Online Supplemental Material for "Doubly Robust Estimation of Standardized Risk Difference and Ratio in the Exposed Population"

1. Estimating functions for the MSM in the exposed

In this section, we show that our proposed estimators are derived straightforwardly using the estimating equation framework in general notation. We define the full-data estimating function (Tsiatis, 2008; Tan, 2010) of a marginal structural model (MSM) in the exposed (Model 1) with the 2-dimensional weight function $w(a) = (1, a)^{\mathrm{T}}$ for each member $i = 1,..., n$:

$$\tau_{1i} = A_i \sum_a w(a)(Y_i^a - \mu_{1a}).$$

These are unbiased (ie, the expectation equals 0-vector) estimating functions because Model 1 implies the following moment conditions:

$$\mathrm{E}[Y^a - \mu_{1a} \mid A = 1] = \mathrm{E}[A(Y^a - \mu_{1a}) \mid A = 1] = \mathrm{E}[A(Y^a - \mu_{1a})]/\mathrm{Pr}[A = 1] = 0,$$

for $a = 0, 1$. Equating the sample means of τ_{1i} to 0 (ie, to the population mean) yields asymptotically unbiased estimators of μ_{1a} for $a = 0, 1$. The full-data estimating equations reduce to

$$\begin{aligned} n^{-1}\sum_i \tau_{1i} &= n^{-1}\sum_i \begin{bmatrix} A_i(Y_i^1 - \mu_{11}) + A_i(Y_i^0 - \mu_{10}) \\ A_i(Y_i^1 - \mu_{11}) \end{bmatrix} \\ &= n^{-1}\sum_i \begin{bmatrix} A_i(Y_i^0 - \mu_{10}) \\ A_i(Y_i^1 - \mu_{11}) \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix}. \end{aligned}$$

Because both the counterfactual variables Y_i^1 and Y_i^0 cannot be observed for every i, the aforementioned full-data estimating equations might suffer from the missing-data problem

(Kang and Schafer, 2007). First, consider the second element of the estimating function, $A_i(Y_i^1 - \mu_{11})$. By consistency assumption, $A_i(Y_i^1 - \mu_{11}) = A_i(Y_i - \mu_{11})$ for all i, and thus E[$A(Y - \mu_{11})$] also equals to 0. Therefore, we can nonparametrically estimate μ_{11} from observed data without any other assumptions.

By contrast, μ_{10} cannot be estimated through the first element $A_i(Y_i^0 - \mu_{10})$, because Y_i^0 is unavailable for population members whose $A_i = 1$; moreover, $A_i(Y_i^0 - \mu_{10}) = A_i(Y_i - \mu_{10}) \equiv 0$ for $A_i = 0$, which contains no information on μ_{10}. To recover the information for μ_{10} lost due to missing Y_i^0, we weight the unexposed population by the inverse odds of the probability of their exposure ($A_i = 0$) based on the propensity score (PS) $\Pr[A = 1|X = x_i] = \pi(x_i)$, given their confounder value x_i (Sato and Matsuyama, 2003; Hernán and Robins, 2006). One can readily show that

$$\mathrm{E}\left[\frac{(1-A)\pi(X)}{1-\pi(X)}(Y-\mu_{10})\right] = \mathrm{E}[A(Y^0 - \mu_{10})] = 0$$

under the conditional exchangeablity assumption for Y_i^0 (ie, $Y^0 \coprod A|X$). Thus, unbiased "observed" estimating equations that only rely on observed data but contain causal target parameters μ_{1a} can be obtained as

$$n^{-1}\sum_i \begin{bmatrix} A_i(Y_i - \mu_{11}) \\ (1-A_i)\dfrac{\hat{\pi}_i}{1-\hat{\pi}_i}(Y_i - \mu_{10}) \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix},$$

where $\pi(x; \alpha)$ is a parametric model of PS $\pi(x)$ and $\hat{\pi}_i = \pi(X_i; \hat{\alpha})$ is a fitted PS value for member i calculated from X_i and maximum likelihood estimates of α.

The above weighted estimating functions can be improved from the viewpoints of efficiency and robustness. In line with the argument of Tan (2010) that targets causal parameters in the total population, given any full-data estimating function targeting the population with exposure level a^* for parameter θ, that is, $I[A_i = a^*]\Sigma_a \eta_{a,i}(\theta)$, whose expectation is 0 (e.g., in our τ_{1i}, $a^* = 1$ and $\eta_{a,i}(\mu_{1a}) = (1, a)^{\mathrm{T}}(Y_i^a - \mu_{1a})$), the following "observed" estimating function must be also mean-zero:

$$\frac{\pi_{a^*}(X_i)}{\pi_{A_i}(X_i)}\eta_{A_i}(\theta) - \left\{\frac{\pi_{a^*}(X_i)}{\pi_{A_i}(X_i)}\varpi_{A_i}(X_i) - \sum\nolimits_a I[A_i = a^*]\varpi_a(X_i)\right\},$$

where $I[\cdot]$ is an indicator function, $\pi_a(x) = \Pr[A = a \mid X = x]$, and $\varpi_a(x) = \mathrm{E}[\eta_a(\theta) \mid X = x]$.

The estimating function is doubly robust (DR), because for unknown $\pi_a(x)$ and $\varpi_a(x)$, its mean is still 0 if either $\pi_a(x)$ or $\varpi_a(x)$, but not necessarily both, is correctly specified as a parametric model. The first term of the augmented estimating function is the aforementioned weighted estimating function when the exposure level A can only take 0 or 1 and we choose the target as $a^* = 1$ (ie, exposed). If conditional exchangeability for Y^a (for all a) is correct, then $\mathrm{E}[Y^a|X = x] = \mathrm{E}[Y^a|A = a, X = x] = \mathrm{E}[Y|A = a, X = x]$, which may be modeled by regression functions $m_a(x; \beta)$. Applying this augmentation to our notation in the main text, we obtain the following augmented estimating equations:

$$\begin{cases} n^{-1}\sum_i \left[\begin{array}{l} A_i(Y_i - \mu_{11}) + (1 - A_i)\dfrac{\hat{\pi}_i}{1 - \hat{\pi}_i}(Y_i - \mu_{10}) \\ \quad - \left\{\begin{array}{l} A_i[m_1(X_i;\hat{\beta}) - \mu_{11}] + \dfrac{(1 - A_i)\hat{\pi}_i}{1 - \hat{\pi}_i}[m_0(X_i;\hat{\beta}) - \mu_{10}] \\ \qquad - A_i[m_1(X_i;\hat{\beta}) - \mu_{11}] - A_i[m_0(X_i;\hat{\beta}) - \mu_{10}] \end{array}\right\} \end{array}\right] = 0, \\ n^{-1}\sum_i A_i(Y_i - \mu_{11}) = 0. \end{cases}$$

$$\Leftrightarrow \begin{cases} n^{-1}\sum_i \left[\begin{array}{l} (1-A_i)\dfrac{\hat{\pi}_i}{1-\hat{\pi}_i}(Y_i-\mu_{10}) \\ \quad -\left\{(1-A_i)\dfrac{\hat{\pi}_i}{1-\hat{\pi}_i}[m_0(X_i;\hat{\beta})-\mu_{10}]-A_i[m_0(X_i;\hat{\beta})-\mu_{10}]\right\} \end{array} \right] = 0, \\ n^{-1}\sum_i A_i(Y_i-\mu_{11}) = 0. \end{cases}$$

