1. Web Appendix A: Causal Assumptions

We use a counterfactual framework for defining causal effects. Let $T_i(z)$ be the potential time to the first event when subject i is assigned treatment z, so that $T_i = (1-Z_i)T_i(0)+Z_iT_i(1)$; similarly, define $J_i(z)$ as the potential type of event under this same treatment assignment. We assume that the treatment assignment of the ith individual does not affect the potential outcomes of other individuals and that there are no multiple forms of treatment, i.e. that the single unit treatment value assumption (SUTVA) holds. We also adopt the axiom of consistency: if a unit is assigned to treatment z, then it must be that $(T_i, J_i) = (T_i(z), J_i(z))$. The full data we would like to see for each subject are denoted $X_i(z) = (T_i(z), J_i(z), W_i)$.

Our goal is estimation of the cumulative incidence of events of type $J = 1$ at a fixed time t_0 under fixed treatment assignment z_0:

$$F_{z_0}(t_0, 1) = P(T(z_0) \leqslant t_0, J(z_0) = 1) \quad \text{for} \quad z_0 = 0, 1. \tag{1}$$

To identify this causal quantity based on the observed data, we must make several non-testable assumptions:

(1) Conditionally independent treatment assignment: $(Z \perp T(z)) \mid W$ for $z = 0, 1$

(2) Positivity: $P(U(z) > t_0 \mid Z = z, W = w) > 0$ for$P_0 -$ almost every w and for $z = 0, 1$

(3) Coarsening at random (CAR): $P_{O|Z=z,X(z)} = P_{O|Z=z,W}$ for $z = 0, 1$

Assumption 1 would be satisfied in any trial where treatment probabilities are known based on baseline covariates W, as would be in a randomized trial. Assumption 2 calls for a positive probability of remaining uncensored up to time t_0 within each strata of W; this ensures counterfactual event times are well defined (Petersen et al., 2010). Assumption 3 requires there be no unmeasured confounders of $T(z)$ and $U(z)$ (van der Laan and Robins, 2003; Tsiatis, 2007). If we are unwilling to make untestable assumptions, as perhaps we would be

in an observational study, the statistical parameter defined in Section 2 Equation (1) may still be an interesting parameter for assessing treatment efficacy.

2. Web Appendix B: Proof of Theorem 1

We provide two derivations of the efficient influence curve: the first via projections onto tangent spaces generated by fluctuation submodels for Q, the second as a direct application of Theorem 1 in van der Laan and Gruber (2012). These two proofs arrive at different, but equivalent representations of the efficient influence curve.

2.1 *Representation 1*

The first method of obtaining the canonical gradient $D^*(P)$ of the pathwise derivative of $\Psi : \mathcal{M} \rightarrow [0,1]$ at P involves projecting an initial gradient, $D_{IPCW}(P)$, onto the tangent space of the model at P, denoted $T(P)$. Because Ψ is only a function of P through the Q portion of the likelihood, we can consider the submodel $\mathcal{M}(g)$ of $\mathcal{M}$ where g is treated as fixed and known. Due to the factorization of $P = Qg$, it follows that the canonical gradient in this smaller model will be the same as in the full model (van der Laan and Robins, 2003). The tangent space in the reduced model is a sum space consisting of the tangent spaces generated by submodels $P(\epsilon)$ that only vary the conditional distributions of W, $dN_1(t)$, and $dN_2(t)$; we denote these tangent spaces as $T_{dN_1(t)}(P), T_{dN_2(t)}(P), t = 1, ..., \tau$, and $T_W(P)$, respectively.

These tangent spaces are defined as the mean-square closure of the space spanned by the scores of regular parametric submodels through $P(O(t) \mid Pa(O(t)))$. Thus, we can write the

tangent spaces as

$$\begin{aligned} T_W(P) &= \{v(W) \in L_2^0(P) : E[v(W)] = 0\} \\ T_{dN_1(t)}(P) &= \{v(dN_1(t), Pa(dN_1(t))) \in L_2^0(P) : \\ &\qquad E[v(dN_1(t), Pa(dN_1(t))) \mid Pa(dN_1(t))] = 0\} \\ T_{dN_2(t)}(P) &= \{v(dN_2(t), Pa(dN_2(t))) \in L_2^0(P) : \\ &\qquad E[v(dN_2(t), Pa(dN_2(t))) \mid Pa(dN_2(t))] = 0\}, \end{aligned}$$

where $L_2^0(P)$ is the Hilbert space of mean zero functions with finite variance with respect to P equipped with inner product $< f, g >:= P(fg)$. The projection operator is given by

$$\begin{aligned} &\Pi(D(P) \mid T_{O(t)}(P))(o(t), pa(o(t))) \\ &\quad = E[D(P)(O) \mid O(t) = o(t), Pa(O(t)) = pa(o(t))] - E[D(P)(O) \mid Pa(O(t)) = pa(o(t))]. \end{aligned}$$

For the binary variables $dN_1(t), dN_2(t)$, this projection operator takes the special form

$$\Pi(D(P) \mid T_{O(t)}(P)) = H_{O(t)}(Pa(O(t)))[O(t) - P(O(t) = 1 \mid Pa(O(t)))], \tag{2}$$

where

$$\begin{aligned} H_{O(t)}(pa(o(t))) &:= E[D(P)(O) \mid O(t) = 1, Pa(O(t)) = pa(o(t))] \\ &\quad -E[D(P)(O) \mid O(t) = 0, Pa(O(t)) = pa(o(t))]. \end{aligned} \tag{3}$$

See van der Laan and Rose (2011) (p 538-542) for proofs of these results.

We note that this derivation uses a parameterization of the distribution of $(dN_1(t), dN_2(t))$ considering the distribution of $dN_1(t)$ given $Pa(dN_1(t))$ and of $dN_2(t)$ given $Pa(dN_2(t))$ and $dN_1(t)$. A different, but equivalent representation could be arrived at by changing the ordering of this parameterization.

Now consider the gradient in $\mathcal{M}(g)$ at P evaluated at a standard data unit o,

$$D_{IPCW}(P)(o) = H_g(t_0, o(t_0 - 1))n_1(t_0) - F(t_0, 1 \mid Z = z_0), \tag{4}$$

where $H_g(t_0, o(t_0-1))$ is as defined in Theorem 1. We now project this gradient onto $T_W(P), T_{dN_1(t)}(P)$, and $T_{dN_2(t)}(P)$.

