Supplemental Appendix to

Specifying the correlation structure in inverse-probability-weighted estimation for repeated measures

For brevity, we focus the discussion on a simple two-occasion dropout example. The observed data is given by (X, Y_1, R, RY_2) where (X, Y_1, R) is observed on all individuals and R indicates whether Y_2 is observed. X is a vector of baseline variables, Y_j is the continuous outcome at occasion $j = 1, 2$. In the following, we let $X^* = (1, X^T)^T$. We wish to estimate β in the marginal mean regression model:

$$E(Y_j) = \beta^T X^* \ , \ j = 1, 2 \tag{1}$$

under the standard assumption that dropout is ignorable, that is:

$$\Pr(R = 1 | X, Y_1, Y_2) = \pi(X, Y_1)$$

only depends on the observed past. We further simplify the presentation by assuming that $\pi(X, Y_1)$ is known. Additionally, suppose that the conditional correlation function $\rho = corr(Y_1, Y_2|X)$ and the conditional variance function $\sigma^2 = var(Y_j|X)$ both do not depend on X. Nowadays, a number of statistical software packages, including SAS, R and Stata, have capabilities for incorporating inverse-probability weights into generalized estimating equations. Proc GENMOD in SAS is arguably the most common software package used in epidemiologic practice to achieve this task and the software package is very well documented. For this reason we chose to focus primarily on the method implemented in Proc Genmod. In our example, the approach entails first computing occasion-specific weights, with the weight for the first occasion set equal to $W_1 = 1$ since Y_1 is observed on all individuals, whereas for the second occasion, the weight is set equal to

$W_2 = \pi(X, Y_1)^{-1}$, which accounts for the dependence of R on Y_1.[1–3] Under our assumptions, the correlation matrix for the pair of observations (Y_1, Y_2) is guaranteed to be exchangeable. Below, we provide a technical description of the weighted-least squares estimator $\widehat{\beta}(\rho^*, \sigma^*)$ computed in the Proc GENMOD procedure in SAS for a fixed (possibly incorrect) value (ρ^*, σ^*).[4] A reason for the specific approach used by Proc GENMOD to incorporate weights $W_{1,i}$ and $W_{2,i}$ is to ensure that the interpretation of $\rho^* = \rho$ and $\sigma^* = \sigma$ is retained irrespective of weighting, as respectively the correlation and the standard deviation for the original outcomes $(Y_{1,i}, Y_{2,i})$; this is essentially achieved by pre-multiplying the standard deviation σ^* of the first and second measurement, by $W_{1,i}^{-1/2}$ and $W_{2,i}^{-1/2}$ respectively (see equation (4) below) . However, this property only holds when the weights strictly depend on covariates also included in the main regression function. Unfortunately, the weighting strategy implemented in Proc GENMOD can induce bias, when the weights are used to account for dependent dropout by incorporating information on variables not included in the regression model. In fact, below we establish the following result:

Result : $\widehat{\beta}(\rho^*, \sigma^*)$ *generally converges (in probability) to a vector* $\beta^* \neq \beta$, *and is therefore biased unless at least one of the following conditions holds:*

Condition 1. $\rho^* = 0$ *and therefore* Y_1 *and* Y_2 *are assumed to be uncorrelated, or*

Condition 2. $\pi(X, Y_1) = \pi(X)$ *does not depend on* Y_1 *and therefore* W_2 *does not depend on* Y_1.

Below, we further establish that the above result applies to a larger class of weighted-generalized estimating equations, which includes the weighted-least squares estimator as a special case, but which generally allows for the nonlinear link functions typically used for binary or count outcomes. Thus, we establish that weighted-generalized estimating equations as implemented in Proc GENMOD can fail to produce a consistent estimator of the coefficients of a mean regression function. The result states that this can happen whenever occasion-specific weights are used in conjunction

with a working correlation matrix to construct generalized estimating equations in Proc GENMOD irrespective of the choice of a link function. According to the more general result, bias in coefficient estimates of such weighted-generalized estimating equations is likely to be present unless at least one of conditions 1 or 2 holds.