Note that the resulting equations only contain $m_0(x; \beta)$ among $m_a(x; \beta)$, we need weaker conditional exchangeability for Y^0 for our standardized estimator in the exposed. Solving the equations provides our proposed DR standardization estimators, $\hat{\mu}_{11,\mathrm{DR}} = \hat{\mu}_{11}$ and $\hat{\mu}_{10,\mathrm{DR}}$.

2. Influence functions and variance estimators

Our estimators $\hat{\mu}_{1a,\text{IPW-RW}}$ and $\hat{\mu}_{1a,\mathrm{DR}}$ possess their own influence functions because they are asymptotically linear estimators (Tsiatis, 2006). To derive their asymptotic variance estimators, we first obtain the influence functions of $\hat{\mu}_{1a,\text{IPW-RW}}$ and $\hat{\mu}_{1a,\mathrm{DR}}$.

The influence function $\phi_{11,i}$ of member i for $\hat{\mu}_{11,\text{IPW-RW}} = \hat{\mu}_{11,\mathrm{DR}} = \overline{Y}_1$ is obtained as $(n/n_1)A_i(Y_i - \mu_{11})$, where $n_1 = \sum_i A_i$ is the size of the exposed group, because

$$n^{1/2}(\hat{\mu}_{11}-\mu_{11}) = n^{-1/2}\sum_i (n/n_1)A_i(Y_i-\mu_{11}) = n^{-1/2}\sum_i \phi_{11,i}\,.$$

Suppose that the probability limit of n_1/n is present and lies between 0 and 1. Using the central limit theorem and Slutsky's theorem, the asymptotic variance of $\hat{\mu}_{11}$ is obtained as $(1/n)\mathrm{var}(\phi_{11})$, which is estimated using $n_1^{-1}\sum_i [A_i(Y_i-\overline{Y}_1)]^2$, the (maximum likelihood) sample-variance estimator in the exposed. If T_i might be censored by C_i and Y_i cannot be observed for all members, one may instead use the inverse-probability of censoring (IPC)-weighted influence function $I(C_i \neq T_i^*)\phi_{11,i}/K_C(T_i^*)$, where $T_i^* = \min(t, T_i, C_i)$ is the observed time and $K_C(t)$ is the probability of remaining uncensored at t (estimated using the Kaplan–Meier method for C_i). This does not yield an efficient influence function when censoring is viewed as a monotone coarsening problem (Tsiatis, 2006, section 9.3). However, the resulting estimator is exactly the Kaplan–Meier estimator (Satten and Datta, 2001) adopted in the main text.

Next, we consider the influence function $\phi_{10,\text{IPW-RW},i}$ of member i for $\hat{\mu}_{10,\text{IPW-RW}}$. Because the weighted estimator is the solution of the estimating equation

$$n^{-1}\sum_i U_i(\hat{\mu}_{10,\text{IPW-RW}}, \hat{\alpha}) = n^{-1}\sum_i (1-A_i)\frac{\hat{\pi}_i}{1-\hat{\pi}_i}(Y_i - \hat{\mu}_{10,\text{IPW-RW}}) = 0,$$

applying the m-estimation theory (Tsiatis, 2007, p. 30, p. 207) followed by the Taylor series expansion reveals that

$$\begin{aligned} 0 &= \sum_i U_i(\hat{\mu}_{10,\text{IPW-RW}}, \hat{\alpha}) \\ &= \sum_i U_i(\mu_{10}, \alpha) + \left\{\sum_i \frac{\partial U_i(\mu_{10}, \alpha)}{\partial \mu_{10}}\right\}(\hat{\mu}_{10,\text{IPW-RW}} - \mu_{10}) + \left\{\sum_i \frac{\partial U_i(\mu_{10}, \alpha)}{\partial \alpha^{\mathrm{T}}}\right\}(\hat{\alpha} - \alpha) + o_p(1). \end{aligned}$$

Because α is estimated using maximum likelihood, standard theory shows that

$$n^{1/2}(\hat{\alpha} - \alpha) = n^{-1/2}\sum_i \mathrm{E}\left[-\frac{\partial S_\alpha(X)}{\partial \alpha^{\mathrm{T}}}\right]^{-1} S_\alpha(X_i) + o_p(1),$$

where $S_\alpha(X_i) = \partial \log \pi_{A_i}(X_i;\ \alpha)/\partial\alpha$ is member i's score function of a PS model. Therefore, it converges in probability,

$$n^{1/2}(\hat{\mu}_{10,\text{IPW-RW}} - \mu_{10}) = -n^{-1/2}\sum_i \phi_{10,\text{IPW-RW},i} + o_p(1)$$

$$= -n^{-1/2}\,\mathrm{E}\left\{\frac{\partial U(\mu_{10},\alpha)}{\partial \mu_{10}}\right\}^{-1}\sum_i\left[U_i(\mu_{10},\alpha) - \mathrm{E}\left\{\frac{\partial U(\mu_{10},\alpha)}{\partial \alpha^{\mathrm{T}}}\right\}\mathrm{E}\left[\frac{\partial S_\alpha(X)}{\partial \alpha^{\mathrm{T}}}\right]^{-1} S_\alpha(X_i)\right] + o_p(1).$$