The projection onto $T_W(P)$ is easily seen to be

$$\begin{aligned}\Pi[D_{IPCW}(P) \mid T_W(P)](o) &= E[D_{IPCW}(P)(o) \mid W = w] \\ &= F(t_0, 1 \mid Z = z_0, W = w) - F(t_0, 1 \mid Z = z_0).\end{aligned}$$

To calculate the projection onto $T_{dN_1(t)}(P)$, we must calculate (3) for $O(t) = dN_1(t)$. Note that (2) will be 0 if $N_1(s-1) = 1$, $N_2(s-1) = 1$, or $C(s-1) = 1$, for any $s < t$. It follows that we can condition on $N_1(t-1) = N_2(t-1) = C(t-1) = 0$ in our calculations; we denote this condition as $\tilde{T} \geqslant t$.

The first term in (3) for $t \leqslant t_0$ is

$$\begin{aligned}&E[D_{IPCW}(P)(o) \mid dN_1(t) = 1, \tilde{T} \geqslant t, Pa(dN_1(t)) = pa(dn_1(t))] \\ &= H_g(t, o(t-1)) - F(t_0, 1 \mid Z = z_0),\end{aligned}$$

For $t > t_0, E[D_{IPCW}(P)(O) \mid dN_1(t) = 1, Pa(dN_1(t)) = pa(dn_1(t))] = -F(t_0, 1 \mid Z = z_0)$. The second term in (3) for $t < t_0$ is,

$$\begin{aligned}&E[D_{IPCW}(P)(o) \mid dn_1(t) = 0, \tilde{T} \geqslant t, Pa(dn_1(t))] \\ &= H_g(t, o(t-1))(1 - \check{Q}_2^{z_0,0}(t, w)) \sum_{s=t+1}^{t_0} \bar{Q}_1^{z_0,0}(s, w) \prod_{m=t+1}^{s-1} (1 - \bar{Q}_1^{z_0,0}(m, w) - \bar{Q}_2^{z_0,0}(m, w)) \\ &= H_g(t, o(t-1))(1 - \check{Q}_2^{z_0,0}(t, w))R(t, w).\end{aligned}$$

For $t \geqslant t_0, E[D_{IPCW}(P)(O) \mid dN_1(t) = 0, \tilde{T} \geqslant t, Pa(dN_1(t)) = pa(dn_1(t))] = -F(t_0, 1 \mid Z = z_0)$. Combining these results, we conclude that

$$H_{dN_1(t)}(t, o(t-1)) = H_g(t, o(t-1))[I(t < t_0)[1 - (1 - \check{Q}_2^{z_0,0}(t, w))R(t, w)] + I(t = t_0)].$$

We can similarly use (2) and (3) to determine the projection of $D_{IPCW}(P)$ onto $T_{dN_2(t)}(P)$. The projection will be 0 when $N_1(t) = 1, N_2(t-1) = 1$, or $C(t-1) = 1$; we use $\{\tilde{T} \geqslant$

$t, dN_1(t) = 0\}$ to denote the complement of this condition. The first term in (3) is $-F(t_0, 1 \mid Z = z_0) \ \ \forall \ \ t \in \{1, ..., t_0\}$. The second term in (3) for $t < t_0$ is,

$$\begin{aligned}
&E[D_{IPCW}(P)(o) \mid dN_2(t) = 0, Pa(dN_2(t)) = pa(dN_2(t)), \tilde{T} \geqslant t, dn_1(t) = 0] \\
&\quad = H_g(t, o(t-1)) \sum_{s=t+1}^{t_0} \bar{Q}_1^{z_0,0}(s, w) \prod_{m=t+1}^{s-1} (1 - \bar{Q}_1^{z_0,0}(m, w) - \bar{Q}_2^{z_0,0}(m, w)) - F(t_0, 1 \mid Z = z_0) \\
&\quad = H_g(t, o(t-1))R(t, w) - F(t_0, 1 \mid Z = z_0).
\end{aligned}$$

For $t \geqslant t_0$, this term is $-F(t_0, 1 \mid Z = z_0)$. Combining these we conclude that

$$H_{dN_2(t)}(t, o(t-1)) = -I(t < t_0) H_g(t, (o-1))R(t, w)$$

Now, by summing over all the projections, we have the first form of the efficient influence curve:

$$\begin{aligned}
D^*(P)(o) \ \ = \ \ & \sum_{t=1}^{\tau} H_g(t, o(t-1)) H_{dN_1(t)}(t, o(t-1))(dn_1(t) - \bar{Q}_1^{z_0,0}(t, w)) \\
& + \sum_{t=1}^{\tau} H_g(t, o(t-1)) H_{dN_2(t)}(t, o(t-1))(dn_2(t) - \check{Q}_2^{z_0,0}(t, w)) \\
& + F(t_0, 1 \mid Z = z_0, W = w) - F(t_0, 1 \mid Z = z_0). \qquad (5)
\end{aligned}$$

It can be easily shown that this form is equivalent to the form given in the theorem. □

2.2 *Representation 2*

Theorem 1 in van der Laan and Gruber (2012) gives the form of the efficient influence curve for longitudinal parameters under multiple time point interventions as established in Bang and Robins (2005). By viewing censoring as a sequentially randomized treatment (equivalent to assuming CAR, see e.g. van der Laan and Rose (2011) Appendix 5) and the event process $N_2(\cdot)$ as a time-varying confounder, we may directly apply the results of this theorem.

We begin by showing that the conditional expectation representation of the cumulative incidence curve does in fact hold. To see this, note that for $m = 1, ..., t_0$, $\tilde{Q}_m^{z_0,0}(w)$ can be

written

$$\begin{aligned}\tilde{Q}_m^{z_0,0}(w) =& I(n_1(m-1) = n_2(m-1) = 0) \\ & \times \left[\sum_{t=m}^{t_0} \bar{Q}_1^{z_0,0}(t,w) \prod_{s=m}^{t-1} (1 - \bar{Q}_1^{z_0,0}(s,w) - \bar{Q}_2^{z_0,0}(s,w))\right] + I(n_1(m-1) = 1) \\ =& I(n_1(m-1) = n_2(m-1) = 0)R(m-1) + I(n_1(m-1) = 1),\end{aligned}$$

taking empty products to be 1 and empty sums to be 0 as necessary so that this form applies for all $m = 1, ..., t_0$. Letting $m = 1$ we see that $\tilde{Q}_1^{z_0,0}(w) = F(t_0, 1 \mid Z = z_0, W)$. Taking the expected value of this with respect to Q_W gives the unconditional cumulative incidence.

