Supporting Information for "Parametric g-formula implementations for causal survival analyses" by L. Wen, J. G. Young, J. M. Robins and M. A. Hernán

A. The IP Weighted estimator

There are at least three algebraically equivalent representations of the g-formula: (1) an expectation weighted by the joint densities of covariates, (2) an iterative conditional expectation over time, and (3) an inverse probability weighted expectation. All three representations are nonparametrically identical, and in this section we will focus on the inverse-probability weighted (IP weighted) estimator under representation (3).

The IP weighted estimator has been used, for example, to estimate the effects of lifestyle on coronary heart disease (Young *and others*, 2019); the effects of aggressive glucose-lowering strategies on the progression of albuminuria (Neugebauer *and others*, 2016); the effects of antiretroviral therapy on mortality (Cain *and others*, 2010), the effect of joint monitoring and treatment strategies on mortality (Caniglia *and others*, 2019) and the effect of treatment switching strategies (Cain *and others*, 2016) in HIV positive individuals.

The IP weighted estimator upweights the outcomes of those who followed treatment strategy g to account for those who did not follow the strategy. For simplicity, we will describe IP weighted estimators for deterministic treatment strategies, but generalization to any user-specified intervention will be discussed at the end. In what follows, we will discuss two IP weighted estimators for the cumulative counterfactual risk under intervention g or $\mathrm{E}(Y_J^g)$.

One estimator is an Horvitz-Thompson IP weighted estimator $\hat{\mu}_{IPW,HT}$, which can be obtained from the following:

$$\hat{\mu}_{IPW,HT}=\sum_{j=1}^{J}\mathbb{P}_n\left\{\frac{I(\bar{A}_{j-1}=\bar{A}_{j-1}^g)(1-C_j)(1-Y_{j-1})}{\prod_{k=0}^{j-1}f(A_k|\bar{A}_{k-1},\bar{L}_k,Y_k=0)P(C_{k+1}=0|C_k=0,\bar{A}_k,\bar{L}_k,Y_k=0)}Y_j\right\} \tag{A.1}$$

where $\mathbb{P}_n(X)=n^{-1}\sum_{i=1}^{n}(X_i)$ for any random variable X. Since each jth component ($j=0,\ldots,J-1$) in the last equation is *not* a convex combination of the observed outcome values, the point estimates obtained from the Horvitz-Thompson IP weighted estimator may be unbounded (i.e. it is not guaranteed to fall between 0 and 1) (see Robins, 2007). Note that alternatively, we could estimate $\mathrm{E}(Y_J^g)$ by calculating

$$1-\mathbb{P}_n\left\{\frac{I(\bar{A}_{J-1}=\bar{A}_{J-1}^g)(1-C_J)}{\prod_{k=0}^{J-1}f(A_k|\bar{A}_{k-1},\bar{L}_k,Y_k=0)P(C_{k+1}=0|C_k=0,\bar{A}_k,\bar{L}_k,Y_k=0)}(1-Y_J)\right\}.$$

Note the estimator $\mathbb{P}_n\left\{\frac{I(\bar{A}_{J-1}=\bar{A}^g_{J-1})(1-C_J)}{\prod_{k=0}^{J-1}f(A_k|\bar{A}_{k-1},\bar{L}_k,Y_k=0)P(C_{k+1}=0|C_k=0,\bar{A}_k,\bar{L}_k,Y_k=0)}Y_J\right\}$ is not consistent for $\mathrm{E}(Y^g_J)$ since A_{J-1} is only defined when $Y_{J-1}=0$.

The following estimator requires slightly more modeling effort. This IP weighted estimator $\hat{\mu}_{IPW,Haz}$ can be obtained as a function of J discrete-time hazard of death or

$$\hat{\mu}_{IPW,Haz}=\sum_{j=1}^{J}\left\{\Lambda_{IPW,j}\prod_{k=1}^{j-1}(1-\Lambda_{IPW,k})\right\}.$$

Here, $\hat{\Lambda}_{IPW,j}$ is the estimated conditional probability of death at time j given survival at time $j-1$ under treatment strategy g and can be obtained by solving for $\Lambda_{IPW,j}$ in the following estimating equations:

$$\mathbb{P}_n\left\{(1-Y_{j-1})\frac{I(\bar{A}_{j-1}=\bar{A}^g_{j-1})(1-C_j)}{\prod_{k=0}^{j-1}f(A_k|\bar{A}_{k-1},\bar{L}_k,Y_k=0)P(C_{k+1}=0|C_k=0,\bar{A}_k,\bar{L}_k,Y_k=0)}(Y_j-\Lambda_{IPW,j})\right\}=0 \tag{A.2}$$

$\hat{\Lambda}_{IPW,j}$ will always be bounded between 0 and 1 since it is always a convex combination of the observed Y_{j+1}-values and thus $\mathrm{E}(Y^g_J)$ will always be bounded between 0 and 1.

Nonparametric estimation of the probability of treatment and the probability of censoring may not be feasible when L_j is high-dimensional, but we can impose working models to estimate both. Estimated values for the probability of treatment and probability of censoring can be obtained by fitting logistic regression models from the observed data. For general user-specified intervention distribution, we would replace $I(\bar{A}_{j-1}=\bar{A}^g_{j-1})$ by $\prod_{k=0}^{j-1}f^{int}(A_j|\bar{A}_{j-1},\bar{L}_j,Y_j=0)$ in Equations (A.1) and (A.2). In all of the simulation studies (see Web Appendix E), we provide results from $\hat{\mu}_{IPW,HT}$ and $\hat{\mu}_{IPW,Haz}$.

B. Proof that Expression (1) is algebraically equivalent to Expression (2)

For simplicity, we prove for two time points that Expressions (1) and (2) in the main text are equivalent. This result can be generalized for any J. We obtain from Expression (1):

$$\sum_{\forall\bar{a}_{j-1}}\sum_{\forall\bar{l}_{j-1}}\sum_{k=1}^{2}\mathrm{P}(Y_k=1|Y_{k-1}=C_k=0,\bar{L}_{k-1}=\bar{l}_{k-1},\bar{A}_{k-1}=\bar{a}_{k-1})\times$$
$$\prod_{s=0}^{k-1}\mathrm{P}(Y_s=0|Y_{s-1}=C_s=0,\bar{L}_{s-1}=\bar{l}_{s-1},\bar{A}_{s-1}=\bar{a}_{s-1})f(l_s|Y_s=C_s=0,\bar{l}_{s-1},\bar{a}_{s-1})f^{\mathrm{int}}(a_s|Y_s=C_s=0,\bar{l}_s,\bar{a}_{s-1})$$
$$=\sum_{\forall l_0}\sum_{\forall a_0}\left[\left\{\sum_{\forall l_1}\sum_{\forall a_1}P(Y_2=1|Y_1=C_2=0,\bar{L}_1=\bar{l}_1,\bar{A}_1=\bar{a}_1)f^{\mathrm{int}}(a_1|Y_1=C_1=0,\bar{l}_1,a_0)f(l_1|Y_1=C_1=0,l_0,a_0)\right.\right.$$
$$\left.\left.P(Y_1=0|C_1=0,L_0=l_0,A_0=a_0)f^{\mathrm{int}}(a_0|l_0)f(l_0)\right\}+P(Y_1=1|C_1=0,L_0=l_0,A_0=a_0)f^{\mathrm{int}}(a_0|l_0)f(l_0)\right]$$