Next, we consider two straightforward strategies that allow more careful use of estimating equations to obtain an asymptotically unbiased estimate of β. The first approach simply entails imposing condition 1 of the Result and altogether ignoring the correlation structure for point estimation, i.e. by setting $\rho^* = 0$ in equation (3), to obtain $\widehat{\beta}(0, \sigma^*)$. Although the independence correlation structure is likely mis-specified in the longitudinal context, according to the result, this approach leads to a consistent estimate of β.[1,2] The approach is akin to pooling together multiple artificial studies, each study ending at a different follow-up time with corresponding dropout weights, and ignoring for the purposes of point estimation the fact that the same individual may contribute to multiple such artificial studies. An alternative equally simple approach only uses data on individuals with fully observed follow-up, i.e. $R_i = 1$ and sets $W_{1,i} = W_{2,i}$.[1,2] This approach is equivalent to applying a single weight, proportional to $W_{2,i}$, to all person-time contributions of an individual i with complete follow-up. In both strategies outlined above, robust standard errors or the bootstrap can be used for inference. Both strategies easily extend to a more general longitudinal study in which an individual's maximum follow-up includes $J > 2$ consecutive measurements (details omitted). However, because the finite sample is restricted to individuals with complete follow-up, the performance of the second strategy will generally be inferior to that of the first, particularly in studies with lengthy follow-up and substantial attrition.

Implications for related weighted-longitudinal analyses

Our results can be extended to the estimation of the parameters of marginal structural mean model for a repeated measures outcome from longitudinal data. A marginal structural mean model is

a model for the mean of a counterfactual outcome as a function of exposure history. Using the well-known relation between the potential outcome or counterfactual theory of causal inference and missing or coarsened data theory[1,2,5] Robins and Tchetgen Tchetgen[6] show that results analogous to those above apply when estimating marginal structural mean models via inverse-probability-of-treatment-weighting in Proc GENMOD. Like us, they describe two classes of consistent estimators. One class of estimators, introduced in Robins[5], applies the same weight to all of a subject's person-time contributions. This weight is equal to the inverse-probability-of-treatment actually received by the individual throughout the entire followup (or a stabilized version there of). Robins[5] shows one can then specify a non-independence working correlation matrix without inducing bias. This reflects the fact that in the re-weighted sample (i.e. pseudo-population), as in an ordinary randomized experiment, the treatment process is external or ancillary - that is, neither past outcome nor past covariate history are predictors of current treatment. Robins[5] and Robins et al[7] (see Section 4) prove that standard generalized estimating equations are valid if the treatment process is ancillary.

The second-class of estimators uses occasion-specific weights and an "independence" working covariance matrix. When occasion-specific weights are used, the treatment process in the weighted pseudo-population is no longer ancillary, essentially because individuals are differentially re-weighted at different times. Robins et al[7] show that for non-ancillary treatments processes, generalized estimating equations are inconsistent, unless an independence working correlation matrix is used. It follows that the occasion-specific weighted-generalized estimating equations estimators proposed by Hernán et al[8] are therefore inconsistent, except, when, as in their empirical examples, an "independence" working covariance matrix is used

Finally, we note that unless one of the two strategies outlined above is followed, the potential for bias in using a non-independence working correlation structure, remains even under the

sharp null hypothesis that the exposure history does not have a causal effect on the longitudinal outcome. Somewhat surprisingly, although estimators that use occasion-specific weights with an "independence" working covariance matrix do not explicitly incorporate an estimate of the true correlation structure of the outcomes, nonetheless, the information contained in these correlations can ultimately be recovered via the estimated inverse-probability weights. Indeed, Robins et al[6] prove that both our classes of consistent estimators contain a fully efficient estimator. A careful study of the finite sample relative efficiency of the two strategies will be published elsewhere.