Now we show the form of the efficient influence curve for this representation of the parameter. A direct application of Theorem 1 in van der Laan and Gruber (2012) gives the efficient influence curve at P to be $\sum_{t=0}^{t_0} D_t^*(P)$, where

$$\begin{aligned} D_{t_0}^*(P)(o) &= H_g(t_0, o(t_0-1))[dN_1(t_0) - \bar{Q}_1^{z_0,0}(t_0, w)] \\ D_t^*(P)(o) &= H_g(t, o(t-1))[\tilde{Q}_{t+1}^{z_0,0}(w) - \tilde{Q}_t^{z_0,0}(w)], \quad \text{for } t = t_0 - 1, ..., 1 \\ D_0^*(P)(o) &= F(t_0, 1 \mid Z = z_0, W = w) - F(t_0, 1 \mid Z = z_0) \end{aligned}$$

First note that for $t = t_0$, we have $D_{t_0}^*(P)(o) = D_1(t_0, o) + D_2(t_0, o) = D_1(t_0, o)$ from our Theorem 1. Now for all $t = t_0 - 1, ..., 1$, we have the following equalities that will be used:

$$\begin{aligned} R(t) - R(t-1) &= -(1 - R(t))\bar{Q}_1^{z_0,0}(t, w) + \bar{Q}_2^{z_0,0}(t, w)R(t) \\ I(N_1(t) = 1) - I(N_1(t-1) = 1) &= dN_1(t); \\ I(N_1(t) = N_2(t) = 0) &= (1 - dN_1(t) - dN_2(t))I(N(t-1) = 0); \\ I(N(t-1) = 0)dN_1(t) &= dN_1(t). \end{aligned}$$

For $t = t_0 - 1, ..., 1$, we can then write

$$
\begin{aligned}
&(\tilde{Q}_{t+1}^{z_0,0}(w) - \tilde{Q}_t^{z_0,0}(w)) \\
&= (I(n(t) = 0)R(t, w) + I(n_1(t) = 1) - I(n_1(t-1) = 0)R(t-1, w) - I(n_1(t-1) = 1)) \\
&= (dn_1(t) + [1 - dn_1(t) - dn_2(t)]I(n(t-1) = 0)R(t) - I(n(t-1) = 0)R(t-1)) \\
&= I(n(t-1) = 0)(dn_1(t) - R(t)dn_1(t) - R(t)dN_2(t) + R(t) - R(t-1)) \\
&= I(n(t-1) = 0)([1 - R(t)]dn_1(t) - [1 - R(t)]\bar{Q}_1^{z_0,0}(t, w) - R(t)dn_2(t) + \bar{Q}_2^{z_0,0}(t, w)R(t)) \\
&= I(n(t-1) = 0)([1 - R(t)][dn_1 - \bar{Q}_1^{z_0,0}(t, w)] - R(t)[dn_2(t) - \bar{Q}_2^{z_0,0}(t, w)]),
\end{aligned}
$$

which is equivalent to $D_1(t, o) + D_2(t, o)$ from Theorem 1. Finally note that $D_0^*(P) = D_W(P)$ from Theorem 1, which completes the proof.

3. Web Appendix C: Regularity Conditions for Lemma 1

Lemma 1 will hold for estimators Q_n and g_n that satisfy the following conditions:

(1) $(P_n - P_0)[D^*(Q_n, g_n) - D^*(Q_0, g_0)] = o_p(n^{-1/2})$

(2) $U(Q_0, Q_n, g_0, g_n) = o_p(n^{-1/2})$,

where

$$U(Q_0, Q_n, g_0, g_n) = \Psi(Q_n) - \Psi(Q_0) + P_0 D^*(Q_n, g_n),$$

is the remainder term resulting from the linearization of $\Psi(Q)$ and involves second- and higher-order differences between Q_n and Q_0 and between g_n and g_0. Condition (1) would be satisfied if there exists a P_0-Donsker class $\mathcal{F}$ such that $D^*(Q_n, g_n) \in \mathcal{F}$ with probability tending to 1 and $P_0[D^*(Q_n, g_n) - D^*(Q_0, g_0)]^2 \to_p 0$. For a detailed proof see van der Laan and Rose (2011) Appendix 18, Theorem A5.

4. Web Appendix D: Constuction of TMLEs

4.1 *Hazard-based TMLE*

The hazard-based TMLE can be constructed as follows:

(1) Generate initial estimates, $g_{z_0,n}(w)$ and $G_{C,n}(\cdot, w)$, of the treatment and censoring mechanisms. These may be obtained through standard methods (e.g. parametric regression for $g_{z_0,n}(w)$; Kaplan-Meier for $G_{C,n}(\cdot, w)$), or using machine learning techniques, possibly combined data-adaptively using the Super Learner.

(2) Generate initial estimates, $\bar{Q}^{z_0,0}_{1,n,k}(\cdot, w)$ and $\check{Q}^{z_0,0}_{2,n,k}(\cdot, w)$, of the cause-specific hazards. Again, these could be obtained through standard logistic regression, but ideall would make use of more data adaptive methods. Let $k = 0$.

(3) Use the current estimates $\bar{Q}^{z_0,0}_{1,n,k}(\cdot, w)$ and $\check{Q}^{z_0,0}_{2,n,k}(\cdot, w)$ to compute $H_{1,n,k}(\cdot, o)$.

(4) Obtain $\epsilon_{1,n,k}$ as the estimated coefficient in a logistic regression model of outcome $dn_1(t)$ on covariate $H_{1,n,k}(t, o)$ with offset $\bar{Q}^{z_0,0}_{1,n,k}(t, w)$ in the subset of data with $Z = z_0, C(t - 1) = 0$. Set $\bar{Q}^{z_0,0}_{1,n,k+1}(t, w) := \bar{Q}^{z_0,0}_{1,n,k}(t, w)(\epsilon_{1,n,k})$.

(5) Use $\bar{Q}^{z_0,0}_{1,n,k+1}(\cdot, w)$ and $\check{Q}^{z_0,0}_{2,n,k}(\cdot, w)$ to compute $H_{2,n,k+1}(\cdot, o)$.

(6) Obtain $\epsilon_{2,n,k}$ as the estimated coefficient in a logisitic regression model of outcome $dn_2(t)$ on covariate $H_{2,n,k+1}(t, o)$ with offset $\text{logit}(\check{Q}^{z_0,0}_{2,n,k}(t, w)[1 - \bar{Q}^{z_0,0}_{1,n,k+1}(t, w)])$ in the subset of data with $Z = z_0, C(t - 1) = 0$. Set $\check{Q}^{z_0,0}_{2,n,k+1}(t, w) := \check{Q}^{z_0,0}_{2,n,k}(t, w)(\epsilon_{2,n,k})$.

(7) Iterate steps 3-6 until $\epsilon_{1,n,k} \approx 0$ and $\epsilon_{2,n,k} \approx 0$. Let $\bar{Q}^{z_0,0,*}_{1,n}(t, w)$, $\check{Q}^{z_0,0,*}_{2,n}(t, w)$ denote the estimates at the final iteration.