The asymptotic variance of $\hat{\mu}_{10,\text{SM}}$ is estimated using the sample variance of $\phi_{10,\text{SM},i}$ (estimates are substituted for unknowns), which can be rewritten as the sandwich variance formula $(1/n)\mathrm{B}^{-1}\mathrm{M}\mathrm{B}^{-1,\mathrm{T}}$, where

$$\mathrm{B}^{-1}\text{ (bread)} = 1 \Big/ n^{-1}\sum_i \frac{\partial U_i(\hat{\mu}_{10,\text{IPW-RW}},\hat{\alpha})}{\partial \mu_{10}},$$

$$\mathrm{M}\text{ (meat)} = n^{-1}\sum_i\left[\begin{array}{l} U_i(\hat{\mu}_{10,\text{IPW-RW}},\hat{\alpha}) \\ \quad -\left\{n^{-1}\sum_i \frac{\partial U_i(\hat{\mu}_{10,\text{IPW-RW}},\hat{\alpha})}{\partial \alpha^{\mathrm{T}}}\right\}\left[n^{-1}\sum_i \frac{\partial S_{\hat{\alpha}}(X_i)}{\partial \alpha^{\mathrm{T}}}\right]^{-1} S_{\hat{\alpha}}(X_i)\end{array}\right]^2.$$

If a logistic model is used for the PS, then the sandwich formula reduces to

$$\frac{\displaystyle\sum_i\left\{\begin{array}{l}(1-A_i)\dfrac{\hat{\pi}_i}{1-\hat{\pi}_i}(Y_i - \hat{\mu}_{10,\text{IPW-RW}}) \\ \quad + \hat{\mathrm{E}}\left[(1-A_i)\dfrac{\hat{\pi}_i}{1-\hat{\pi}_i}(Y_i - \hat{\mu}_{10,\text{IPW-RW}})\mathbf{X}_i^{\mathrm{T}}\right]\hat{\mathrm{E}}\left[\hat{\pi}_i(1-\hat{\pi}_i)\mathbf{X}_i\mathbf{X}_i^{\mathrm{T}}\right]^{-1}(A_i - \hat{\pi}_i)\mathbf{X}_i\end{array}\right\}^2}{\left\{-\displaystyle\sum_i (1-A_i)\frac{\hat{\pi}_i}{1-\hat{\pi}_i}\right\}^2},$$

in which $\mathbf{X}_i = (1, X_i^{\mathrm{T}})^{\mathrm{T}}$ is member i's covariate-vector with intercept, and we use the shorthand notation $\hat{\mathrm{E}}[V_i] = n^{-1}\sum_i V_i$. In the case of censored outcome data, one may replace $\hat{\mu}_{10,\text{IPW-RW}}$ with $\hat{\mu}_{10,\text{IPW-RW}}(t)$ and $U_i(\hat{\mu}_{10,\text{IPW-RW}},\hat{\alpha})$ with IPC-weighted $I[C_i \neq T_i^*]U_i(\hat{\mu}_{10,\text{IPW-RW}}(t),\hat{\alpha})\big/\hat{K}_C(T_i^*)$ in each calculation.

In a similar manner, the influence function $\phi_{10,\mathrm{DR},i}$ of member i for $\hat{\mu}_{10,\mathrm{DR}}$ is

$$\mathrm{E}\left\{\frac{\partial U(\mu_{10},\alpha,\beta)}{\partial\mu_{10}}\right\}^{-1}\left[\begin{aligned}U_i(\mu_{10},\alpha,\beta)+\mathrm{E}\left\{\frac{\partial U(\mu_{10},\alpha,\beta)}{\partial\alpha^{\mathrm{T}}}\right\}\mathrm{E}\left[-\frac{\partial S_\alpha(X)}{\partial\alpha^{\mathrm{T}}}\right]^{-1}S_\alpha(X_i)\\+\mathrm{E}\left\{\frac{\partial U(\mu_{10},\alpha,\beta)}{\partial\beta^{\mathrm{T}}}\right\}\mathrm{E}\left[-\frac{\partial S_\beta(X)}{\partial\beta^{\mathrm{T}}}\right]^{-1}S_\beta(X_i)\end{aligned}\right],$$

where

$$\begin{aligned}U_i(\mu_{10},\alpha,\beta)&=(1-A_i)\frac{\pi(X_i;\alpha)}{1-\pi(X_i;\alpha)}(Y_i-\mu_{10})\\&\quad-\left\{(1-A_i)\frac{\pi(X_i;\alpha)}{1-\pi(X_i;\alpha)}[m_0(X_i;\beta)-\mu_{10}]-A_i[m_0(X_i;\beta)-\mu_{10}]\right\}\end{aligned}$$

and $S_\beta(X_i)$ is a score function of an outcome regression model. The sample variance of the influence function $\phi_{10,\mathrm{DR},i}$, or equivalently the sandwich estimator $(1/n)\mathrm{B}^{-1}\mathrm{M}\mathrm{B}^{-1,\mathrm{T}}$, is

$$\text{B (bread)} = -\,n^{-1}\sum_i A_i = n_1/n,$$

$$\text{M (meat)} = n^{-1}\sum_i\left\{U_i(\hat{\mu}_{10,\mathrm{DR}},\hat{\alpha},\hat{\beta})+Q_{\alpha,i}+Q_{\beta,i}\right\}^2,$$

$$Q_{\alpha,i}=\hat{\mathrm{E}}\left[(1-A_i)\frac{\hat{\pi}_i}{1-\hat{\pi}_i}(Y_i-\hat{m}_{0,i})\mathbf{X}_i^{\mathrm{T}}\right]\hat{\mathrm{E}}\left[\hat{\pi}_i(1-\hat{\pi}_i)\mathbf{X}_i\mathbf{X}_i^{\mathrm{T}}\right]^{-1}(A_i-\hat{\pi}_i)\mathbf{X}_i$$

$$Q_{\beta,i}=\hat{\mathrm{E}}\left[\frac{\hat{\pi}_i-A_i}{1-\hat{\pi}_i}\hat{m}_{0,i}(1-\hat{m}_{0,i})\mathbf{X}_i^{\mathrm{T}}\right]\hat{\mathrm{E}}\left[(1-A_i)\hat{m}_{0,i}(1-\hat{m}_{0,i})\mathbf{X}_i\mathbf{X}_i^{\mathrm{T}}\right]^{-1}(1-A_i)(Y_i-\hat{m}_{0,i})\mathbf{X}_i$$

if both PS and outcome regression (in the unexposed, $A_i = 0$) are modeled with logistic regression, with the shorthand notation being $\hat{m}_{0,i}=m_0(X_i;\hat{\beta})$. If Cox models are used as in the main text (Appendix), one may instead use the estimating function