We obtain from Expression (2):

$$\mathrm{E}_{f_{L_0}}\left(\mathrm{E}_{f^{\mathrm{int}}_{A_0}}\left[\mathrm{E}_{f_{Y_1}}\left\{Y_1+\mathrm{E}_{f_{L_1}}\left(\mathrm{E}_{f^{\mathrm{int}}_{A_1}}\left[\mathrm{E}_{f_{Y_2}}\{Y_2(1-Y_1)|Y_1,C_2=0,\bar{L}_1,\bar{A}_1\}|Y_1,C_1=0,\bar{L}_1,A_0\right]|Y_1,C_1=0,L_0,A_0\right)\middle|C_1=0,L_0,A_0\right\}\middle|L_0\right]\right)$$

Taking the inner-most expectation w.r.t. Y_2 given $(Y_1,\{C_2=0\},\bar{L}_1,\bar{A}_1)$, then w.r.t. A_1 (**under** $f^{\mathbf{int}}$) given $(Y_1,\{C_1=0\},\bar{L}_1,A_0)$, and then w.r.t. L_1 given $(Y_1,\{C_1=0\},L_0,A_0)$ yields:

$$\mathrm{E}_{f_{L_0}}\left(\mathrm{E}_{f^{\mathrm{int}}_{A_0}}\left[\mathrm{E}_{f_{Y_1}}\left\{Y_1+\mathrm{E}_{f_{L_1}}\left(\mathrm{E}_{f^{\mathrm{int}}_{A_1}}\left[\mathrm{E}_{f_{Y_2}}\{Y_2(1-Y_1)|Y_1,C_2=0,\bar{L}_1,\bar{A}_1\}|Y_1,C_1=0,\bar{L}_1,A_0\right]|Y_1,C_1=0,L_0,A_0\right)\middle|C_1=0,L_0,A_0\right\}\middle|L_0\right]\right)$$

$$=\mathrm{E}_{f_{L_0}}\left(\mathrm{E}_{f^{\mathrm{int}}_{A_0}}\left[\mathrm{E}_{f_{Y_1}}\left\{Y_1+(1-Y_1)\sum_{\forall l_1}\sum_{\forall a_1}\mathrm{E}(Y_2|Y_1=C_2=0,L_0,A_0,l_1,a_1)f^{\mathrm{int}}(a_1|Y_1=C_1=0,L_0,A_0,l_1)f(l_1|Y_1=C_1=0,L_0,A_0)\right.\right.\right.$$
$$\left.\left.\left.\middle|C_1=0,L_0,A_0\right\}\middle|L_0\right]\right)$$

Then taking expectation w.r.t. Y_1 given $(\{C_1=0\},L_0,A_0)$, then w.r.t. A_0 (**under** $f^{\mathbf{int}}$) given (L_0), and finally w.r.t. L_0 yields:

$$\mathrm{E}_{f_{L_0}}\left(\mathrm{E}_{f^{\mathrm{int}}_{A_0}}\left[\mathrm{E}_{f_{Y_1}}\left\{Y_1|C_1=0,L_0,A_0\right\}\middle|L_0\right]\right)+\mathrm{E}_{f_{L_0}}\left(\mathrm{E}_{f^{\mathrm{int}}_{A_0}}\left[\mathrm{E}_{f_{Y_1}}\left\{(1-Y_1)\sum_{\forall l_1}\sum_{\forall a_1}\mathrm{E}(Y_2|Y_1=C_2=0,L_0,A_0,l_1,a_1)\right.\right.\right.$$
$$\left.\left.\left.f^{\mathrm{int}}(a_1|Y_1=C_1=0,L_0,A_0,l_1)f(l_1|Y_1=C_1=0,L_0,A_0)\middle|C_1=0,L_0,A_0\right\}\middle|L_0\right]\right)$$
$$=\sum_{\forall l_0}\sum_{\forall a_0}P(Y_1=1|C_1=0,L_0=l_0,A_0=a_0)f^{\mathrm{int}}(a_0|l_0)f(l_0)+\sum_{\forall \bar{l}_1}\sum_{\forall \bar{a}_1}P(Y_2=1|Y_1=C_2=0,\bar{L}_1=\bar{l}_1,\bar{A}_1=\bar{a}_1)$$
$$f^{\mathrm{int}}(a_1|Y_1=C_1=0,\bar{l}_1,a_0)f(l_1|Y_1=C_1=0,l_0,a_0)P(Y_1=0|C_1=0,L_0=l_0,A_0=a_0)f^{\mathrm{int}}(a_0|l_0)f(l_0)$$
$$=\sum_{\forall l_0}\sum_{\forall a_0}\left[\left\{\sum_{\forall l_1}\sum_{\forall a_1}P(Y_2=1|Y_1=C_2=0,\bar{L}_1=\bar{l}_1,\bar{A}_1=\bar{a}_1)f^{\mathrm{int}}(a_1|Y_1=C_1=0,\bar{l}_1,a_0)f(l_1|Y_1=C_1=0,l_0,a_0)\right.\right.$$
$$\left.\left.P(Y_1=0|C_1=0,L_0=l_0,A_0=a_0)f^{\mathrm{int}}(a_0|l_0)f(l_0)\right\}+P(Y_1=1|C_1=0,L_0=l_0,A_0=a_0)f^{\mathrm{int}}(a_0|l_0)f(l_0)\right]$$

which is what we obtain from Expression (1). Similarly, to show that Expressions (1.d) and (2.d) are algebraically equivalent, we note that for two time points Expression (1.d) equals:

$$\sum_{\forall \bar{l}_1}\sum_{k=1}^{2}\mathrm{P}(Y_k=1|Y_{k-1}=C_k=0,\bar{L}_{k-1}=\bar{l}_{k-1},\bar{A}_{k-1}=\bar{a}^g_{k-1})\times$$
$$\prod_{s=0}^{k-1}\mathrm{P}(Y_s=0|Y_{s-1}=C_s=0,\bar{L}_{s-1}=\bar{l}_{s-1},\bar{A}_{s-1}=\bar{a}^g_{s-1})f(l_s|Y_s=C_s=0,\bar{l}_{s-1},\bar{a}^g_{s-1})$$
$$=\sum_{\forall l_0}\left[\left\{\sum_{\forall l_1}P(Y_2=1|Y_1=C_2=0,\bar{L}_1=\bar{l}_1,\bar{A}_1=\bar{a}^g_1)f(l_1|Y_1=C_1=0,l_0,a^g_0)P(Y_1=0|C_1=0,L_0=l_0,A_0=a^g_0)f(l_0)\right\}\right.$$
$$\left.+P(Y_1=1|C_1=0,L_0=l_0,A_0=a^g_0)f(l_0)\right]$$