(8) Apply mapping in (**??**) using $\bar{Q}^{z,\bar{0}*}_{1,n}(\cdot, w)$, $\bar{Q}^{z_0,0,*}_{1,n}(t, w)$, $\check{Q}^{z_0,0,*}_{2,n}(t, w)$, and $Q_{W,n}(\cdot)$ to obtain the estimator $\psi^*_n := F^*_n(t_0, 1|Z = z_0)$.

4.2 *Iterative mean-based TMLE*

We begin by briefly discussing how to obtain initial estimators of $\tilde{Q}^{z_0,0}_{\cdot}(w)$ before detailing the TMLE algorithm. Consider the conditional mean at the last time point, $\tilde{Q}^{z_0,0}_{t_0}(w)$. The

initial estimate of this quantity should trivially assign 1 to subjects who have failed due to cause 1 prior to t_0 and 0 to subjects who have failed due to cause 2. It remains to estimate the conditional probability of a failure at time t_0 given that a subject has not already failed. This can be achieved through standard parametric regression or more data adaptive methods, but in each case we only use data for subjects at risk and uncensored at $t_0 - 1$. Moving to the next time point, $t_0 - 1$, the initial estimate of $\tilde{Q}^{z_0,0}_{t_0-1}(w)$ now assigns 1 to subjects who have failed due to cause 1 prior to $t_0 - 1$ and 0 to subjects who have failed due to cause 2. We then estimate the probability of a type 1 event at either $t_0 - 1$ or t_0 in the subset of subjects who have not failed due to any cause by time $t_0 - 2$. Again, this can be done using standard parametric regression or more data adaptive methods. An example of a standard regression approach would be a logistic regression model with $\tilde{Q}^{z_0,0}_{t_0-1,n}(w)$ as the outcome and functions of w as predictors, fit using the subset of the data for which $Z = z_0, C(t_0 - 2) = N(t_0 - 2) = 0$. We iterate this estimation process at each time and note that the subset of uncensored subjects used in the estimation at each step is getting larger. Eventually at $t = 1$, all subjects are used to obtain an estimate of $F(t_0, 1 \mid Z = z_0, W = w)$. We would then average that estimate over covariate levels to obtain an estimate of the unconditional cumulative incidence function.

The TMLE procedure follows the estimation procedure outlined above, but adds in a targeting step to estimation at each time point. Thus, the procedure is as follows:

(1) Generate initial estimates, $g_{z_0,n}(w)$ and $G_{C,n}(\cdot, w)$, of the treatment and censoring mechanisms. Use these to compute $H_{g,n}(\cdot, o)$.

(2) Generate an initial estimate $\tilde{Q}^{z_0,0}_{t_0,n}(w)$ of the first conditional mean as outlined above.

(3) Obtain $\epsilon_{t_0,n}$ by fitting logistic regression with $dn_1(t_0)$ as outcome, $\tilde{Q}^{z_0,0}_{t_0,n}(w)$ as offset, and $H_{g,n}(t_0, o)$ as covariate in the subset of data with $Z = z_0, C(t_0 - 1) = N_1(t_0 - 1) = N_2(t_0 - 1) = 0$. Set $\tilde{Q}^{z_0,0,*}_{t_0,n}(w) = \tilde{Q}^{z_0,0}_{t_0,n}(w)(\epsilon_{t_0,n})$. Let $t = t_0 - 1$.

(4) Generate an initial estimate $\tilde{Q}_{t,n}^{z_0,0}(w)$ of the t-th conditional mean as outlined above using $\tilde{Q}_{t+1,n}^{z_0,0}(w)$ as the outcome in e.g. a logistic regression model or a data adaptive learning algorithm.

(5) Obtain $\epsilon_{t,n}$ by fitting logistic regression with $\tilde{Q}_{t+1,n}^{z_0,0,*}(w)$ as outcome, $\tilde{Q}_{t,n}^{z_0,0}(w)$ as offset, and $H_{g,n}(t,o)$ as covariate in the subset of data with $Z = z_0, C(t-1) = N_1(t-1) = N_2(t-1) = 0$.

(6) Iterate steps 4-5 for $t = t_0 - 2, ..., 1$.

(7) Use the empirical distribution of covariates to average $\tilde{Q}_{1,n}^{z_0,0,*}(w) = F(t_0, 1 \mid Z = z_0, W = w)$ over w, obtaining estimator $\psi_n^* = F_n^*(t_0, 1 \mid Z = z_0)$.

4.3 *Practical implementation*

We first note that the targeting in the mean-based TMLE is completed in a single step, as opposed to the hazard-based approach, which requires iteration. However, the mean-based TMLE requires estimation of $\tilde{Q}_t^{z_0,0}$ at each time, whereas there were only two steps of initial estimation in the hazard-based approach. If the initial estimation is being done with a computationally intensive method, e.g. Super Learner, then the mean-based approach might be much slower to implement in practice. However, for a modest number of discrete time points, as in the simulation studies presented, we found that the methods take approximately the same amount of time to execute.

5. Web Appendix E: Simulation Details

5.1 *Simulation parameters*

Web Table 1 contains values for parameters used in the first simulation study. Additionally, $\beta_0 = -3.876$, $exp(\beta_1) = 0.474$, and $exp(\beta_3) = 1$ for all scenarios in simulation 1. The exponentiated parameters can be interpreted as a cause-specific hazard ratio for a unit difference in the associated covariate.

Table 2 contains values for parameters used in the second simulation study, with $\beta_0 = -3.876$, $exp(\beta_1) = 3.123$, $exp(\gamma_1) = 1.557$ for all scenarios.

[Table 1 about here.]

[Table 2 about here.]

5.2 *Super Learner libraries*

Web Table 3 shows the algorithms included in the Super Learner library for cause-specific hazards (TMLE-1), iterated conditional means (TMLE-2) and censoring (both TMLE-1 and TMLE-2) in the simulation studies. The generalized additive models were computed using the gam R package (Hastie, 2013) and the Super Learner algorithm was executed using the SuperLearner R package (Polley and van der Laan, 2013).

[Table 3 about here.]

5.3 *Standard error estimates*

Web Figure 1 shows the performance of the influence curve-based standard error estimates of $\hat{F}(8, 1 \mid Z = z_0)$ for $z_0 = 0, 1$ from simulation study 1. The estimates perform well in all settings.

[Figure 1 about here.]