$$U_i(\hat{\mu}_{10,\mathrm{DR}}, \hat{\alpha}, \hat{\beta}) = \frac{I[C_i \neq T_i^*]}{\hat{K}_C(T_i^*)}(1-A_i)\frac{\hat{\pi}_i}{1-\hat{\pi}_i}(Y_i - \hat{\mu}_{10,\mathrm{DR}})$$
$$-\left\{(1-A_i)\frac{\hat{\pi}_i}{1-\hat{\pi}_i}[\hat{R}_{0,i}(t) - \hat{\mu}_{10,\mathrm{DR}}] - A_i[\hat{R}_{0,i}(t) - \hat{\mu}_{10,\mathrm{DR}}]\right\},$$

and hence the sandwich estimator $(1/n)\mathrm{B}^{-1}\mathrm{M}\mathrm{B}^{-1,\mathrm{T}}$ is

$$\mathrm{B} = \hat{\mathrm{E}}\left\{\left.\frac{\partial U(\mu_{10}, \hat{\alpha}, \hat{\beta})}{\partial \mu_{10}}\right|_{\mu_{10}=\hat{\mu}_{10,\mathrm{DR}}}\right\} = -\frac{I[C_i \neq T_i^*]}{\hat{K}_C(T_i^*)}(1-A_i)\frac{\hat{\pi}_i}{1-\hat{\pi}_i} + \left\{(1-A_i)\frac{\hat{\pi}_i}{1-\hat{\pi}_i} - A_i\right\},$$

$$\mathrm{M} = n^{-1}\sum_i \left\{U_i(\hat{\mu}_{10,\mathrm{DR}}, \hat{\alpha}, \hat{\beta}) + Q_{\alpha,i} + Q_{\beta,i}\right\}^2, \text{ in which}$$

$$Q_{\alpha,i} = \hat{\mathrm{E}}\left[\frac{I[C_i \neq T_i^*]}{\hat{K}_C(T_i^*)}(1-A_i)\frac{\hat{\pi}_i}{1-\hat{\pi}_i}[Y_i - \hat{R}_{0,i}(t)]\mathbf{X}_i^{\mathrm{T}}\right]\hat{\mathrm{E}}\left[\hat{\pi}_i(1-\hat{\pi}_i)\mathbf{X}_i\mathbf{X}_i^{\mathrm{T}}\right]^{-1}(A_i - \hat{\pi}_i)\mathbf{X}_i,$$

$$Q_{\beta,i} = \hat{\mathrm{E}}\left[\frac{\hat{\pi}_i - A_i}{1-\hat{\pi}_i}(1-\hat{R}_{0,i}(t))\log(1-\hat{R}_{0,i}(t))X_i^{\mathrm{T}}\right]\hat{\Im}(\hat{\beta})^{-1}(1-A_i)\Delta_i(\hat{\beta}),$$

where $\hat{R}_{0,i}(t) = 1 - \hat{S}(t \mid A=0, X=0; \hat{\beta})^{\exp(\hat{\beta}^{\mathrm{T}} X_i)}$ is the product-limit estimator (Kalbfleisch and Prentice, 2002, pp. 114–116) and $\hat{\Im}(\hat{\beta})$ is an observed information matrix and $\Delta_i(\hat{\beta})$ is a score residual defined by

$$Y_i\left\{X_i - \frac{\sum_{j \text{ in risk set at } T_i} X_j \exp(X_j^{\mathrm{T}}\hat{\beta})}{\sum_{j \text{ in risk set at } T_i} \exp(X_j^{\mathrm{T}}\hat{\beta})}\right\}$$
$$-\sum_{k \text{ s.t. } i \text{ in risk set at } T_k} Y_k \frac{\exp(X_i^{\mathrm{T}}\hat{\beta})}{\sum_{j \text{ in risk set at } T_k} \exp(X_j^{\mathrm{T}}\hat{\beta})}\left\{X_i - \frac{\sum_{j \text{ in risk set at } T_k} X_j \exp(X_j^{\mathrm{T}}\hat{\beta})}{\sum_{j \text{ in risk set at } T_k} \exp(X_j^{\mathrm{T}}\hat{\beta})}\right\}.$$

Despite their lengthy algebraic form, the inverse of the observed information matrix and the score residual parts are obtained as building blocks of the robust sandwich formula in off-the-shelf software (eg, PHREG in SAS).

The asymptotic variance of risk difference and ratio estimators is estimated through their influence functions. For standardized risk difference, the influence functions of $\hat{\gamma}_{\text{IPW-RW}} = \hat{\mu}_{11} - \hat{\mu}_{10,\text{IPW-RW}}$ and $\hat{\gamma}_{\text{DR}} = \hat{\mu}_{11} - \hat{\mu}_{10,\text{DR}}$ for member i are $\phi_{11,i} - \phi_{10,\text{IPW-RW},i}$ and $\phi_{11,i} - \phi_{10,\text{DR},i}$, respectively. For standardized risk ratios $\exp(\hat{\delta}_{\text{IPW-RW}}) = \hat{\mu}_{11} / \hat{\mu}_{10,\text{IPW-RW}}$ and $\exp(\hat{\delta}_{\text{DR}}) = \hat{\mu}_{11} / \hat{\mu}_{10,\text{DR}}$, the corresponding influence functions are obtained as $(1/\mu_{10})\phi_{11,i} - (\mu_{11}/\mu_{10}^2)\phi_{10,\text{IPW-RW},i}$ and $(1/\mu_{10})\phi_{11,i} - (\mu_{11}/\mu_{10}^2)\phi_{10,\text{DR},i}$ by applying the Taylor series expansion. The asymptotic variance estimates are obtained using the sample variance of these influence functions, which is evaluated at the estimate of each parameter. For risk ratios, in particular, logarithmic transformation would improve approximation to normal distribution; using the delta method, the asymptotic variance of log risk ratio $\hat{\delta}$ is obtained as $\delta^{-2}\,\text{var}[\exp(\hat{\delta})] = \text{var}(\phi_{11,i}/\mu_{11} - \phi_{10,i}/\mu_{10})$, where $\text{var}[\exp(\hat{\delta})]$ is the variance of the influence functions for the risk ratio.