We obtain from Expression (2.d):

$$\mathrm{E}_{f_{L_0}}\left[\mathrm{E}_{f_{Y_1}}\left\{Y_1+\mathrm{E}_{f_{L_1}}\left(\mathrm{E}_{f_{Y_2}}\{Y_2(1-Y_1)|Y_1,C_2=0,\bar{L}_1,\bar{A}_1=\bar{A}_1^g\}|Y_1,C_1=0,L_0,A_0=A_0^g\right)\Big|C_1=0,L_0,A_0=A_0^g\right\}\right]$$

Taking the inner-most expectation w.r.t. Y_2 given $(Y_1,\{C_2=0\},\bar{L}_1,\{\bar{A}_1=\bar{A}_1^g\})$, and then w.r.t. L_1 given $(Y_1,\{C_1=0\},L_0,\{A_0=A_0^g\})$ yields:

$$\mathrm{E}_{f_{L_0}}\left[\mathrm{E}_{f_{Y_1}}\left\{Y_1+\mathrm{E}_{f_{L_1}}\left(\mathrm{E}_{f_{Y_2}}\{Y_2(1-Y_1)|Y_1,C_2=0,\bar{L}_1,\bar{A}_1=\bar{A}_1^g\}|Y_1,C_1=0,L_0,A_0=A_0^g\right)\Big|C_1=0,L_0,A_0=A_0^g\right\}\right]$$

$$=\mathrm{E}_{f_{L_0}}\left[\mathrm{E}_{f_{Y_1}}\left\{Y_1+(1-Y_1)\sum_{\forall l_1}\mathrm{E}(Y_2|Y_1=C_2=0,L_0,l_1,\bar{A}_1=\bar{A}_1^g)f(l_1|Y_1=0,L_0,A_0^g)\Big|C_1=0,L_0,A_0=A_0^g\right\}\right]$$

Taking expectation w.r.t. Y_1 given $(\{C_1=0\},L_0,A_0)$, and finally w.r.t. L_0 yields:

$$\mathrm{E}_{f_{L_0}}\left[\mathrm{E}_{f_{Y_1}}\left\{Y_1|C_1=0,L_0,A_0=A_0^g\right\}\right]+\mathrm{E}_{f_{L_0}}\left[\mathrm{E}_{f_{Y_1}}\left\{(1-Y_1)\sum_{\forall l_1}\mathrm{E}(Y_2|Y_1=C_2=0,L_0,l_1,\bar{A}_1=\bar{A}_1^g)f(l_1|Y_1=0,L_0,A_0^g)\Big|C_1=0,L_0,A_0=A_0^g\right\}\right]$$

$$=\sum_{\forall l_0}P(Y_1=1|L_0=l_0,A_0=a_0^g)f(l_0)+\sum_{\forall \bar{l}_1}P(Y_2=1|Y_1=0,\bar{L}_1=\bar{l}_1,\bar{A}_1=\bar{a}_1^g)f(l_1|Y_1=0,l_0,a_0^g)P(Y_1=0|L_0=l_0,A_0=a_0^g)f(l_0)$$

$$=\sum_{\forall l_0}\Bigg[\left\{\sum_{\forall l_1}P(Y_2=1|Y_1=C_2=0,\bar{L}_1=\bar{l}_1,\bar{A}_1=\bar{a}_1^g)f(l_1|Y_1=C_1=0,l_0,a_0^g)P(Y_1=0|C_1=0,L_0=l_0,A_0=a_0^g)f(l_0)\right\}$$
$$+P(Y_1=1|C_1=0,L_0=l_0,A_0=a_0^g)f(l_0)\Bigg]$$

which is what we obtain from Expression (1.d).

C. Details on the hazard-extended ICE estimator

C.1 Deterministic treatment strategies

In this section, we show that Expressions (2) and (2.d) equal $\mathrm{E}(Y_J^g)$. Since we have shown in Web Appendix B that Expressions (2) and (2.d) equal Expressions (1) and (1.d), respectively, Expressions (1) and (1.d) also equal $\mathrm{E}(Y_J^g)$. Again, for simplicity, we show this for two time points, but the result can be generalized for any J.

We obtain from Expression (2.d) and from the observed data:

$$\mathrm{E}_{f_{L_0}}\left[\mathrm{E}_{f_{Y_1}}\left\{Y_1+\mathrm{E}_{f_{L_1}}\left(\mathrm{E}_{f_{Y_2}}\{Y_2(1-Y_1)|Y_1,C_2=0,\bar{L}_1,\bar{A}_1=\bar{A}_1^g\}|Y_1,C_1=0,L_0,A_0=A_0^g\right)\Big|C_1=0,L_0,A_0=A_0^g\right\}\right]$$

$$=\mathrm{E}_{f_{L_0}}\left\{\mathrm{E}_{f_{Y_1}}\left[Y_1+(1-Y_1)\mathrm{E}_{f_{L_1}}\left\{\underbrace{P(Y_2=1|Y_1=C_2=0,\bar{L}_1,\bar{A}_1=\bar{A}_1^g)}_{h_{2,1}^g}|Y_1=C_1=0,L_0,A_0=A_0^g\right\}\Big|C_1=0,L_0,A_0=A_0^g\right]\right\}$$

$$=\mathrm{E}_{f_{L_0}}\left[\underbrace{P(Y_1=1|C_1=0,L_0,A_0=A_0^g)}_{h_{1,0}^g}+P(Y_1=0|C_1=0,L_0,A_0=A_0^g)\mathrm{E}_{f_{L_1}}\left(h_{2,1}^g|Y_1=C_1=0,L_0,A_0=A_0^g\right)\right]$$

$$=\mathrm{E}_{f_{L_0}}\left[\underbrace{h_{1,0}^g+(1-h_{1,0}^g)\mathrm{E}_{f_{L_1}}\left(h_{2,1}^g|Y_1=C_1=0,L_0,A_0=A_0^g\right)}_{h_{2,0}^g}\right]=\mathrm{E}_{f_{L_0}}(h_{2,0}^g)$$

The first line in the equations can be shown to equal $\mathrm{E}(Y_2^g)$ under consistency, positivity and exchangeability. This is because under these assumptions, $h_{2,1}^g=\mathrm{E}(Y_2^g|Y_1=C_2=0,\bar{L}_1,\bar{A}_1=\bar{A}_1^g)=\mathrm{E}(Y_2^g|Y_1=C_1=0,\bar{L}_1,A_0=A_0^g)$, and $h_{2,1}^g+(1-h_{2,1}^g)\mathrm{E}_{f_{L_1}}\left(h_{2,1}^g|Y_1=C_1=0,L_0,A_0=A_0^g\right)=\mathrm{P}(Y_1^g=1|C_1=0,L_0,A_0=A_0^g)+\mathrm{P}(Y_2^g=1,Y_1^g=0|C_1=0,L_0,A_0=A_0^g)=\mathrm{P}(Y_2^g=1|C_1=0,L_0,A_0=A_0^g)=\mathrm{P}(Y_2^g=1|L_0)$ for deterministic treatment strategies.