6. Web Appendix F: Additional Simulation Studies

We examined the performance of our estimators when covariates are measured with error, e.g. self-reported risk behavior in an HIV vaccine efficacy trial. To generate data with measurement error, we suppose there exists a latent variable $L_1 \sim \mathrm{N}(0, \sigma^2_{L_1})$ and W_1 is a version of L_1 subject to measurement error,

$$W_1 = L_1 + e_1, \quad e_1 \sim \mathrm{N}(0, \sigma^2_{e_1}), \quad e_1 \perp L_1,$$

so that $W_1 \sim \mathrm{N}(0, \sigma_1^2)$, with $\sigma_1^2 = \sigma_{L_1}^2 + \sigma_{e_1}^2$.

We suppose there exists a second latent variable, $L_2^\star$, with

$$L_2^\star = L_2^\dagger + e_2, \quad L_2^\dagger \sim N(0, \sigma_{L_2}^2), \quad e_2 \sim \mathrm{N}(0, \sigma_{e_2}^2), \quad e_2 \perp L_2^\dagger,$$

so that $L_2^\star \sim \mathrm{N}(0, \sigma_2^2)$, with $\sigma_2^2 = \sigma_{L_2}^2 + \sigma_{e_2}^2$. W_2 is a dichotomized version of $L_2^\star$, i.e. $W_2 \equiv I(L_2^\star > \theta)$, which implies

$$W_2 \sim \mathrm{Bernoulli}(\Phi(-\theta/\sigma_{obs})),$$

where Φ is the standard normal CDF. We let $L_2 \equiv I(L_2^\dagger > \theta)$ be the true variable we would like to have measured. We denote $\rho_1 \equiv 1 - \sigma_{e_1}^2/\sigma_1^2$ as the fraction of variability of W_1 explained by L_1; similarly, $\rho_2 \equiv 1 - \sigma_{e_2}^2/\sigma_2^2$ is the fraction of variability in $L_2^\star$ explained by $L_2^\dagger$. We let $\rho_1 = 0.9$ and repeated simulation study 1. We then let $\rho_1 = \rho_2 = 0.9$ and $\theta = 0$ and repeated simulation study 2.

6.1 *Simulation 1 Results*

Web Figures 2, 3, and 4 show the results for the first simulation study when covariates are measured with measurement error. We see that when covariates are measured with error TMLE estimates have some bias and efficiency gains are less than when there is no error in covariate measurement. In general, when covariates are not prognostic TMLE still performs as well as Aalen-Johansen.

[Figure 2 about here.]

[Figure 3 about here.]

[Figure 4 about here.]

6.2 *Simulation 2 Results*

Web Figure 5 shows the results for simulation study 2 when covariates are measured with error. We see that TMLE still provides some gains in power over Aalen-Johansen for moder-

ately effective vaccines. When there was no measurement error, TMLE offered a larger gain in efficiency in the effect modification scenario. However, we see that the gain in power in the effect modification scenario is smaller than the gain in the no effect modification scenario.

[Figure 5 about here.]

7. Web Appendix G: HVTN 505 Results

Web Figure 6 shows results from an analysis of the HVTN 505 HIV vaccine efficacy trial using TMLE. We implemented the mean-based TMLE using a the Super Learner to generate our initial estimates at each time point as well as the censoring. We considered self-reported sexual risk behavior (number of sexual partners, unprotected sex, drug/alcohol use, and a derived sexual risk score. The Super Learner library included generalized linear models, Bayesian generalized linear models, generalized additive models, and stepwise selection procedures. The point estimates given by TMLE were very similar to those obtain via Aalen-Johansen, while the estimated variances were generally 1-3% smaller (Web Figure 7).

[Figure 6 about here.]

[Figure 7 about here.]

8. Web Appendix H: R Code

R code for implementing the methods in this article is provided online. The functions tmle.cmprsk1 and tmle.cmprsk2 execute the hazard-based and mean-based TMLE's respectively. For each method, the code calculates estimates of $F(t_0, 1 \mid Z = z_0), z_0 = 0, 1$ for a user-specified time point t_0. The code has been annotated for easy referencing against the description of the methods in this paper. The code for generating data and producing plots shown in the simulation studies is also included. Due to confidentiality agreements, the real RV144 data set could not be made available. We have instead included an illustration of our

method on a similarly-structured data set and have included code to reproduce the RV144 plots included in the body of this paper.

References

Bang, H. and Robins, J. M. (2005). Doubly robust estimation in missing data and causal inference models. *Biometrics* **61,** 962–973.

Hastie, T. (2013). *gam: Generalized Additive Models*. R package version 1.09.1.

Petersen, M. L., Porter, K. E., Gruber, S., Wang, Y., and van der Laan, M. J. (2010). Diagnosing and responding to violations in the positivity assumption. *Statistical Methods in Medical Research* **21,** 31–54.

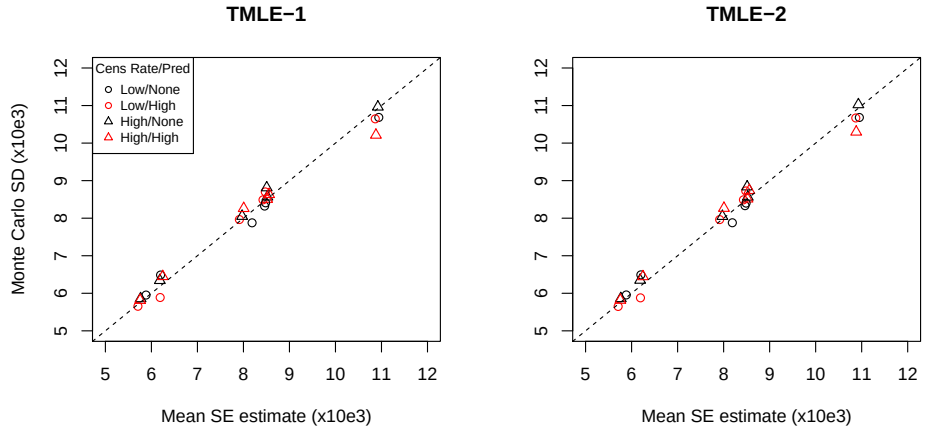
Polley, E. and van der Laan, M. (2013). *SuperLearner: Super Learner Prediction*. R package version 2.0-10.

Tsiatis, A. (2007). *Semiparametric theory and missing data*. Springer.

