same cluster and covariate interference for the outcome and the missing data generating process. Bold arrows represent the covariate interference of subject 2 over subject 1 and the covariate interference of subject 1 over subject 2. A is the treatment, X is a covariate which is here also a interfering covariate, Y is the primary outcome correlated in a cluster through U, and R is the missingness indicator.

Figure 2: Histograms of estimates values for GEE, IPW, AUG and DR with a data generation process described in Section 5.2 of the manuscript for 1000 replicates. True value of the treatment effect is 5.73 and is materialized by a vertical line.

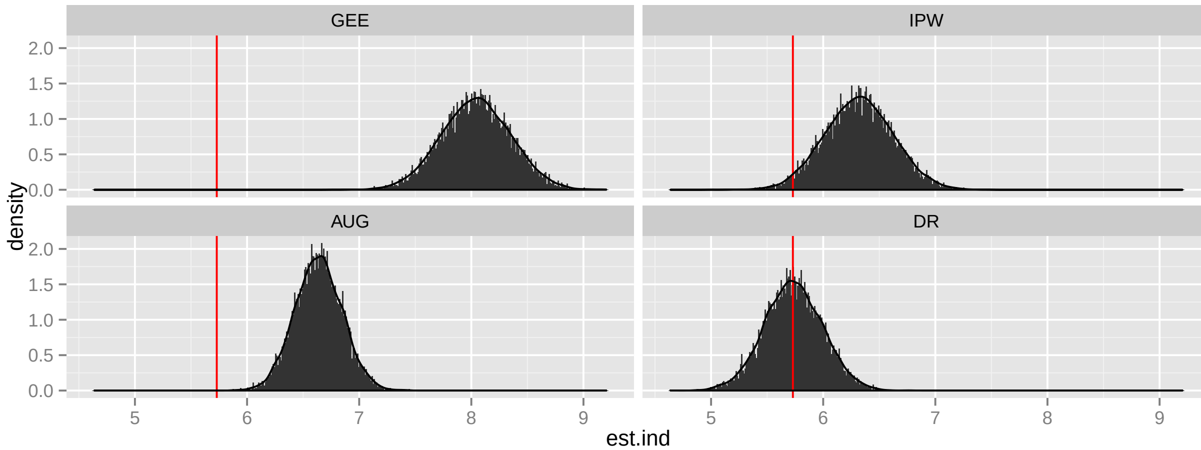

Table 1: Properties for the Doubly robust estimator with $\boldsymbol{W}_i^{1/2}(\boldsymbol{X}_i, A_i, \boldsymbol{\eta}_W)\boldsymbol{V}_i^{-1}\boldsymbol{W}_i^{1/2}(\boldsymbol{X}_i, A_i, \boldsymbol{\eta}_W)$ implementation of weights using the data generation mechanism shown below and in Equation 4 for small and large samples. Statistics for 1000 replicates are the bias compared to $M_E^* = 2.0$, the empirical standard errors over the replicates, the mean asymptotic standard error adjusted for nuisance parameters estimation in the OM and PS and the coverage for GEE and DR with independence (-I) and exchangeable (-E) working correlation matrix.

$$\begin{cases} Y_{ij} & = & 1 + A_i + X1_{ij} + \overline{X1}_{i.} + A_i X1_{ij} + \epsilon_i^O + \epsilon_{ij}^O \\ logit(P(R_{ij}=0)) & = & \frac{1}{2}(-6 + A_i + X1_{ij} + \overline{X1}_{i.} + A_i X1_{ij}) \end{cases}$$

				Standard Error (SE)						Coverage			
		Bias		Empirical		Robust		Fay's		Robust		Fay's	
	M_E^*	-I	-E	-I	-E	-I	-E	-I	-E	-I	-E	-I	-E
Small sample $M = 10, n_i = (10, 20, 30)$ with probability 1/3 each, **Low correlation**													
GEE (no missing)	2.0	-0.028	-0.028	0.497	0.497	0.440	0.440	0.527	0.527	88.8	88.8	93.4	93.4
GEE	2.0	-1.768	-1.766	0.434	0.432	0.393	0.321	0.450	0.452	3.0	0.6	4.5	4.5
IPW.NONE	2.0	-1.059	-1.238	0.677	1.025	0.518	0.746	0.624	0.967	42.4	51.2	52.0	64.7
IPW.TRUE	2.0	-0.019	0.527	0.851	1.181	0.663	0.872	0.853	1.396	84.0	89.8	89.9	97.7
AUG.NONE	2.0	-1.812	-1.812	0.179	0.179	0.393	0.321	0.450	0.452	7.1	7.1	17.4	17.4
AUG.TRUE	2.0	-1.800	-1.800	0.292	0.292	0.829	0.829	0.874	0.874	23.9	23.9	36.7	36.7
OM.MISS.PS.MISS	2.0	-1.774	-1.763	0.435	0.432	0.381	0.510	0.402	0.538	1.9	12.9	2.4	14.3
OM.MISS.PS.TRUE	2.0	-0.016	409.7	1.216	325.4	1.143	72.56	1.205	286.8	94.7	18.1	95.7	92.5
OM.TRUE.PS.MISS	2.0	-0.000	0.001	0.126	0.129	0.100	0.202	0.106	0.213	84.9	98.0	86.5	98.3
OM.TRUE.PS.TRUE	2.0	0.002	0.642	0.138	12.21	0.109	1.583	0.115	1.669	85.0	96.7	87.4	97.0
OM.NONE.PS.NONE	2.0	-0.018	0.923	0.231	9.758	0.190	2.264	0.201	2.386	85.6	99.8	87.7	99.9
Large sample $M = 100, n_i = (90, 100, 110)$ with probability 1/3 each, **Low correlation**													
GEE (no missing)	2.0	-0.002	-0.002	0.117	0.117	0.113	0.113	0.115	0.115	93.1	93.1	93.6	93.6
GEE	2.0	-1.739	-1.738	0.103	0.103	0.098	0.073	0.099	0.099	0.0	0.0	0.0	0.0
IPW.NONE	2.0	-0.997	-1.248	0.160	0.272	0.147	0.255	0.150	0.263	0.6	2.5	0.6	2.8
IPW.TRUE	2.0	0.003	0.899	0.353	0.573	0.259	0.411	0.272	0.470	92.8	32.7	93.1	40.9
AUG.NONE	2.0	-1.802	-1.802	0.039	0.039	0.249	0.249	0.250	0.250	0.0	0.0	0.0	0.0
AUG.TRUE	2.0	-1.801	-1.801	0.065	0.065	0.255	0.255	0.256	0.256	0.0	0.0	0.0	0.0
OM.MISS.PS.MISS	2.0	-1.739	-1.734	0.103	0.103	0.098	0.136	0.098	0.137	0.0	0.0	0.0	0.0
OM.MISS.PS.TRUE	2.0	-0.001	858.4	0.436	223.4	0.407	195.4	0.409	196.3	97.1	0.0	97.6	1.2
OM.TRUE.PS.MISS	2.0	0.001	0.002	0.025	0.028	0.026	0.092	0.026	0.092	94.4	100.0	94.5	100.0
OM.TRUE.PS.TRUE	2.0	0.001	0.374	0.029	4.151	0.029	0.550	0.029	0.552	95.2	99.2	95.4	99.2
OM.NONE.PS.NONE	2.0	-0.004	0.089	0.053	0.159	0.049	0.408	0.050	0.410	93.9	100.0	94.4	100.0