Since ($h_{2,1}^g$, $h_{2,1}^g$, $h_{2,0}^g$) are generally unknown, they need to be estimated. The corresponding predicted values are represented by ($\hat{h}_{1,0}^g$, $\hat{h}_{2,1}^g$, $\hat{h}_{2,0}^g$), respectively, in the algorithm from Sections 4.2.1 and 4.2.2.

To predict ($h_{1,0}^g$, $h_{2,1}^g$, $h_{2,0}^g$), we can solve a set of estimating equations. For the stratified ICE, step 1 amounts to solving for $\theta_{j,j-1}$ (j=1,...,J) in:

$$\mathbb{P}_n\left\{\phi(\bar{L}_{j-1})\left(Y_j-\mathrm{expit}\left\{\theta_{j,j-1}^T\phi(\bar{L}_{j-1})\right\}\right)I(\bar{A}_{j-1}=\bar{A}_{j-1}^g)Y_{j-1}(1-C_j)\right\}=0 \qquad \text{(C.1)}$$

where $\mathbb{P}_n(X)=\frac{1}{n}\sum_{i=1}^n X_i$ and $\mathrm{expit}\{\theta_{j,j-1}^T\phi(\bar{L}_{j-1})\}=\mathrm{E}(Y_j|Y_{j-1}=C_j=0,\bar{L}_{j-1},\bar{A}_{j-1}=\bar{A}_{j-1}^g;\theta_{j,j-1})$, and step 3 amounts to solving for $\theta_{J,k}$ in:

$$\mathbb{P}_n\left\{\phi(\bar{L}_k)\left(\hat{h}_{J,k+1}^g-\mathrm{expit}\{\theta_{J,k}^T\phi(\bar{L}_k)\}\right)I(\bar{A}_k=\bar{A}_k^g)Y_{k+1}(1-C_{k+1})\right\}=0 \qquad \text{(C.2)}$$

where $\mathrm{expit}\{\theta_{j,k}^T\phi(\bar{L}_k)\}=\mathrm{E}(\hat{h}_{j,k+1}^g|Y_{k+1}=C_{k+1}=0,\bar{L}_k,\bar{A}_k=\bar{A}_k^g;\theta_{J,k})$.

Let $\theta=(\theta_{1,0},\theta_{2,1},...,\theta_{J,J-1},\theta_{2,0},...,\theta_{J,J-2},...,\theta_{J,0})$. Under regularity conditions for M-estimators, $\hat{\theta}=(\hat{\theta}_{1,0},\hat{\theta}_{2,1},...,\hat{\theta}_{J,J-1},\hat{\theta}_{2,0},...,\hat{\theta}_{J,J-2},...,\hat{\theta}_{J,0})$ has probability limits $\theta^*=(\theta_{1,0}^*,\theta_{2,1}^*,...,\theta_{J,J-1}^*,\theta_{2,0}^*,...,\theta_{J,J-2}^*,...,\theta_{J,0}^*)$. In addition, it can be shown that the estimator $\hat{\theta}$ for θ is asymptotically linear (Tsiatis, 2006). Suppose that the true values for θ are represented by $\theta^0=(\theta_{1,0}^0,\theta_{2,1}^0,...,\theta_{J,J-1}^0,\theta_{2,0}^0,...,\theta_{J,J-2}^0,...,\theta_{J,0}^0)$. If the model for $\mathrm{E}(Y_j|Y_{j-1}=C_j=0,\bar{L}_{j-1}=\bar{l}_{j-1},\bar{A}_{j-1}=\bar{A}_{j-1}^g)$ ($\forall j\leq J$) is correctly specified, then the estimating equations (C.1) are unbiased when $\theta_{j,j-1}^*=\theta_{j,j-1}^0$. Then under sufficient regularity conditions, $\hat{h}_{j,j-1}^g$ converges in probability to $h_{j,j-1}^g$ ($\forall j\leq J$). In particular, $\hat{h}_{J,J-1}^g$ converges in probability

to $h^g_{J,J-1}$. By induction, suppose that the probability limit of $\hat{h}^g_{J,k+1}$ equals $h^g_{J,k+1}$ under correctly specified outcome regression models from times $J-2$ to $k+1$. If the model for $\mathrm{E}(h^g_{J,k+1}|Y_{k+1}=C_{k+1}=0,\bar{L}_k=\bar{l}_k,\bar{A}_k=\bar{A}^g_k)$ is correctly specified, then the estimating equations (C.2), with $\hat{h}^g_{J,k+1}$ replaced with its limit, are unbiased when $\theta^*_{J,k}=\theta^0_{J,k}$. By the continuous mapping theorem, $\hat{h}^g_{J,k}$ converges in probability to $h^g_{J,k}$.

The consistency of the pooled ICE can be proven in the same way. However, we note that in the stratified ICE, step 1 amounts to solving for $\theta_{j,j-1}$ (j=1,...,J) in:

$$\mathbb{P}_n\left\{\phi(\bar{L}_{j-1},\bar{A}_{j-1})\Big(Y_j-\mathrm{expit}\big\{\theta^T_{j,j-1}\phi(\bar{L}_{j-1},\bar{A}_{j-1})\big\}\Big)Y_{j-1}(1-C_j)\right\}=0$$

where $\mathrm{expit}\{\theta^T_{j,j-1}\phi(\bar{L}_{j-1},\bar{A}_{j-1})\}=\mathrm{E}(Y_j|Y_{j-1}=C_j=0,\bar{L}_{j-1},\bar{A}_{j-1};\theta_{j,j-1})$, and step 3 amounts to solving for $\theta_{J,k}$ in:

$$\mathbb{P}_n\left\{\phi(\bar{L}_k,\bar{A}_k)\Big(\hat{h}_{J,k+1}-\mathrm{expit}\{\theta^T_{J,k}\phi(\bar{L}_k,\bar{A}_k)\}\Big)Y_{k+1}(1-C_{k+1})\right\}=0$$

where $\mathrm{expit}\{\theta^T_{J,k}\phi(\bar{L}_k)\}=\mathrm{E}(\hat{h}^g_{J,k+1}|Y_{k+1}=C_{k+1}=0,\bar{L}_k,\bar{A}_k;\theta_{J,k})$. Note that we can also use the general sandwich variance estimator for our ICE estimators, but computation is cumbersome for multiple time points. We recommend using non-parametric bootstrap variance estimator.