Proofs

In the simple linear model (1), Proc GENMOD solves the weighted-generalized estimating equations

$$0 = \sum_i X_i^{**} Q_i \left(\rho^*, \sigma^*\right)^{-1} \varepsilon_i \left(\beta\right) \tag{2}$$

$$\varepsilon_i \left(\beta\right)^T = \left(\varepsilon_{1,i} \left(\beta\right), \varepsilon_{2,i} \left(\beta\right)\right) = \left(Y_{1,i} - \beta^T X_i^*, R_i \left(Y_{2,i} - \beta^T X_i^*\right)\right)$$

to produce the weighted least squares estimator:

$$\widehat{\beta} \left(\rho^*, \sigma^*\right) = \left\{ \sum_i X_i^{**} Q_i \left(\rho^*, \sigma^*\right)^{-1} X_i^{*T} \right\}^{-1} \left\{ \sum_i X_i^{**} Q_i \left(\rho\right)^{-1} Y^{obs} \right\}$$

where $Y^{obs} = \left(Y_1, RY_2\right)^T$; and if $R_i = 1$

$$X_i^{**} = \left(X_i^*, X_i^*\right)$$

$$Q_i \left(\rho^*, \sigma^*\right) = P_i \left(\ \sigma^*\right) S_i \left(\rho^*\ \right) P_i \left(\ \sigma^*\right)^T \tag{3}$$

$$P_i \left(\ \sigma^*\right) = \begin{pmatrix} \sigma^* & 0 \\ 0 & \sigma^* \end{pmatrix} \begin{pmatrix} W_{1,i}^{-1/2} & 0 \\ 0 & W_{2,i}^{-1/2} \end{pmatrix} \tag{4}$$

$$S_i \left(\rho^*\right) = \begin{pmatrix} 1 & \rho^* \\ \rho^* & 1 \end{pmatrix}$$

otherwise, if $R_i = 0$,

$$X_i^{**} = X_i^*$$

$$Q_i \left(\rho^*, \sigma^*\right) = \sigma^{*2} W_{1,i}^{-1}$$

We prove that the Result holds in a more general model in which $\mu_{j,i}(\beta)$ is the mean function of $[Y_{j,i}|X_i]$ such that

$$g(\mu_{j,i}(\beta)) = \beta^T h_j(X_i), j = 1, 2.$$

where $h_j(X_i^*)$ is a known function of X and time; and g is a known link function. Let

$$\varepsilon_i(\beta)^T = (\varepsilon_{1,i}(\beta), \varepsilon_{2,i}(\beta)) = (Y_{1,i} - \mu_{j,i}(\beta), R_i(Y_{2,i} - \mu_{j,i}(\beta))),$$

$$H_i = (h_1(X_i), h_2(X_i))$$

if $R_i = 1$, and

$$H_i = (h_1(X_i))$$

if $R_i = 0$.Thus we wish to show the Result holds for $\widehat{\beta}$ that solves the weighted-generalized-estimating-equations:

$$0 = \sum_i H_i Q_i^{-1}(\rho^*, \sigma^*)\varepsilon_i(\beta)$$

It is sufficient to show that the estimating function on the right-hand side of the above display is generally unbiased only if condition 1 or 2 holds. Some algebra gives

$$\begin{aligned} Q_i^{-1}(\rho^*, \sigma^*)\varepsilon_i(\beta) = & \frac{\sigma^{*-2}}{(1-\rho^{*2})} \begin{pmatrix} \varepsilon_{1,i}(\beta) W_{1,i} \\ \varepsilon_{2,i}(\beta) W_{2,i} \end{pmatrix} R_i \\ & - \frac{\sigma^{*-2}}{(1-\rho^{*2})} \begin{pmatrix} \varepsilon_{2,i}(\beta) W_{1,i}^{1/2} W_{2,i}^{1/2} \rho^* \\ \varepsilon_{1,i}(\beta) W_{1,i}^{1/2} W_{2,i}^{1/2} \rho^* \end{pmatrix} R_i \\ & + \sigma^{*-2} W_{1,i} \varepsilon_{1,i}(\beta)(1 - R_i) \end{aligned}$$