3. Double-robustness property in the presence of censored outcome

A weighted Kaplan–Meier estimator is the maximizer of the log pseudo-likelihood that is the observed log likelihood with corresponding weights (Xie and Liu, 2005). Thus, our weighted estimate $\hat{\mu}_{10,\text{IPW-RW}}(t)$ is regarded as the Kaplan–Meier estimate in the pseudo-population with member i's weight $\dfrac{\pi(X_i;\hat{\alpha})(1-A_i)}{1-\pi(X_i;\hat{\alpha})}$, which therefore converges to

$$\mathrm{E}\left[\frac{\pi(X;\alpha^*)(1-A)}{1-\pi(X;\alpha^*)}Y\right]\Big/\mathrm{E}\left[\frac{\pi(X;\alpha^*)(1-A)}{1-\pi(X;\alpha^*)}\right],$$

where α^* is a limiting value of $\hat{\alpha}$ in the working model $\pi(X;\alpha)$. Conversely, because the product-limit estimator $\hat{S}(t \mid A=0, X=0; \hat{\beta})$ is a maximum likelihood estimator under a Cox model $\Lambda(s|A = 0, X = x; \beta)$ (Kalbfleisch and Prentice, 2002), a limiting value $S_{0t}^*(x; \beta^*)$ of $\hat{S}(t \mid A=0, x; \hat{\beta}) = \hat{S}(t \mid A=0, X=0; \hat{\beta})^{\exp(\hat{\beta}^{\mathrm{T}}x)}$ with a limiting value β^* of $\hat{\beta}$ is present in the working Cox model.

First, suppose that the PS model $\pi(X;\alpha)$ is correctly specified; then, $\pi(x; \alpha^*)$ is equal to $\pi(x) = \Pr[A = 1|X = x]$ and $\hat{\mu}_{10,\text{IPW-RW}}(t)$ is also shown to converge to $\mathrm{E}[Y^0|A = 1] = \mu_{10}$ (Hernán and Robins, 2006). Similarly, the augmentation term $\hat{\mu}_{10,\text{aug}}(t)$ can be shown to converge to

$$\begin{aligned}
&\mathrm{E}\left[1-S_{0t}^*(X;\beta^*)\middle|A=1\right]-\frac{\mathrm{E}\left[(1-A)\dfrac{\pi(X)}{1-\pi(X)}\left\{1-S_{0t}^*(X;\beta^*)\right\}\right]}{\mathrm{E}\left[(1-A)\dfrac{\pi(X)}{1-\pi(X)}\right]}\\
&=\frac{\mathrm{E}\left[A\left\{1-S_{0t}^*(X;\beta^*)\right\}\right]}{\Pr(A=1)}-\frac{\mathrm{E}\left[\pi(X)\left\{1-S_{0t}^*(X;\beta^*)\right\}\right]}{\mathrm{E}\left[\pi(X)\right]}\\
&=0.
\end{aligned}$$

Therefore, the consistency of $\hat{\mu}_{10,\text{DR}}(t) = \hat{\mu}_{10,\text{IPW-RW}}(t) + \hat{\mu}_{10,\text{aug}}(t)$ is confirmed.

Second, suppose that the Cox model $\Lambda(s|A = 0, X = x; \beta)$ is correctly specified; then, $1 - S_{0t}^*(x; \beta^*)$ is equal to $1 - S(t|A = 0, X = x) = \mathrm{E}[Y| A = 0, X = x] = \mathrm{E}[Y^0|X = x] = \mathrm{E}[Y^0|A = 1, X = x]$. Thus, $\hat{\mu}_{10,\text{DR}}(t)$ converges to

$$\frac{\mathrm{E}\left[\frac{\pi(X;\alpha^*)(1-A)}{1-\pi(X;\alpha^*)}Y\right]}{\mathrm{E}\left[\frac{\pi(X;\alpha^*)(1-A)}{1-\pi(X;\alpha^*)}\right]}+\mathrm{E}\big[\mathrm{E}(Y^0\mid A=1,X)\big|A=1\big]-\frac{\mathrm{E}\left[\frac{\pi(X;\alpha^*)(1-A)}{1-\pi(X;\alpha^*)}\mathrm{E}(Y\mid A=0,X)\right]}{\mathrm{E}\left[\frac{\pi(X;\alpha^*)(1-A)}{1-\pi(X;\alpha^*)}\right]}$$

$$=\mathrm{E}\big[Y^0\big|A=1\big]+\frac{\mathrm{E}\left[\frac{\pi(X;\alpha^*)}{1-\pi(X;\alpha^*)}\left\{\mathrm{E}[(1-A)Y\big|X]-\mathrm{E}[1-A\big|X]\,\mathrm{E}[Y\mid A=0,X]\right\}\right]}{\mathrm{E}\left[\frac{\pi(X;\alpha^*)(1-A)}{1-\pi(X;\alpha^*)}\right]}$$

$$=\mu_{10}+\frac{\mathrm{E}\left[\frac{\pi(X;\alpha^*)}{1-\pi(X;\alpha^*)}\left\{\mathrm{E}[Y\big|A=0,X]\Pr(A=0\big|X)-(1-\pi(X))\,\mathrm{E}[Y\mid A=0,X]\right\}\right]}{\mathrm{E}\left[\frac{\pi(X;\alpha^*)(1-A)}{1-\pi(X;\alpha^*)}\right]}$$

$$=\mu_{10}.$$

Therefore, the double robustness of $\hat{\mu}_{10,\mathrm{DR}}(t)$ is demonstrated.

$\hat{\mu}_{10,\mathrm{DR}}(t)$ can be regarded as a censored-outcome analog of the “B-DR” estimator for total population of Robins et al. (2007, p. 546), which takes the form

$$\hat{\mathrm{E}}[m_0(X;\hat{\beta})]+\frac{\hat{\mathrm{E}}\left[\frac{1-A}{1-\pi(X;\hat{\alpha})}\left\{Y-m_0(X_i;\hat{\beta})\right\}\right]}{\hat{\mathrm{E}}\left[\frac{1-A}{1-\pi(X;\hat{\alpha})}\right]}.$$

When targeting the exposed population, the formula can be modified as

$$\hat{\mathrm{E}}[m_0(X;\hat{\beta})\big|A=1]+\frac{\hat{\mathrm{E}}\left[\frac{\pi(X;\hat{\alpha})}{1-\pi(X;\hat{\alpha})}(1-A)\left\{Y-m_0(X_i;\hat{\beta})\right\}\right]}{\hat{\mathrm{E}}\left[\frac{\pi(X;\hat{\alpha})}{1-\pi(X;\hat{\alpha})}(1-A)\right]}.$$

When $m_0(X_i;\beta)$ is replaced with the Cox model and the sample means $\hat{\mathrm{E}}[\cdot]$ is replaced with the Kaplan–Meier (product-limit) estimates at *t*, the “B-DR” estimator reduces to

$\hat{\mu}_{10,\mathrm{DR}}(t)$.

