# A cautionary note on specification of the correlation structure in inverse-probability-weighted estimation for repeated measures

by

Eric J. Tchetgen Tchetgen[a,b]

M. Maria Glymour[c]

Jennifer Weuve[d]

James Robins[a,b]

[a]Department of Epidemiology, Harvard School of Public Health

[b]Department of Biostatistics, Harvard School of Public Health

[c]Department of Society, Human Development and Health, Harvard School of Public Health

[d]Rush Institute for Healthy Aging and Department of Internal Medicine, Rush University Medical Center

**Abstract**

In studies of repeated outcomes, it is customary to account for dependence in the outcomes of a given individual by incorporating a working correlation structure for the individual's outcomes in generalized estimating equations. Inverse-probability weighting is also a common approach used for causal inference and missing or censored data problems in epidemiology. In the absence of inverse-probability weights, it is well known that generalized estimating equations consistently estimate the parameters of a correctly specified regression model, irrespective of whether or not the working correlation structure is correct. In this commentary, we show that the situation is quite different when weights are present, and that regression estimates obtained from generalized estimating equations that are inverse-probability-weighted can be biased, even when the correlation structure is correct. Specifically, we show that weighted-generalized estimating equations as implemented in Proc GENMOD in SAS can produce biased regression estimates even when modeling bias is absent. We discuss possible strategies to avoid this potential bias and illustrate this phenomenon in an epidemiologic application.

Inverse-probability weighting recently has gained popularity as an intuitive and practical approach for estimation in the context of causal inference and missing data problems in epidemiology. Nowadays, inverse-probability weighting an individual's data by the probability density for his or her observed exposure history is most commonly used in epidemiology to account for time-varying confounding when estimating the parameters indexing the joint causal effects of a time-varying exposure in a marginal structural model.[1–4] Inverse-probability-weights for drop-out are similarly incorporated when estimating the regression parameters of a right-censored outcome, or to account for dependent forms of attrition in the analysis of repeated measures.[5–7] In studies of repeated outcomes, it is customary to account for dependence in the outcomes of a given individual by specifying a working correlation structure for the individual's outcomes; and to subsequently estimate the mean regression parameters of main interest using generalized estimating equations which incorporate both the inverse-probability weights and the working correlation structure. In the absence of weights, it is well known that generalized estimating equations consistently estimate the parameters of a correctly specified regression model, irrespective of whether the working correlation structure is correct. In this note, we show that the situation is quite different when weights are present, and that regression estimates obtained from generalized estimating equations that are inverse-probability-weighted can be biased even when the correlation structure is correct. Specifically, we show that weighted-generalized estimating equations as implemented in Proc GENMOD in SAS can produce biased regression estimates even when modeling bias is absent–that is, even though models for both the regression function and inverse-probability-weights are correct. Below, we demonstrate that our theoretical result can have implications for epidemiologic practice, by illustrating the aforementioned bias in a recent analysis of the effects of smoking on cognitive decline in an aging population subject to dependent attrition due to death and other unrelated drop-out.

## Bias of weighted generalized estimating equations

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 $\beta$ in the marginal mean regression model:

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

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

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

only depends on the observed past. We further simplify the presentation by assuming that $\pi\left(X, Y_1\right)$ 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 $W_2 = \pi\left(X, Y_1\right)^{-1}$, which accounts for the dependence of $R$ on $Y_1$.[5–7] Under our assumptions, the correlation matrix for the pair of observations $(Y_1, Y_2)$ is guaranteed to be exchangeable. In the appendix, we provide a technical description of the weighted-least squares estimator $\widehat{\beta}\left(\rho^*, \sigma^*\right)$

computed in the Proc GENMOD procedure in SAS for a fixed (possibly incorrect) value $(\rho^*, \sigma^*)$.[8] 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 $\sigma^*$ of the first and second measurement, by $W_{1,i}^{-1/2}$ and $W_{2,i}^{-1/2}$ respectively (see equation (4) of the Appendix) . However, this property only holds when the weights strictly depend on covariates also included in the main regression function. Unfortunately, we prove in the appendix that 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, 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* $Y_1$.

The second condition in the above result will generally fail to hold in settings where, as we assume throughout, it is believed that the observed past (here, $Y_1$) predicts an individual's chance of attrition. When $\rho^*$ is random, that is when it is estimated from the data, the first condition may be modified to state that for consistency, the estimated within-person correlation must converge with sample size to zero. In either case, whether $\rho^*$ is fixed or random, the true correlation $\rho$ will rarely be zero when $Y_1$ and $Y_2$ are consecutive measures of the same underlying process in a given individual; therefore condition 1 essentially implies incorrectly assuming uncorrelated outcomes in the analysis. Therefore, the result states that the weighted least squares estimator $\widehat{\beta}(\rho^*, \sigma^*)$ will

generally be biased for $\beta$, even when $(\rho^*, \sigma^*) = (\rho, \sigma)$ and model mis-specification is completely absent. In the appendix, we 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 $\beta$. The first approach simply entails imposing condition 1 of the Result and altogether ignoring the correlation structure for 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 $\beta$.[5,6] 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 $W_{1,i} = W_{2,i}$.[5,6] 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. For this reason, the following data example will only consider the first estimation strategy.