C.1.1 An alternative pooled hazard-extended ICE estimator

Suppose that $f^{\mathrm{int}}(A_0|L_0)=I_{\{A^g_0\}}(A_0)$ and $f^{\mathrm{int}}(A_1|Y_1=C_1=0,\bar{L}_1,A_0)=I_{\{A^g_1\}}(A_1)$. Then it is immediate that the following equation holds:

$$\mathrm{E}_{f_{L_0}}\left(\mathrm{E}_{f^{\mathrm{int}}_{A_0}}\left[\mathrm{E}_{f_{Y_1}}\left\{Y_1+\mathrm{E}_{f_{L_1}}\left(\mathrm{E}_{f^{\mathrm{int}}_{A_1}}\left[\mathrm{E}_{f_{Y_2}}\{Y_2(1-Y_1)|Y_1,C_2=0,\bar{L}_1,\bar{A}_1\}|Y_1,C_1=0,\bar{L}_1,A_0\right]|Y_1,C_1=0,L_0,A_0\right)\middle|C_1=0,L_0,A_0\right\}\middle|L_0\right]\right)$$
$$=\mathrm{E}_{f_{L_0}}\left[\mathrm{E}_{f_{Y_1}}\left\{Y_1+\mathrm{E}_{f_{L_1}}\left(\mathrm{E}_{f_{Y_2}}\{Y_2(1-Y_1)|Y_1,C_2=0,\bar{L}_1,\bar{A}_1=\bar{A}^g_1\}|Y_1,C_1=0,L_0,A_0=A^g_0\right)\middle|C_1=0,L_0,A_0=A^g_0\right\}\right]$$

and more generally Expression (2) equals Expression (2.d) for deterministic treatment strategies. Thus under consistency, positivity and exchangeability $\mathrm{E}(Y^g_2)$ can be identified from the observed data.

We can obtain from Expression (2) and from the observed data:

$$\mathrm{E}_{f_{L_0}}\left(\mathrm{E}_{f^{\mathrm{int}}_{A_0}}\left[\mathrm{E}_{f_{Y_1}}\left\{Y_1+\mathrm{E}_{f_{L_1}}\left(\mathrm{E}_{f^{\mathrm{int}}_{A_1}}\left[\mathrm{E}_{f_{Y_2}}\{Y_2(1-Y_1)|Y_1,C_2=0,\bar{L}_1,\bar{A}_1\}|Y_1,C_1=0,\bar{L}_1,A_0\right]|Y_1,C_1=0,L_0,A_0\right)\middle|C_1=0,L_0,A_0\right\}\middle|L_0\right]\right)$$

$$=\mathrm{E}_{f_{L_0}}\left\{\mathrm{E}_{f^{\mathrm{int}}_{A_0}}\left(\mathrm{E}_{f_{Y_1}}\left[Y_1+(1-Y_1)\mathrm{E}_{f_{L_1}}\left\{\underbrace{P(Y_2=1|Y_1=C_2=0,\bar{L}_1,A_1=A^g_1,A_0)}_{h^{a^g_1}_{2,1}}|Y_1=C_1=0,L_0,A_0\right\}\middle|C_1=0,L_0,A_0\right]\middle|L_0\right)\right\}$$

$$=\mathrm{E}_{f_{L_0}}\left[\mathrm{E}_{f^{\mathrm{int}}_{A_0}}\left\{P(Y_1=1|C_1=0,L_0,A_0)+P(Y_1=0|C_1=0,L_0,A_0)\mathrm{E}_{f_{L_1}}\left(h^{a^g_1}_{2,1}|Y_1=C_1=0,L_0,A_0\right)\middle|L_0\right\}\right]$$

$$=\mathrm{E}_{f_{L_0}}\left[\underbrace{P(Y_1=1|C_1=0,L_0,A_0=A^g_0)}_{h^{a^g_0}_{1,0}}+P(Y_1=0|C_1=0,L_0,A_0=A^g_0)\mathrm{E}_{f_{L_1}}\left(h^{a^g_1}_{2,1}|Y_1=C_1=0,L_0,A_0=A^g_0\right)\right]$$

$$=\mathrm{E}_{f_{L_0}}\left[\underbrace{h^g_{1,0}+(1-h^g_{1,0})\mathrm{E}_{f_{L_1}}\left(h^{a^g_1}_{2,1}|Y_1=C_1=0,L_0,A_0=A^g_0\right)}_{h^{a^g_0}_{2,0}}\right]=\mathrm{E}_{f_{L_0}}(h^{a^g_0}_{2,0})$$

Since ($h^{a^g_0}_{1,0}$, $h^{a^g_1}_{2,1}$, $h^{a^g_0}_{2,0}$) are generally unknown, they need to be estimated. We can do so using the following estimator:

Let $\hat{h}^{a^g_{j-1}}_{j,j-1}$ be the predicted outcome for $h^{a^g_{j-1}}_{j,j-1}$ and let $\hat{h}^{a^g_k}_{J,k}$ be the predicted outcome for $h^{a^g_k}_{J,k}$.

1. For each j=1,...,J, regress Y_j on $\bar{L}_{j-1}$ and $\bar{A}_{j-1}$ in those whose Y_{j-1}=C_j=0 with parameter $\theta_{j,j-1}$.

2. For each j=1,...,J, obtain predicted values $\hat{h}^{a^g_{j-1}}_{j,j-1}$ from $\mathrm{E}(Y_j|Y_{j-1}=C_j=0,\bar{L}_{j-1},\bar{A}_{j-2},A_{j-1}=A^g_{j-1};\hat{\theta}_{j,j-1})$ for *all* of those whose Y_{j-1}=C_{j-1}=0 by fixing A_{j-1}=A^g_{j-1}. Set q=2.

3. Let k=J−q: Regress $\hat{h}^{a^g_{k+1}}_{J,k+1}$ from the previous step on $\bar{L}_k$ and $\bar{A}_k$ (observed treatment values) in those whose Y_{k+1}=C_{k+1}=0 with parameter $\theta_{J,k}$. That is, fit the model for:

$$\mathrm{E}(\hat{h}^{a^g_{k+1}}_{J,k+1}|Y_{k+1}=C_{k+1}=0,\bar{L}_k,\bar{A}_k;\theta_{J,k}) \tag{C.3}$$

4. Obtain predicted values $\hat{h}^{a^g_k}_{J,k}$ from

$$\mathrm{E}(\hat{h}^{a^g_{k+1}}_{J,k+1}|Y_{k+1}=C_{k+1}=0,\bar{L}_k,\bar{A}_{k-1},A_k=A^g_k;\hat{\theta}_{J,k})\times(1-\hat{h}^{a^g_k}_{k+1,k})+\hat{h}^{a^g_k}_{k+1,k} \tag{C.4}$$

   by fixing A_k=A^g_k for *all* of those whose Y_k=C_k=0.