4. Details of simulation

First, covariates X_1 (continuous) and X_2 (binary) are generated by using standard normal distribution and Bernoulli distribution with a success probability of 0.5, respectively. Exposure A is assigned according to the true PS of $\Pr(A = 1| X_1, X_2) = \{1 + \exp(-\log(0.5) - \log(2)X_1 - \log(4)X_2 - \log(2)X_1X_2)\}^{-1}$. Potential time-to-event T^a follows an exponential distribution with a rate parameter $X^{\mathrm{T}}\beta_0 = \log(0.1) + \log(2)X_1 + \log(2)X_2 + \log(2)X_1X_2$ under unexposed (T^0) and $X^{\mathrm{T}}\beta_1 = \log(0.4) + \log(4)X_1 + \log(4)X_2 + \log(2)X_1X_2$ under exposed (T^1); this is equivalent to the model $\log T_i^a = -X^{\mathrm{T}}\beta_a + e_i$, where e_i follows extreme value distribution. Time to be censored C is also generated independently for each member by using exponential distribution with a unit rate parameter. Observed time is obtained as $T^* = \min(AT_1 + (1 - A)T_0, C, 1)$ and event indicator Y^* (1 if $T^* = AT_1 + (1 - A)T_0$, 0 otherwise). Consequently, we must estimate the parameters from observed data (X_{1i}, X_{2i}, A_i, T_i^*, Y_i^*), for $i = 1,..., n$. The size of each dataset is $n = 2{,}000$ and estimation is performed in 2,000 Monte Carlo samples. A series of simulation studies are conducted by SAS version 9.4; SAS program for generating simulation data is provided below.

From a large dataset ($n = 10{,}000{,}000$) generated as above, "true" targeted parameters are searched using unobservable outcomes Y^1 (1 if $T^1 < t = 1$, 0 otherwise) and Y^0 (1 if $T^0 < t = 1$, 0 otherwise). Appendix Table summarizes these parameter values (and also provides, as reference, the corresponding parameters in total population). In this dataset, the proportion of members censored before $t = 1$ without event is 0.32 among the exposed (0.47 in total population): if the censored members are treated as nonevents, the risk in the exposed would be underestimated by one-third.

We fit the correct and incorrect PS models

$$\text{(Correct) } \Pr(A = 1 | X_1, X_2; \alpha) = \{1 + \exp(-\alpha_0 - \alpha_1 X_1 - \alpha_2 X_2 - \alpha_{12} X_1 X_2)\}^{-1},$$

$$\text{(Incorrect) } \Pr(A = 1 | X_1, X_2; \alpha) = \{1 + \exp(-\alpha_0 - \alpha_1 X_1 - \alpha_2 X_2)\}^{-1},$$

and the correct and incorrect proportional hazards models for T in $A = 0$

$$\text{(Correct) } \Lambda(t|A = 0, X_1, X_2; \beta) = \Lambda_0(t)\exp(\beta_1 X_1 + \beta_2 X_2 + \beta_{12} X_1 X_2),$$

$$\text{(Incorrect) } \Lambda(t|A = 0, X_1, X_2; \beta) = \Lambda_0(t)\exp(\beta_1 X_1 + \beta_2 X_2),$$

where $\Lambda_0(t)$ is the parametrically unspecified baseline cumulative hazard function. The true model for E[log $T|A = 0, X_1, X_2$] is an exponential regression model, and the aforementioned proportional hazards model includes the true model if the product term X_1X_2 is adjusted for (Kalbfleisch and Prentice, 2002).

Appendix Table. Simulation Parameters Defined in a Large Sample (n = 10,000,000)

Target Population	**E(Y^1|Target) at t = 1**	**E(Y^0|Target) at t = 1**	**Causal RD in Target**	**Log (Causal RR in Target)**	**Probability of Censoring Before t = 1**
Total	0.534	0.195	0.339	1.006	0.465
Exposed	0.701	0.285	0.417	0.902	0.322

RD indicates risk difference; RR, risk ratio.

Summary of simulation results (Tables 2 and 3):

- As expected, the regression model-based and the IP weighted–reweighted estimators produced unbiased estimates only with correct model specification for outcome and propensity score, respectively, whereas the proposed doubly robust estimator yielded unbiased estimates when at least one of these model specifications was correct.
- If the outcome model is correctly specified, regression estimator is the most efficient among competitors, followed by doubly robust estimator in each case.
- Coverage proportions maintain almost nominal 95% probability in the settings where bias is negligible.
- Because the influence function for the target parameter depends only on how the nuisance parameters are estimated, and not on whether the working models (ie, PS and outcome regression models) including the nuisance parameters are correct or not (Tsiatis, 2006), the estimated asymptotic standard errors incorporating maximum likelihood estimates of the nuisance parameters suitably approximate empirical (Monte Carlo) standard errors even if the models are incorrectly specified.

References in the Online Supplemental Material

1. Hernán MA, Robins JM. Estimating causal effects from epidemiological data. *J Epidemiol Community Health*. 2006;60:578–586.
2. Kalbfleisch JD, Prentice RL. *The Statistical Analysis of Failure Time Data*. 2nd ed. New York, NY: Wiley; 2002.
3. Kang JDY, Schafer JL. Demystifying double robustness: a comparison of alternative

strategies for estimating a population mean from incomplete data. *Stat Sci*. 2007;22:523–539.

4. Robins JM, Sued M, Lei-Gomez Q, et al. Comment: performance of double-robust estimators when “inverse probability” weights are highly variable. *Stat Sci*. 2007;22:544–559.
5. Sato T, Matsuyama Y. Marginal structural models as a tool for standardization. *Epidemiology*. 2003;14:680–686.
6. Satten GS, Datta S. The Kaplan-Meier estimator as an inverse-probability-of-censoring weighted average. *Am Statist*. 2001;55:207–210.
7. Tan Z. Nonparametric likelihood and doubly robust estimating equations for marginal and nested structural models. *Can J Stat*. 2010;38:609–632.
8. Tsiatis AA. *Semiparametric Theory and Missing Data*. New York: Springer; 2006.
9. Xie J, Liu C. Adjusted Kaplan–Meier estimator and log-rank test with inverse probability of treatment weighting for survival data. *Stat Med*. 2005;24:3089–3110.