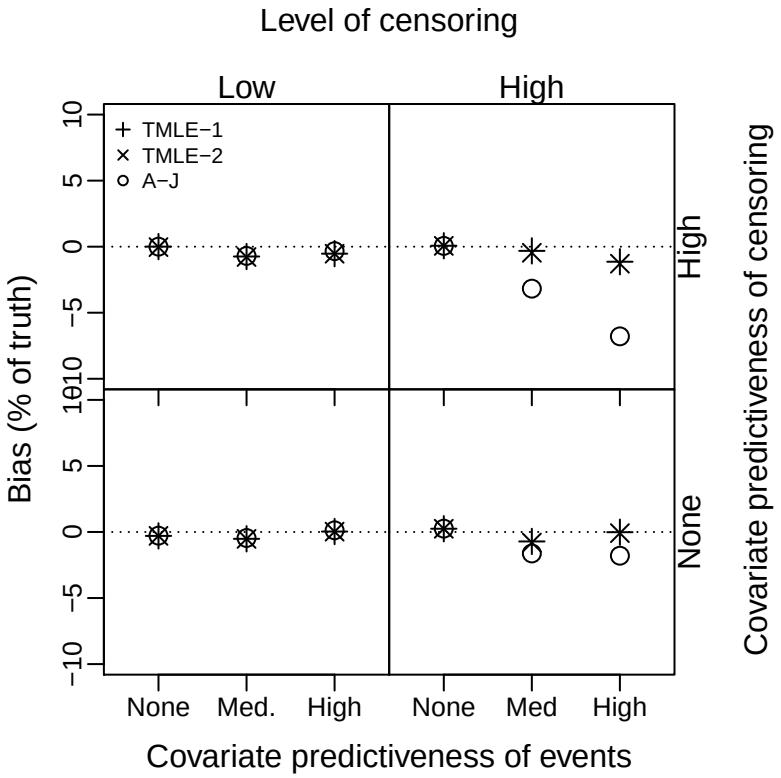
van der Laan, M. J. and Gruber, S. (2012). Targeted minimum loss based estimation of causal effects of multiple time point interventions. *The International Journal of Biostatistics* **8,** 1–34.

van der Laan, M. J. and Robins, J. M. (2003). *Unified Methods for Censored Longitudinal Data and Causality*. Springer.

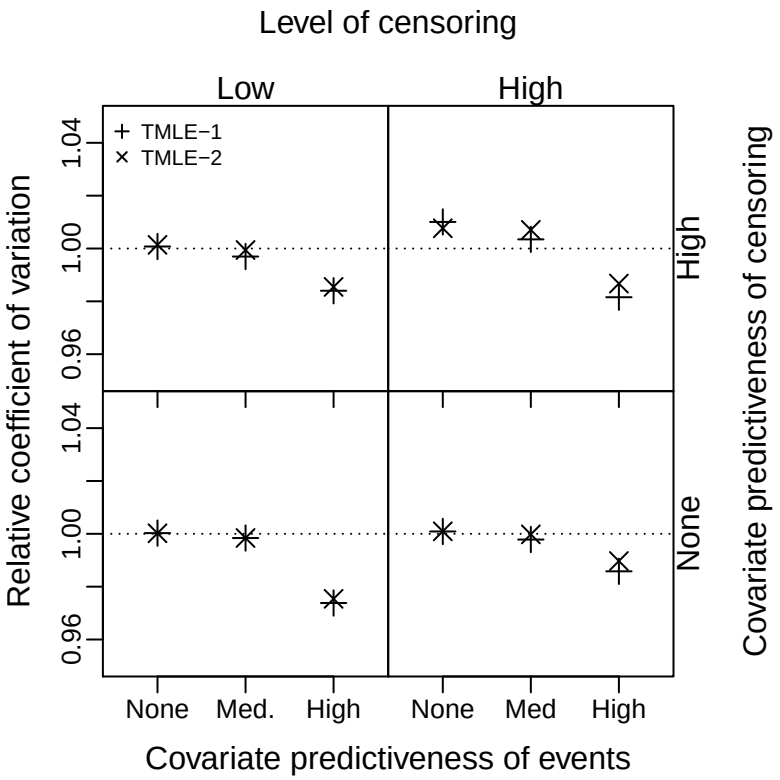
van der Laan, M. J. and Rose, S. (2011). *Targeted Learning: Causal Inference for Observational and Experimental Data*. Springer.



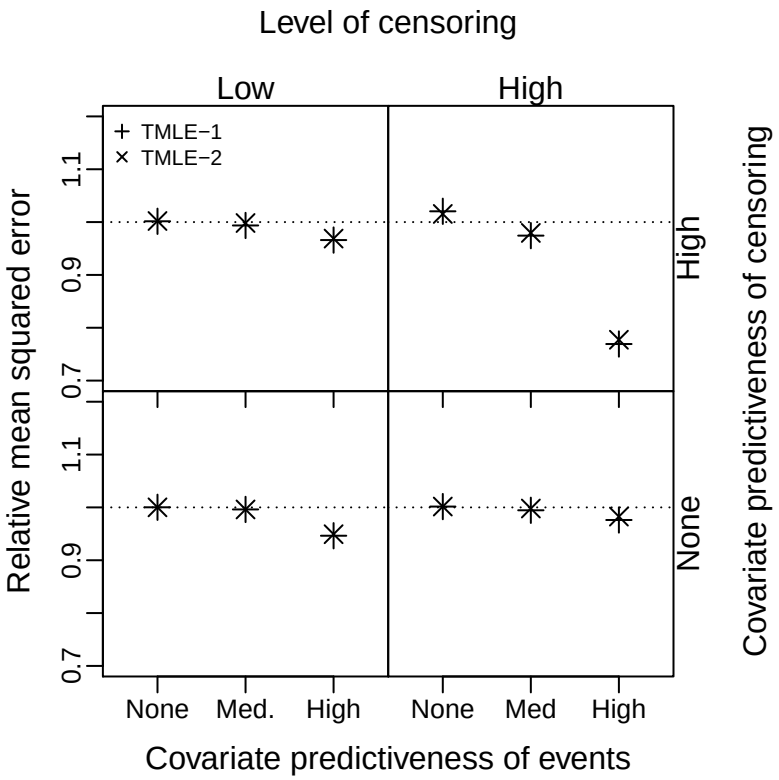
Web Figure 1: Monte Carlo standard deviation (SD) plotted against the mean standard error (SE) estimate for simulation study 1 for the hazard-based TMLE (TMLE-1) and the mean-based TMLE (TMLE-2). The dashed line is the identity line, along which points should fall if estimates of standard error accurately reflect the true standard deviation of the estimates. Circles indicate the low censoring rate setting and triangles represent high censoring rate. Red points indicate high covariate predictiveness of censoring, while black points indicate low covariate predictiveness of censoring.