## A data example

We briefly illustrate the results of the previous section in an application involving a recent analysis of the effects of smoking on cognitive decline in an aging population subject to substantial attrition due to death and drop-out for other reasons.[7] In their paper, Weuve et al noted that selective attrition in this population may introduce bias into analyses of the effects of smoking status measured at the start of follow-up on cognitive decline, mainly due to the facts that:[7]

(1) an individual's evolving health status is likely to be a common cause for attrition and cognitive decline among survivors who do not drop out.

(2) an individual's evolving health status is likely to mediate the causal effect of smoking on cognitive decline.

To appropriately account for (1) and (2) Weuve et al[7] used inverse-probability-of-attrition weights and examined the influence of selective attrition on the estimated association of current smoking (versus never smoking) with cognitive decline in participants of the Chicago Health and Aging Project (n=3,713), aged 65-109, who were current smokers or never-smokers, and underwent cognitive assessments up to 5 times at 3-year intervals. They used pooled logistic regression to fit predictive models of attrition due to death or study drop-out across the follow-up waves using both baseline and time-updated data to construct the inverse-probability-of-attrition weights. We refer

the reader to Weuve et al[7] for additional details on the design and analysis of the study, also see Chaix et al[10] and Tchetgen Tchetgen et al[9] for further considerations of issues related to statistical and causal inference in connection to this analysis. For inference, Weuve et al[7] fit unweighted and weighted, generalized estimating equations for a linear mean regression model contrasting rates of change in cognitive scores in current versus never-smokers, adjusting for the following baseline confounders in the regression: age, sex, race, education, and alcohol consumption. Their analysis assumed a compound symmetry correlation structure, also known as an exchangeable correlation structure, for the 5 serial measurements of cognitive function (coded as z-scores). Confidence intervals are obtained via the bootstrap. In unweighted analyses, current smokers' cognitive scores declined 0.11 standard units per decade more rapidly than never-smokers' (95% CI= -0.20 to -0.02). Weighting to account for attrition yielded an estimate that was twice as large, with smokers' estimated 10-year rate of decline 0.20 units faster than never-smokers' (95% CI= -0.36 to -0.04). The within-subject correlation was estimated to be approximately equal to 0.5, suggesting that condition 1 of the Result is unlikely to hold, further suggesting that the inverse-probability-weighted estimate of the effect of smoking obtained in Weuve et al[7] may be biased. To investigate this possibility, we fit the generalized estimating equations both weighted and unweighted for attrition, assuming that the five outcome measures were mutually uncorrelated, i.e. under the independence working correlation structure. In the new unweighted analysis, current smokers' cognitive scores declined 0.13 standard units per decade more rapidly than never-smokers (95% CI= -0.24 to -0.03), a result that is compatible with the previous estimate (equal to 0.11) obtained by Weuve et al.[7] In contrast, weighted-analyses based on the independence working correlation structure delivered effect estimates that were slightly larger than weighted results obtained by Weuve et al[7], with an estimated increased decline of 0.26 (versus 0.20 obtained by Weuve et al[7]) for smokers versus never-smokers (95% CI: -0.44 to -0.08). Although the point estimates contrasting smokers' and

never-smokers' rates of cognitive decline appear to have been relatively robust to bias induced by the use of an exchangeable working correlation structure, the estimated 10-year cognitive decline for never-smokers was notably more sensitive to the correlation structure used in these analyses. Specifically, under specification of an independence working correlation structure, we obtained an estimated decline of 0.64 for never-smokers (95% CI: -0.81 to -0.47), which was somewhat smaller than the estimated decline of 0.82 for never-smokers (95% CI: -0.97 to -0.66) reported in Weuve et al[7] using an exchangeable working correlation structure.

## Implications for related weighted-longitudinal analyses

In the previous sections, we established using both theoretical arguments and empirical evidence from a real world application, that specifying a working correlation structure in a longitudinal analysis that involves occasion-specific inverse-probability weights for drop-out as implemented in SAS Proc GENMOD, can result in biased estimates of regression coefficients, unless an independence working correlation structure is assumed.

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[1−4] Robins and Tchetgen Tchetgen[11] 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[2], 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[2] 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[2] and Robins et al[12] (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[12] 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[4] 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[11] 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.

Acknowledgment: The authors thank Todd L. Beck for his assistance with the data analysis.

# Appendix

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

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

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 $E\left(\varepsilon_{j,i}(\beta) \mid X_i\right) = 0$, $j = 1, 2$.

# References


[1] Robins JM. (1998). Marginal structural models. 1997 Proceedings of the American Statistical Association. Section on Bayesian Statistical Science, pp. 1-10.

[2] 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.

[3] Robins JM, Hernán M, Brumback B. (2000). Marginal structural models and causal inference in epidemiology. Epidemiology, 11(5):550-560.

[4] 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.

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

[6] 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.

[7] 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; in press.

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

[9] Tchetgen Tchetgen EJ, Glymour MM, Weuve J, Shpitser I. To weight or not to weight:On the relation between inverse-probability-weighting and principal stratification for truncation by death. *Epidemiology*. 2012; in press.

[10] Chaix B, Evans D, Merlo J, Suzuki E. Weighing up the dead and missing: reflections on inverse-probability weighting and principal stratification to address truncation by death. *Epidemiology*. 2012;in press.

[11] Robins JM, Tchetgen Tchetgen EJ (2011). Bias and efficiency properties of semi-parametric estimators of marginal structural models for repeated outcomes. Technical report.

[12] 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.