5. SAS program for censored outcome data

```
/* ---------- Program for Point Estimation ---------- */
%let dsn = /* dataset name (eg, TAMDATA) */;
%let exp = /* exposure variable (eg, TAM) */;
%let target = /* targeted exposure level (eg, 1) */;
%let time = /* observation time variable (eg, RECTIME) */;
%let event = /* event indicator (eg, REC) */;
%let t = /* time at which risk is calculated (eg, 5) */;
%let PSmodel = /* covariates for propensity score model (eg, AGE STAGE LYM MENO) */;
%let ORmodel = /* covariates for outcome regression model (eg, AGE STAGE LYM MENO)
*/;

/* ----- 0. Nonparametric Kaplan-Meier estimate ----- */
ods listing close;
proc phreg data = &dsn;
  model &time * &event(0) = ;
  where &exp = &target;
  baseline out = mu11 survival = _all_ / method = PL;
run;
/* NOTE: PHREG has an advantage on computing time over LIFETEST. */
ods listing;

data mu11;
  set mu11 end = end;
  where &time < &t;
  if end;
  mu11 = 1 - survival;
run;

/* ----- 1. Model-based maximum likelihood estimator ----- */
data &dsn._ML;
  set &dsn;
  targ = (&exp = &target);
```

```
  MLwt = 1 - targ;
/* NOTE: Trick to estimate baseline risk for members in the target in 1-step from
the model fitted in the untargeted. */

ods listing close;
proc phreg data = &dsn._ML;
   model &time * &event(0) = &ORmodel;
   weight MLwt;
   output out = &dsn._ML xbeta = xb;
run;

data cov;
  xb = 0;

proc phreg data = &dsn._ML;
  where MLwt = 1;
  model &time * &event(0)= /offset=xb;
  baseline out = base covariates = cov survival = _all_ / method = PL;
run;

data _null_;
  set base end = end;
  where &time < &t;
  if end then call symput("S0", Survival);
run;

data &dsn._ML;
  set &dsn._ML;
  R0i = 1 - &S0 ** exp(xb);
run;

proc means data = &dsn._ML;
  var R0i;
```

```
  where targ = 1;
  output out = mu10ML mean = mu10ML;
run;
ods listing;

data ML;
  merge mu11(keep=mu:) mu10ML(keep=mu:);
  RD_ML = mu11 - mu10ML;
  RR_ML = mu11 / mu10ML;
run;

/* One can instead use Breslow applox. for getting interval estimates automatically
*/
data cov;
  set &dsn.;
  where &exp = &target;
run;

proc phreg data = &dsn.;
  where &exp = 1 - &target;
  model &time * &event(0) = &ORmodel;
  baseline out = base covariates = cov survival = _all_ / diradj;
run;

data mu10ML2;
  set base end = end;
  where &time < &t;
  mu10ML = 1 - survival;
  se10ML = StdErrSurvival;
  if end;
run;

data ML2;
```

```
  format mu11 se11 mu10ML se10ML RD_ML seRD_ML logRR_ML selogRR_ML;
  merge mu11(rename = (StdErrSurvival = se11) keep = mu: StdErrSurvival) base(keep=
mu: se:);
  RD_ML = mu11 - mu10ML;
  logRR_ML = log(mu11 / mu10ML);
  seRD_ML = sqrt(se11**2 + se10ML**2);
  selogRR_ML = sqrt(mu11**(-2) * se11**2 + mu10ML**(-2) * se10ML**2);
/* NOTE: mu10ML is estimated through parameter estimates of a model fitted in the
unexposed and is independent of mu11, which is estimated in the exposed. */
run;

/* ----- 2. IPW-RW estimator ----- */
ods listing close;
proc logistic data = &dsn desc;
  model &exp = &PSmodel;
  output out = &dsn._SM p = pi;
run;

data &dsn._SM;
  set &dsn._SM;
  SMW = abs((&exp - pi)/(1 - pi));
run;

proc lifetest data = &dsn._SM outsurv = mu10SM;
  time &time * &event(0);
  where &exp = 1 - &target;
  weight SMW;
run;
/* NOTE: The implementation of weights in PROC LIFETEST is based on Xie and Liu (2005,
2011). On the contrary, weights in PROC PHREG is only available for Breslow estimator.
*/

ods listing;
```

```
data mu10SM;
  set mu10SM end = end;
  where &time < &t;
  if end;
  mu10SM = 1 - survival;
run;

data SM;
  merge mu11(keep=mu:) mu10SM(keep=mu:);
  RD_SM = mu11 - mu10SM;
  RR_SM = mu11 / mu10SM;
run;

/* ----- 3. DR estimator ----- */
data &dsn._DR;
  merge &dsn &dsn._ML(keep = R0i) &dsn._SM(keep = pi SMW);

ods listing close;
proc means data = &dsn._DR;
  var R0i;
  weight SMW;
  where &exp = 1 - &target;
  output out = mu10aug mean = mu10ML_SM;
run;
ods listing;

data mu10DR;
  merge mu10SM mu10ML mu10aug;
  mu10DR = mu10SM + mu10ML - mu10ML_SM;

data DR;
  merge mu11(keep=mu:) mu10DR(keep=mu:);
  RD_DR = mu11 - mu10DR;
```

```
  RR_DR = mu11 / mu10DR;
run;

/* ---------- Program for Interval Estimation of Semiparametric Estimators ----------
*/

/* First, calculate the probability of censoring at each observed time (Ti) */
data &dsn._ipc;
  set &dsn.;
  censtime =min(&time, &t);
  if &time < &t and &event = 0 then cens = 1;
  else cens = 0;
run;

ods listing close;
proc phreg data = &dsn._ipc;
  model censtime * cens(0) = ;
  output out = &dsn._IPC survival = pc /method = PL;
run;
ods listing;

/* Get numbers (total and exposed sample sizes) for the influence function (IF) of
mu11 (= phi11) */
ods listing close;
proc means data=&dsn.;
  var &exp;
  output out =_ sum = n1 n = n;
run;
ods listing;

data _null_;
  set _;
  n&target = n1*(&target) + (n - n1)*(1 - &target);
```

```
  call symput("n", n);
  call symput("n&target", n&target);
run;
%put &n;
%put &&n&target;

/* ----- 1. IF for IPW-RW estimator ----- */
data _null_;
  set sm;