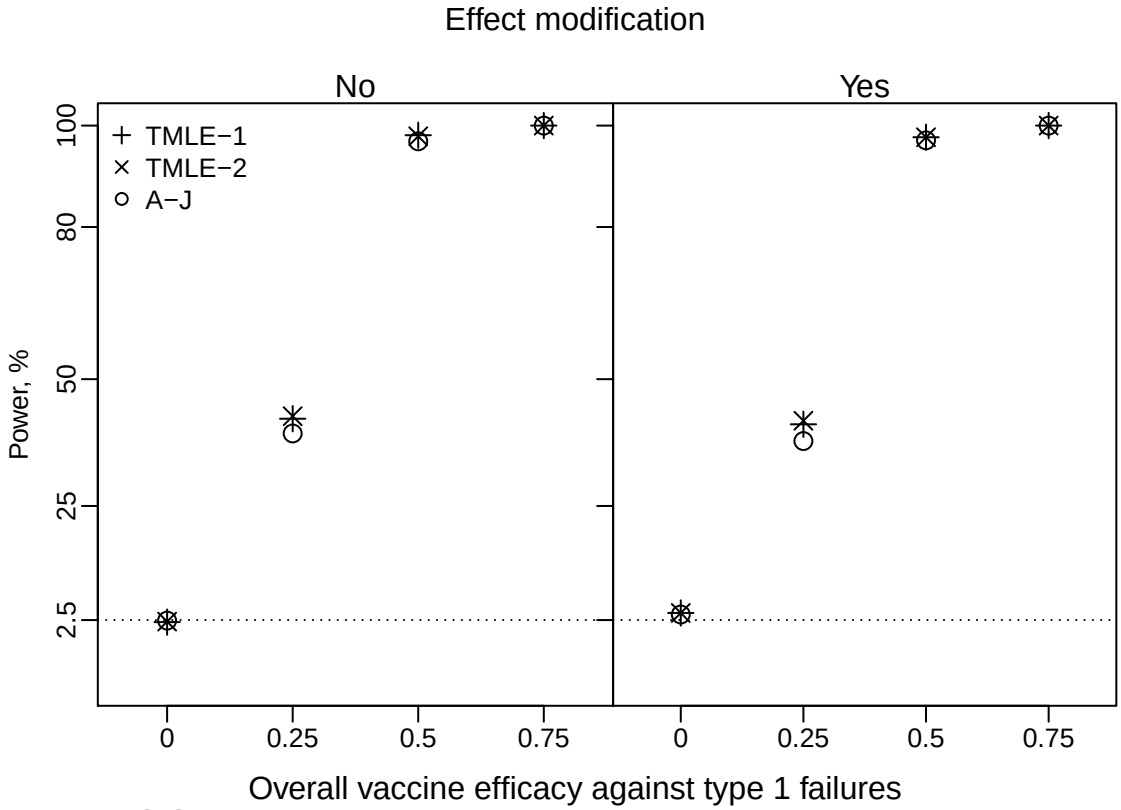
Web Figure 2: Bias of the hazard-based TMLE (TMLE-1), mean-based TMLE (TMLE-2), and the Aalen-Johansen (A-J) estimator when covariates are measured with error. The bias is presented as a percentage of the true cumulative incidence function.



Web Figure 3: Relative coefficient of variation of the hazard-based (TMLE-1) and mean-based (TMLE-2) TMLE estimators compared to the Aalen-Johansen (A-J) estimator when covariates are measured with error. Values below 1.0 indicate a smaller coefficient of variation for the TMLE estimator.



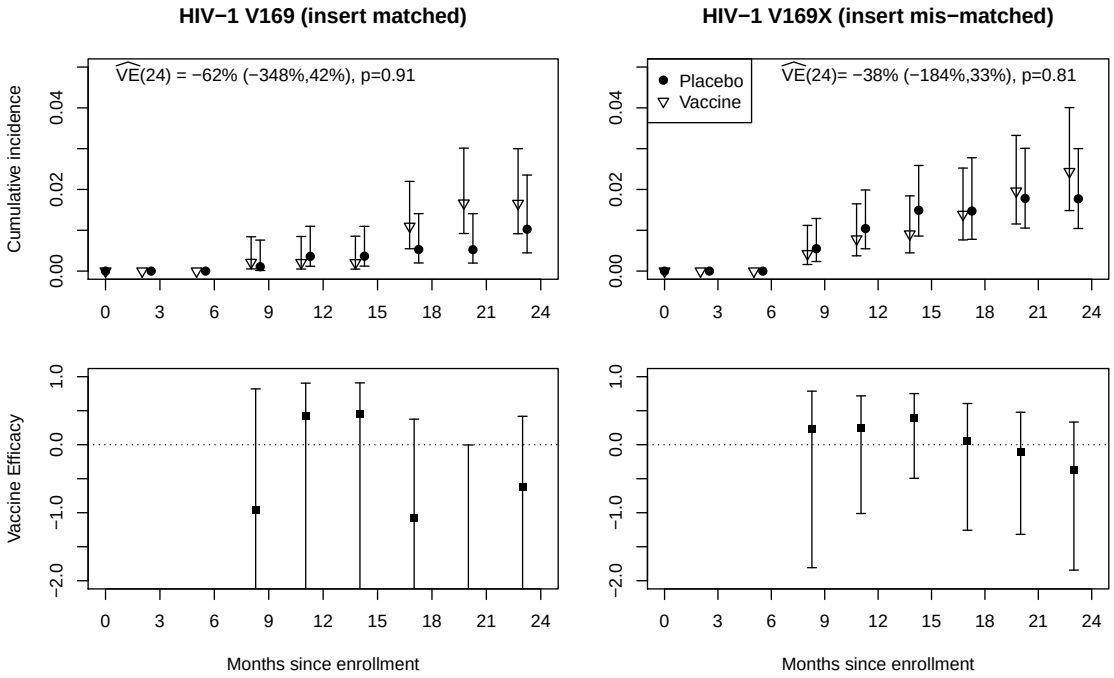
Web Figure 4: Relative mean squared error of the hazard-based (TMLE-1) and mean-based (TMLE-2) estimators compared to the Aalen-Johansen (A-J) estimator when covariates are measured with error. Values below 1.0 indicate a smaller mean squared error for the TMLE estimator.