5. If q<J, then let q=q+1 and return to Step 3.

By the end of this algorithm, we obtain predicted values $\hat{h}^{a^g_0}_{J,0}$ and average over L_0 to obtain an estimate for $\mathrm{E}(Y^g_J)$. Note that for NICE, under the Monte Carlo simulation, estimating Expression (1) is equivalent to estimating Expression (1.d), so there is only one way of estimating Expression (1).

C.2 Random treatment strategies

If one were to carry out a study under a random treatment intervention $f^{\text{int}}(a_j|Y_j{=}C_j{=}0,\bar{l}_j,\bar{a}_{j-1})$, by probability rules, the risk under that intervention of interest is precisely given by g-formula (1) and (2). In addition, by design a random treatment intervention satisfies the exchangeability and consistency assumptions for a set of $g{\in}\mathcal{G}$ that are observable under $f^{\text{int}}(a_j|Y_j{=}C_j{=}0,\bar{l}_j,\bar{a}_{j-1})$. To identify the risk from the observed data, we need the following.

Under our exchangeability and consistency assumptions, the g-formula (1.d) and (2.d) for the risk associated with any deterministic dynamic regimes satisfying positivity is equal to the counterfactual risk had all subjects followed that regime. It follows that the risk under the random dynamic regime with intervention distribution $f^{\text{int}}(a_j|Y_j{=}C_j{=}0,\bar{l}_j,\bar{a}_{j-1})$ is given by (1) and (2) and algebraically equals a weighted average of risk associated with the set of deterministic dynamic regimes satisfying positivity under $f^{\text{int}}(a_j|Y_j{=}C_j{=}0,\bar{l}_j,\bar{a}_{j-1})$ that only depend on past covariate history (Robins, 1986; Young *and others*, 2011). Hence, if the positivity, consistency and exchangeability holds for the aforementioned set of deterministic dynamic regimes, then we will be able to identify the risk under the random intervention of interest from the observed data.

We now define $h^{\text{int}}_{2,1}$ and $h^{\text{int}}_{2,0}$ for two time points. From Expression (2), we obtain:

$$\mathrm{E}_{f_{L_0}}\left(\mathrm{E}_{f^{\text{int}}_{A_0}}\left[\mathrm{E}_{f_{Y_1}}\left\{Y_1{+}\mathrm{E}_{f_{L_1}}\left(\mathrm{E}_{f^{\text{int}}_{A_1}}\left[\mathrm{E}_{f_{Y_2}}\{Y_2(1{-}Y_1)|Y_1,C_2{=}0,\bar{L}_1,\bar{A}_1\}|Y_1,C_1{=}0,\bar{L}_1,A_0\right]|Y_1,C_1{=}0,L_0,A_0\right)\middle|C_1{=}0,L_0,A_0\right\}\middle|L_0\right]\right)$$

$$=\mathrm{E}_{f_{L_0}}\left(\mathrm{E}_{f^{\text{int}}_{A_0}}\left[\mathrm{E}_{f_{Y_1}}\left\{Y_1{+}(1{-}Y_1)\mathrm{E}_{f_{L_1}}\left(\mathrm{E}_{f^{\text{int}}_{A_1}}\left[\mathrm{E}_{f_{Y_2}}(Y_2|Y_1{=}C_2{=}0,\bar{L}_1,\bar{A}_1)|Y_1{=}C_1{=}0,\bar{L}_1,A_0\right]|Y_1{=}C_1{=}0,L_0,A_0\right)\middle|C_1{=}0,L_0,A_0\right\}\middle|L_0\right]\right)$$

$$=\mathrm{E}_{f_{L_0}}\left[\mathrm{E}_{f^{\text{int}}_{A_0}}\left\{P(Y_1{=}1|C_1{=}0,L_0,A_0){+}P(Y_1{=}0|C_1{=}0,L_0,A_0)\mathrm{E}_{f_{L^*_1}}\left(h^{\text{int}}_{2,1}|Y_1{=}C_1{=}0,L_0,A_0\right)\middle|L_0\right\}\right]=\mathrm{E}_{f_{L^*_0}}(h^{\text{int}}_{2,0})$$

As argued previously, the first line in the equations above can be shown to equal the risk at time 2 under the treatment intervention provided that the conditions required for identifiability (as stated above) hold. **Note that the expectations taken w.r.t. A_1 and A_0 are under f^{int}.** Thus, $h^{\text{int}}_{2,1}$, $h^{\text{int}}_{2,0}$, L^*_1 and L^*_0 depend on the random treatment strategy. We consider two examples.

Example 1: "independently at each month, treat individuals with probability 0.3"

$$\mathrm{E}_{f_{L_1}}\left(\mathrm{E}_{f^{\text{int}}_{A_1}}\left[\mathrm{E}_{f_{Y_2}}(Y_2|Y_1{=}C_2{=}0,\bar{L}_1,\bar{A}_1)|Y_1{=}C_1{=}0,\bar{L}_1,A_0\right]|Y_1{=}C_1{=}0,L_0,A_0\right)$$

$$=\sum_{\forall l_1}\left\{\mathrm{E}(Y_2|Y_1{=}C_2{=}0,L_1{=}l_1,A_1{=}1,L_0,A_0){\times}0.3{+}\mathrm{E}(Y_2|Y_1{=}C_2{=}0,L_1{=}l_1,A_1{=}0,L_0,A_0){\times}0.7\right\}f(l_1|L_0,A_0,Y_1{=}C_1{=}0)$$

$$=\mathrm{E}_{f_{L^*_1}}\left\{\underbrace{\mathrm{E}(Y_2|Y_1{=}C_2{=}0,\bar{L}_1,A_1{=}1,A_0){\times}0.3{+}\mathrm{E}(Y_2|Y_1{=}C_2{=}0,\bar{L}_1,A_1{=}0,A_0){\times}0.7}_{h^{\text{int}}_{2,1}}\middle|L_0,A_0,Y_1{=}C_1{=}0)\right\}$$

where $L^*_1{=}L_1$. Thus, in this example $h^{\text{int}}_{2,1}{=}\mathrm{E}_{f^{\text{int}}_{A_1}}\left[\mathrm{E}_{f_{Y_2}}(Y_2|Y_1{=}C_2{=}0,\bar{L}_1,\bar{A}_1)|Y_1{=}C_1{=}0,\bar{L}_1,A_0\right]$.
Similarly, $h^{\text{int}}_{2,0}{=}\mathrm{E}_{f^{\text{int}}_{A_0}}\left\{P(Y_1{=}1|C_1{=}0,L_0,A_0){+}P(Y_1{=}0|C_1{=}0,L_0,A_0)\mathrm{E}_{f_{L^*_1}}\left(h^{\text{int}}_{2,1}|Y_1{=}C_1{=}0,L_0,A_0\right)\middle|L_0\right\}$.