  call symput("mu11", mu11);
  call symput("mu10SM", mu10SM);
run;

data &dsn._iml;
  merge &dsn._sm &dsn._ipc(keep = pc cens);
  int = 1;
  Yt = (&event = 1)*(&time < &t);
  IPCW = 1/pc;
run;

proc iml;
  use &dsn._iml;
  read all var {int &PSmodel} into X; *nxp;
  read all var {SMW} into SMW; *nx1;
  read all var {pi} into pi; *nx1;
  read all var {cens} into cens; *nx1;
  read all var {IPCW} into IPCW; *nx1;
  read all var {Yt} into Yt; *nx1;
  read all var {&exp} into A; *nx1;
  U = (1 - cens) # IPCW # (1 - A) # SMW # (Yt - &mu10SM); *nx1;
  dU_dmu10 = (-(1 - A) # SMW # (1 - cens) # IPCW) [:];
  dU_dalpha = ((1 - cens) # IPCW # (1 - A) # SMW # (Yt - &mu10SM))` * X; *1xp;
  Iinv = ginv((pi # (1 - pi) # X)` * X); *pxp;
```

```
  score = (A - pi) # X; *nxp;
  phi10SM = ginv(dU_dmu10) * (U + ((dU_dalpha * Iinv) * score`)`);
  name = {'phi10SM'};
  create phi10SM from phi10SM[colname = name];
  append from phi10SM;
  close phi10SM;
quit;

data &dsn._iml;
  merge &dsn._iml phi10SM;
  phi11 = &exp * (Yt - &mu11) * (&n/&&n&target)/sqrt(&n) * (1 - cens) * IPCW;
  phi10SM = phi10SM/sqrt(&n);
  phiRD_SM = phi11 - phi10SM;
  philogRR_SM = (phi11/&mu11) - (phi10SM/&mu10SM);
run;

proc means data = &dsn._iml N mean var std;
  var phi11 phi10SM phiRD_SM philogRR_SM;
run;

/* ----- 2. IF for DR estimator ----- */
data _null_;
  set dr;
  call symput("mu11", mu11);
  call symput("mu10DR", mu10DR);
run;

data &dsn._iml2;
  merge &dsn._DR &dsn._ipc(keep = pc cens);
  int = 1;
  Yt = (&event = 1)*(&time < &t);
  IPCW = 1/pc;
run;
```

```
ods listing close;
ods output covb = cov;
proc phreg data = &dsn._ML;
  model &time * &event(0) = &ORmodel / covb;
  weight MLwt;
  output out = res ressco = res1-res1000;
run;

proc contents data = res(keep = res:) out = nvar;
run;
ods listing;

data _null_;
  set nvar;
  call symput("nvar", _n_);
run;

data res;
  set res(keep = res:);
  array res{*} res1 - res%eval(&nvar);
  do i = 1 to &nvar;
    if res{i} = . then res{i} = 0;
  end;
  drop i;
run;

proc iml;
  use &dsn._iml2;
  read all var {int &PSmodel} into X_PS; * nx(p+1);
  read all var {&ORmodel} into X_OR; * nxp;
  read all var {SMW} into SMW; * nx1;
  read all var {pi} into pi; * nx1;
```

```
  read all var {R0i} into R0; * nx1;
  read all var {cens} into cens; * nx1;
  read all var {IPCW} into IPCW; * nx1;
  read all var {Yt} into Yt; * nx1;
  read all var {&exp} into A; * nx1;
  use cov;
  read all var {&ORmodel} into cov;
  use res;
  read all into res;
  U = (1 - cens) # IPCW # (1 - A) # SMW # (Yt - &mu10SM) - ((1 - A) # SMW # (R0 -
&mu10DR) - A # (R0 - &mu10DR) ); * nx1;
  dU_dmu10 = (-(1 - A) # SMW # (1 - cens) # IPCW + (1 - A) # SMW - A) [:]; * 1x1;
  dU_dalpha = ((1 - cens) # IPCW # (1 - A) # SMW # (Yt - R0))` * X_PS; * 1x(p+1);
  Iinv_alpha = ginv((pi # (1 - pi) # X_PS)` * X_PS); * 1x(p+1);
  score_alpha = (A - pi) # X_PS; * 1x(p+1);
  S0 = max(1 - R0, 0.0000000001);
  dU_dbeta = ((pi - A) # 1/(1 - pi) # (S0) # log(S0))` * X_OR; * 1xp;
/* NOTE: Avoiding undefined operation log(0). Because limit xlog(x) = 0 (as x goes
to +0), dU_dbeta of those with S0 = 1 goes to 0 */
  phi10DR = ginv(dU_dmu10) * (U + ((dU_dalpha * Iinv_alpha) * score_alpha`)` +
((dU_dbeta * cov) * res`)`);
  name = {'phi10DR'};
  create phi10DR from phi10DR[colname = name];
  append from phi10DR;
  close phi10DR;
quit;

data &dsn._iml2;
  merge &dsn._iml2 phi10DR;
  phi11 = &exp * (Yt - &mu11) * (&n/&&n&target)/sqrt(&n) * (1 - cens) * IPCW;
  phi10DR = phi10DR/sqrt(&n);
  phiRD_DR = phi11 - phi10DR;
  philogRR_DR = (phi11/&mu11) - (phi10DR/&mu10DR);
```

```
run;

proc means data = &dsn._iml2 N mean var std;
  var phi11 phi10DR phiRD_DR philogRR_DR;
run;

/* ---------- Program for Simulating Data ---------- */
* Simulation size;
%let nsamp = 2000;
%let sampsiz = 2000;
* Time t;
%let t = 1;
* PS model;
%let alpha0 = log(0.5);
%let alpha1 = log(2);
%let alpha2 = log(4);
%let alpha12 = log(2);
* OR model;
%let beta0 = log(0.1);
%let beta1 = log(2);
%let beta2 = log(2);
%let beta12 = log(2);
%let betaA = log(4); * Exposure effect;
%let betaA1 = log(2); * Exposure effect;
%let betaA2 = log(2); * Exposure effect;
%let betaA12 = log(1); * Exposure effect;

data sim;
call streaminit(20141203);
        do replicate = 1 to &nsamp;
        do i = 1 to &sampsiz;

        * Covariates;
```

```
        X1 = rand("Normal");
        X2 = rand("Bernoulli", 0.5);

        * Exposure;
        do A = rand("Bernoulli", 1/(1 + exp(- &alpha0 - &alpha1*X1 - &alpha2*X2 -
&alpha12*X1*X2)));

        * Outcome;
        UT = rand("Uniform");
        ErrT = log(-log(1 - UT));
        T0 = exp(- &beta0 - &beta1*X1 - &beta2*X2 - &beta12*X1*X2 + ErrT);
        T1 = exp(- &beta0 - &beta1*X1 - &beta2*X2 - &beta12*X1*X2 - &betaA - &betaA1*X1
- &betaA2*X2 - &betaA12*X1*X2 + ErrT);
        Y0= (T0 < &t);
        Y1= (T1 < &t);
        T = (A = 1)*T1 + (A = 0)*T0;
        C = rand("Exponential", 1);
        time = min(T, C, &t);
        event = (time = T);

        output;
        end;
        end;
        end;
run;
```