and therefore

$$
\begin{aligned}
&E\left\{H_i Q_i^{-1}\left(\rho^*,\sigma^*\right)\varepsilon_i\left(\beta\right)\right\}\\
&= E\left[\pi\left(X_i,Y_{1,i}\right)\left\{\frac{\sigma^{*-2}\rho^{*2}}{(1-\rho^{*2})}h_1(X_i)\varepsilon_{1,i}\left(\beta\right)W_{1,i}+\frac{\sigma^{*-2}}{(1-\rho^{*2})}h_2(X_i)\varepsilon_{2,i}\left(\beta\right)W_{2,i}\right.\right.\\
&\left.\left.-\frac{\sigma^{*-2}\ \rho^* W_{1,i}^{1/2}W_{2,i}^{1/2}}{(1-\rho^{*2})}\left(h_1(X_i)\varepsilon_{2,i}\left(\beta\right)+h_2(X_i)\varepsilon_{1,i}\left(\beta\right)\right)\right\}\right.\\
&\left.+\sigma^{*-2}W_{1,i}\varepsilon_{1,i}\left(\beta\right)\right]\\
&= E\left[\pi\left(X_i,Y_{1,i}\right)\left\{\frac{\sigma^{*-2}\rho^{*2}}{(1-\rho^{*2})}h_1(X_i)\varepsilon_{1,i}\left(\beta\right)W_{1,i}\right.\right.\\
&\left.\left.-\frac{\sigma^{*-2}\ \rho^* W_{1,i}^{1/2}W_{2,i}^{1/2}}{(1-\rho^{*2})}\left(h_1(X_i)\varepsilon_{2,i}\left(\beta\right)+h_2(X_i)\varepsilon_{1,i}\left(\beta\right)\right)\right\}\right]
\end{aligned}
$$

is equal to zero provided that either $\rho^* = 0$ or $\pi\left(X_i, Y_{1,i}\right)$ does not depend on $Y_{1,i}$. In the first case, the proof is immediate; in the second case, the proof follows from the fact that $E\left(\varepsilon_{j,i}\left(\beta\right)|X_i\right) = 0$, $j = 1, 2$.

References

[1] Robins JM. and Rotnitzky, A. (1995). Semiparametric efficiency in multivariate regression models with missing data. Journal of the American Statistical Association, 90:122-129.

[2] Robins JM, Rotnitzky A, Zhao L-P. (1995). Analysis of semiparametric regression models for repeated outcomes in the presence of missing data. Journal of the American Statistical Association, 90:106-121.

[3] Weuve J, Tchetgen Tchetgen EJ, Glymour MM, Beck TL, Aggarwal NT, Wilson RS, Evans DA, Mendes de Leon CF. Accounting for bias due to selective attrition: The example of smoking and cognitive decline. *Epidemiology*, 2012; 23:119-128.

[4] SAS/STAT(R) 9.2 User's Guide, Second Edition.

[5] Robins JM. (1999). Marginal Structural Models versus Structural Nested Models as Tools for Causal Inference. Statistical Models in Epidemiology: The Environment and Clinical Trials. M.E. Halloran and D. Berry, Editors, IMA Volume 116, NY: Springer-Verlag, pp. 95-134.

[6] Robins JM, Tchetgen Tchetgen EJ (2011). On Bias and efficiency properties of semi-parametric estimators of marginal structural models for repeated outcomes:theoretical considerations. In progress.

[7] Robins JM, Greenland S, Hu F-C. (1999). Estimation of the causal effect of a time-varying exposure on the marginal mean of a repeated binary outcome. Journal of the American Statistical Association - Applications and Case Studies, 94:687-700.

[8] Hernán MA, Brumback B, Robins JM. (2002). Estimating the causal effect of zidovudine on CD4 count with a marginal structural model for repeated measures. Statistics in Medicine, 21:1689-1709.