Example 2: "independently at each month, treat all individuals whose CD4$\leq x$, but do not intervene on treatment if an individual's CD4$>x$"

$$\mathrm{E}_{f_{L_1}}\Big(\mathrm{E}_{f^{\mathrm{int}}_{A_1}}\big[\mathrm{E}_{f_{Y_2}}(Y_2|Y_1{=}C_2{=}0,\bar{L}_1,\bar{A}_1)|Y_1{=}C_1{=}0,\bar{L}_1,A_0\big]|Y_1{=}C_1{=}0,L_0,A_0\Big)$$

$$=\sum_{\forall l_1}\Big\{\mathrm{E}(Y_2|Y_1{=}C_2{=}0,l_1,A_1{=}1,L_0,A_0)I(l_1{\leq}x)+\sum_{\forall a_1}\mathrm{E}(Y_2|Y_1{=}C_2{=}0,l_1,a_1,L_0,A_0)I(l_1{>}x)f(a_1|Y_1{=}C_1{=}0,l_1,L_0,A_0)\Big\}f(l_1|L_0,A_0,Y_1{=}C_1{=}0)$$

$$=\sum_{\forall l_1,a_1}\Big\{\mathrm{E}(Y_2|Y_1{=}C_2{=}0,l_1,A_1{=}1,L_0,A_0)I(l_1{\leq}x)+\mathrm{E}(Y_2|Y_1{=}C_2{=}0,l_1,a_1,L_0,A_0)I(l_1{>}x)\Big\}f(a_1|Y_1{=}C_1{=}0,l_1,L_0,A_0)f(l_1|L_0,A_0,Y_1{=}C_1{=}0)$$

$$=\mathrm{E}_{f_{L_1^*}}\Big\{\underbrace{\mathrm{E}(Y_2|Y_1{=}C_2{=}0,\bar{L}_1,A_1{=}1,A_0)I(L_1{\leq}x)+\mathrm{E}(Y_2|Y_1{=}C_2{=}0,\bar{L}_1,\bar{A}_1)I(L_1{>}x)}_{h^{\mathrm{int}}_{2,1}}\Big|L_0,A_0,Y_1{=}C_1{=}0\Big\}$$

where $L_1^*{=}(L_1,A_1)$. Similarly, it can be shown that

$$h^{\mathrm{int}}_{2,0}{=}\Big\{P(Y_1{=}1|C_1{=}0,L_0,A_0{=}1)+P(Y_1{=}0|C_1{=}0,L_0,A_0{=}1)\mathrm{E}_{f_{L_1^*}}\big(h^{\mathrm{int}}_{2,1}|Y_1{=}C_1{=}0,L_0,A_0{=}1\big)\Big\}I(L_0{\leq}x)+$$
$$\Big\{P(Y_1{=}1|C_1{=}0,L_0,A_0)+P(Y_1{=}0|C_1{=}0,L_0,A_0)\mathrm{E}_{f_{L_1^*}}\big(h^{\mathrm{int}}_{2,1}|Y_1{=}C_1{=}0,L_0,A_0\big)\Big\}I(L_0{>}x)$$

and that

$$\mathrm{E}_{f_{L_0^*}}(h^{\mathrm{int}}_{2,0}){=}\sum_{\forall l_0,a_0}h^{\mathrm{int}}_{2,0}f(a_0|l_0)f(l_0)$$

where $L_0^*{=}(L_0,A_0)$.

The consistency of the stratified ICE estimator for the random treatment strategy can be realized by noting that it can be written as a weighted average of a set of deterministic treatment strategies. The consistency of the pooled ICE estimator can be proven in a similar way as for deterministic treatment strategies.

C.3 Examples of the pooled hazard-extended ICE estimators

Example 1: "independently at each month, treat individuals with probability 0.3"

In Step 2, predict $\hat{h}^{\mathrm{int}}_{J,J-1}$ from

$$0.3{\cdot}\mathrm{E}(Y_J|Y_{J-1}{=}C_J{=}0,\bar{L}_{J-1},\bar{A}_{J-2},A_{J-1}{=}1;\hat{\theta}_{J,J-1})+0.7{\cdot}\mathrm{E}(Y_J|Y_{J-1}{=}C_J{=}0,\bar{L}_{J-1},\bar{A}_{J-2},A_{J-1}{=}0;\hat{\theta}_{J,J-1})$$

for those whose $Y_{J-1}{=}C_{J-1}{=}0$. In Step 4, predict $\hat{h}^{\mathrm{int}}_{J,k}$ from

$$0.3{\cdot}\Big\{\mathrm{E}(\hat{h}^{\mathrm{int}}_{J,k+1}|Y_{k+1}{=}C_{k+1}{=}0,\bar{L}_k,\bar{A}_{k-1},A_k{=}1;\hat{\theta}_{J,k}){\times}\{1{-}E(Y_{k+1}|Y_k{=}C_{k+1}{=}0,\bar{L}_k,\bar{A}_{k-1},A_k{=}1;\hat{\theta}_{k+1,k})\}+$$
$$E(Y_{k+1}|Y_k{=}C_{k+1}{=}0,\bar{L}_k,\bar{A}_{k-1},A_k{=}1;\hat{\theta}_{k+1,k})\Big\}+$$
$$0.7{\cdot}\Big\{\mathrm{E}(\hat{h}^{\mathrm{int}}_{J,k+1}|Y_{k+1}{=}C_{k+1}{=}0,\bar{L}_k,\bar{A}_{k-1},A_k{=}0;\hat{\theta}_{J,k}){\times}\{1{-}E(Y_{k+1}|Y_k{=}C_{k+1}{=}0,\bar{L}_k,\bar{A}_{k-1},A_k{=}0;\hat{\theta}_{k+1,k})\}+$$
$$E(Y_{k+1}|Y_k{=}C_{k+1}{=}0,\bar{L}_k,\bar{A}_{k-1},A_k{=}0;\hat{\theta}_{k+1,k})\Big\}$$

for those whose Y_k=C_k=0.