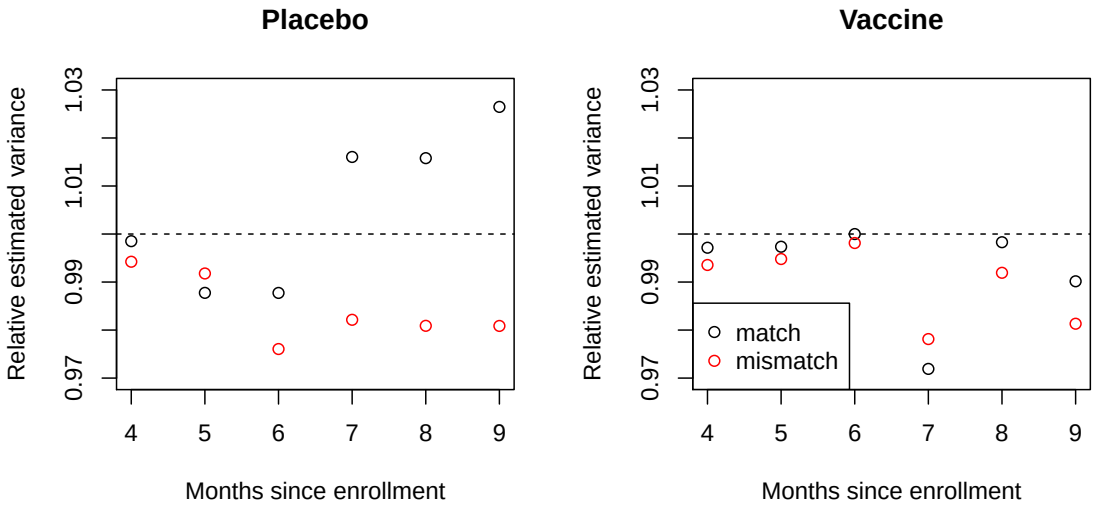
Power relative to A−J

TMLE−1	1.07	1.01	1.00	1.09	1.01	1.00
TMLE−2	1.09	1.01	1.00	1.11	1.01	1.00

Web Figure 5: Power of a one-sided level $\alpha = 0.025$ (dotted line) Wald test of the null hypothesis of no vaccine efficacy against type 1 events at the end of follow up when covariates are measured with error. Tests are based on hazard-based TMLE (TMLE-1), mean-based TMLE (TMLE-2), and the Aalen-Johansen (A-J) estimators. The left panel shows power with no effect modification; the right panel shows power with effect modification.



Web Figure 6: Results from the HVTN 505 HIV vaccine efficacy trial. The top row shows estimated cumulative incidence over time using the hazard-based TMLE for 169-matched and mismatched HIV-infections with point-wise 95% confidence intervals. The bottom row shows estimated vaccine efficacy (VE) and 95% confidence interval (CI) for each type of infection.



Web

Figure 7: Comparison of the estimated variance of Aalen-Johansen estimator and TMLE estimator across time points in HVTN-505. Points below 1 indicate a smaller estimated variance for TMLE estimators.

Pred. cens.	Cens. rate	Pred. event	$exp(\beta_2)$	γ_0	$exp(\gamma_1)$	n
None	Low	None	1	-4.125	1	2180
None	Low	Medium	1.557	-4.115	1	2150
None	Low	High	3.123	-4.100	1	1570
None	High	None	1	-2.887	1	2630
None	High	Medium	1.557	-2.875	1	2450
None	High	High	3.123	-2.840	1	1780
High	Low	None	1	-4.210	1.557	2340
High	Low	Medium	1.557	-4.200	1.557	2180
High	Low	High	3.123	-4.130	1.557	1610
High	High	None	1	-2.935	1.557	2650
High	High	Medium	1.557	-2.920	1.557	2530
High	High	High	3.123	-2.850	1.557	1900

Web Table 1: Parameters for simulation study 1. The first three columns indicate the scenario considered, consisting of all combinations of none/high covariate predictiveness of censoring (Pred. cens.), low/high censoring rates (Cens. rate), none/medium/high covariate predictiveness of events (Pred. event).

Eff. Mod.	VE	$exp(\beta_1)$	$exp(\beta_3)$	γ_0	n
No	0	1	1	-4.170	1640
No	0.25	0.726	1	-4.175	1880
No	0.5	0.474	1	-4.190	2190
No	0.75	0.225	1	-4.206	2630
Yes	0	1.620	0.280	-4.120	1200
Yes	0.25	1	0.474	-4.170	1880
Yes	0.5	0.830	0.220	-4.200	2150
Yes	0.75	0.375	0.250	-4.215	2630

Web Table 2: Parameters for simulation study 2. The first two columns indicate the scenario considered, consisting of combinations of effect modification yes/no (Eff. mod.), and vaccine efficacy 0,0.25,0.5,0.75 (VE).

Model type	Time	Covariates
Hazard estimates (TMLE-1)		
`glm`	$\emptyset$	Z
`glm`	$\emptyset$	$Z + W_1 + W_2$
`glm`	$\emptyset$	$Z + W_1 + W_2 + Z * W_2$
`glm`	`factor(t)`	$Z + W_1 + W_2$
`glm`	`factor(t)`	$Z + W_1 + W_2 + Z * W_2$
`gam`	$\emptyset$	$Z + \mathtt{s}(W_1, df = 3) + W_2$
`gam`	$\mathtt{s}(t, df = 3)$	$Z + W_1 + W_2$
`gam`	$\emptyset$	$Z + \mathtt{s}(W_1, df = 3) + W_2 + Z * W_2$
`gam`	$\mathtt{s}(t, df = 3)$	$Z + W_1 + W_2 + Z * W_2$
Conditional mean estimates (TMLE-2)		
`glm`	$\emptyset$	Z
`glm`	$\emptyset$	$Z + W_1 + W_2$
`glm`	$\emptyset$	$Z + W_1 + W_2 + Z * W_2$
`gam`	$\emptyset$	$Z + \mathtt{s}(W_1, df = 3) + W_2$
`gam`	$\emptyset$	$Z + \mathtt{s}(W_1, df = 3) + W_2 + Z * W_2$
Censoring estimates (TMLE-1 & TMLE-2)		
`glm`	$\emptyset$	$\emptyset$
`glm`	$\emptyset$	$Z + W_1 + W_2$
`glm`	$\emptyset$	$Z + W_1 + W_2 + Z * W_2$
`glm`	`factor(t)`	$Z + W_1 + W_2$
`glm`	`factor(t)`	$Z + W_1 + W_2 + Z * W_2$
`gam`	$\emptyset$	$Z + \mathtt{s}(W_1, df = 3) + W_2$
`gam`	$\mathtt{s}(t, df = 3)$	$Z + W_1 + W_2$
`gam`	$\emptyset$	$Z + \mathtt{s}(W_1, df = 3) + W_2 + Z * W_2$
`gam`	$\mathtt{s}(t, df = 3)$	$Z + W_1 + W_2 + Z * W_2$

Web Table 3: Models included in Super Learner libraries for simulation studies. The columns indicate what type of model was used (`glm`=generalized linear model, `gam` = generalized additive model), how time was modeled ($\emptyset$ denotes time was omitted from the model, `factor(t)` indicates dummy variables were used), and what covariates were included ($x * y$ indicates a cross product between covariates x and y). We use $\mathtt{s}(x, df = d)$ to denote that variable x was modeled using a polynomial spline of degree d.