Marginal model for the GEE:

$\mu_{ij}(\boldsymbol{\beta}, A_i) = \beta_0 + \beta_A A_i$

OM is fitted for each treatment group $A_i = a$:

OM.TRUE $B_{ij}(\boldsymbol{X}_i, A_i = a) = \gamma_0^a + \gamma_1^a X1_{ij} + \gamma_2^a \overline{X1}_{i.}$

OM.MISS $B_{ij}(\boldsymbol{X}_i, A_i = a) = \gamma_0^a + \gamma_1^a X2_{ij}$

OM.NONE $B_{ij}(\boldsymbol{X}_i, A_i = a) = \gamma_0^a + \gamma_1^a X1_{ij}$

PS is fitted for the whole dataset:

PS.TRUE $\pi_{ij}(\boldsymbol{X}_i, A_i) = expit\left(\gamma_0^M + \gamma_A^M A_i + \gamma_1^M X1_{ij} + \gamma_2^M \overline{X1}_{i.} + \gamma_3^M A_i X1_{ij}\right)$

PS.MISS $\pi_{ij}(\boldsymbol{X}_i, A_i) = expit\left(\gamma_0^M + \gamma_A^M A_i + \gamma_1^M X2_{ij}\right)$

PS.NONE $\pi_{ij}(\boldsymbol{X}_i, A_i) = expit\left(\gamma_0^M + \gamma_A^M A_i + \gamma_1^M X1_{ij}\right)$

Table 2: Variable selected in the stepwise regression PS for primary and secondary outcomes analysis with the DR estimator. The signs of the regression coefficients are displayed to allow the interpretation of the direction of the association.

PS	R_Y (overall)	R_{Y^1} (main vaginal)	R_{Y^2} (casual vaginal)	R_{Y^3} (main anal)	R_{Y^4} (casual anal)
Treated	+	+	+	+	+
Age	+		+		
Employment					
Married			+		+
Education					
Number of children	+			+	+
Wealth					
Social Desirability	-			-	-
Religiosity	-			-	
HIV/STI Knowledge	-	-	-	-	-
Condom Behavior					
Condom Knowledge	+				+
Condom Efficacy			+	+	
Condom Peer Norm					
Never had HIV test					
Sexual Activity					
Eating Attitude			-		
Exercise		-	-		
CAGE <2		+	+		
Health		-			

Table 3: Variable selected in the stepwise regression OM in the treated arm for primary and secondary outcomes analysis with the DR estimator. The signs of the regression coefficients are displayed to allow the interpretation of the direction of the association.

OM (treated)	Y (overall)	Y^1 (main vaginal)	Y^2 (casual vaginal)	Y^3 (main anal)	Y^4 (casual anal)
Age	-				
Employment					
Married		-	+		
Education					
Number of children		-	-		-
Wealth				+	
Social Desirability			+		+
Religiosity		+			
HIV/STI Knowledge			+	+	+
Condom Behavior					
Condom Knowledge				-	
Condom Efficacy	+	+			
Condom Peer Norm					
Never had HIV test					
Sexual Activity	-	-			
Eating Attitude	+	+			
Exercise	+	+	+	+	
CAGE <2	+	+			
Health				-	

Table 4: Variable selected in the stepwise regression OM in the control arm for primary and secondary outcomes analysis with the DR estimator. The signs of the regression coefficients are displayed to allow the interpretation of the direction of the association.

OM (control)	Y (overall)	Y^1 (main vaginal)	Y^2 (casual vaginal)	Y^3 (main anal)	Y^4 (casual anal)
Age				+	
Employment	-		-	-	-
Married	-				
Education					
Number of children				-	
Wealth			+		
Social Desirability					
Religiosity					
HIV/STI Knowledge					
Condom Behavior	+	+	+		
Condom Knowledge					+
Condom Efficacy			+		
Condom Peer Norm					
Never had HIV test					-
Sexual Activity			-		
Eating Attitude				-	
Exercise					
CAGE < 2	+	+			
Health				-	