Example 2: "independently at each month, treat all individuals whose CD4$\leq x$, but do not intervene on treatment if an individual's CD4$>x$"

In Step 2, predict $\hat{h}^{\text{int}}_{J,J-1}$ from $\text{E}(Y_J|Y_{J-1}{=}C_J{=}0,\bar{L}_{J-1},\bar{A}_{J-2},A_{J-1}{=}1;\hat{\theta}_{J,J-1})$ in those whose $CD4_{J-1}\leq x$ and Y_{J-1}=C_{J-1}=0, and predict $\hat{h}^{\text{int}}_{J,J-1}$ from $\text{E}(Y_J|Y_{J-1}{=}C_J{=}0,\bar{L}_{J-1},\bar{A}_{J-1};\hat{\theta}_{J,J-1})$ in those whose $CD4_{J-1}>x$ and Y_{J-1}=C_{J-1}=0. In Step 4, predict $\hat{h}^{\text{int}}_{J,k}$ from

$$\text{E}(\hat{h}^{\text{int}}_{J,k+1}|Y_{k+1}{=}C_{k+1}{=}0,\bar{L}_k,\bar{A}_{k-1},A_k{=}1;\hat{\theta}_{J,k})\times\{1-E(Y_{k+1}|Y_k{=}C_{k+1}{=}0,\bar{L}_k,\bar{A}_{k-1},A_k{=}1;\hat{\theta}_{k+1,k})\}+ E(Y_{k+1}|Y_k{=}C_{k+1}{=}0,\bar{L}_k,\bar{A}_{k-1},A_k{=}1;\hat{\theta}_{k+1,k})$$

in those whose $CD4_k\leq x$ and Y_k=C_k=0, and predict $\hat{h}^{\text{int}}_{J,k}$ from

$$\text{E}(\hat{h}^{\text{int}}_{J,k+1}|Y_{k+1}{=}C_{k+1}{=}0,\bar{L}_k,\bar{A}_k;\hat{\theta}_{J,k})\times\{1-E(Y_{k+1}|Y_k{=}C_{k+1}{=}0,\bar{L}_k,\bar{A}_k;\hat{\theta}_{k+1,k})\}+ E(Y_{k+1}|Y_k{=}C_{k+1}{=}0,\bar{L}_k,\bar{A}_k;\hat{\theta}_{k+1,k})$$

in those whose $CD4_k>x$ and Y_k=C_k=0.

D. Classical ICE estimator

D.1 Deterministic treatment strategies

D.1.1 Procedure 1: stratifying on treatment history

The following procedure can be implemented in existing an R package called ltmle (Lendle *and others*, 2017). Iteratively, the algorithm for time J is defined as follows:

1. Regress Y_J on $\bar{L}_{J-1}$ in those whose Y_{J-1}=C_J=0 and who followed the regime $\bar{A}_{J-1}$=$\bar{A}^g_{J-1}$ with parameter $\theta_{J,J-1}$.
2. Obtain predicted values $\hat{h}^g_{J,J-1}$ from E(Y_J|Y_{J-1}=C_J=0,$\bar{L}_{J-1}$,$\bar{A}_{J-1}$=$\bar{A}^g_{J-1}$;$\hat{\theta}_{J,J-1}$) by fixing A_{J-1}=A^g_{J-1} in those whose (Y_{J-1},C_{J-1},$\bar{A}_{J-2}$)=(0,0,$\bar{A}^g_{J-2}$). Set q=2.
3. Let k=J−q: Let $\hat{Q}^g_{J,k+1}$=$\hat{h}^g_{J,k+1}$ for those whose Y_{k+1}=0 and let $\hat{Q}^g_{J,k+1}$=1 for those whose Y_{k+1}=1. Regress $\hat{Q}^g_{J,k+1}$ on $\bar{L}_k$ in those whose Y_k=C_{k+1}=0 and who followed the regime $\bar{A}_k$=$\bar{A}^g_k$ with parameter $\theta_{J,k}$.
4. Obtain predicted values $\hat{h}^g_{J,k}$ from E($\hat{Q}^g_{J,k+1}$|Y_k=C_{k+1}=0,$\bar{L}_k$,$\bar{A}_k$=$\bar{A}^g_k$;$\hat{\theta}_{J,k}$) by fixing A_k=A^g_k in those whose (Y_k,C_k,$\bar{A}_{k-1}$)=(0,0,$\bar{A}^g_{k-1}$).
5. If q<J, then let q=q+1 and return to step 3.
6. Average over $\hat{h}^g_{J,0}$.

D.1.2 Procedure 2: pooling over treatment history

1. Regress Y_J on $\bar{L}_{J-1}$ and $\bar{A}_{J-1}$ in those whose Y_{J-1}=C_J=0 with parameter $\theta_{J,J-1}$.
2. Obtain predicted values $\hat{h}^g_{J,J-1}$ from E(Y_J|Y_{J-1}=C_J=0,$\bar{L}_{J-1}$,$\bar{A}_{J-1}$=$\bar{A}^g_{J-1}$;$\hat{\theta}_{J,J-1}$) by fixing $\bar{A}_{J-1}$=$\bar{A}^g_{J-1}$ for *all* of those whose Y_{J-1}=C_{J-1}=0. Set q=2.
3. Let k=J−q: Let $\hat{Q}^g_{J,k+1}$=$\hat{h}^g_{J,k+1}$ for those whose Y_{k+1}=0 and let $\hat{Q}^g_{J,k+1}$=1 for those whose Y_{k+1}=1. Regress $\hat{Q}^g_{J,k+1}$ on $\bar{L}_k$ and $\bar{A}_k$ in those whose Y_k=C_{k+1}=0 with parameter $\theta_{J,k}$.
4. Obtain predicted values $\hat{h}^g_{J,k}$ from E($\hat{Q}^g_{J,k+1}$|Y_k=C_{k+1}=0,$\bar{L}_k$,$\bar{A}_k$=$\bar{A}^g_k$;$\hat{\theta}_{J,k}$) by fixing $\bar{A}_k$=$\bar{A}^g_k$ for *all* of those whose Y_k=C_k=0.
5. If q<J, then let q=q+1 and return to step 3.
6. Average over $\hat{h}^g_{J,0}$.

Setting $\hat{Q}^g_{J,k+1}$ to be 1 if Y_k=1 does not effectively utilize information on the hazard of the event at time k and is consequently inefficient. Hence, this estimation procedure is less

efficient than the ICE procedure in Section 4.2.2, as seen in our simulation studies.

D.2 Random treatment strategies

1. Regress Y_J on $\bar{L}_{J-1}$ and $\bar{A}_{J-1}$ in those whose $Y_{J-1}{=}C_J{=}0$ with parameter $\theta_{J,J-1}$.

2.(a) If $f^{\text{int}}(a_{J-1}|Y_{J-1}{=}C_{J-1}{=}0,\bar{l}_{J-1},\bar{a}_{J-2})$ deterministically equals 1 for some $(\bar{l}_{J-1},\bar{a}_{J-2})$: then using the fitted model $\text{E}(Y_J|Y_{J-1}{=}C_J{=}0,\bar{L}_{J-1},\bar{A}_{J-1};\hat{\theta}_{J,J-1})$, predict $\hat{h}^{\text{int}}_{J,J-1}$ by fixing $A_{J-1}{=}a_{J-1}$ consistent with the treatment strategy for those with this observed history of $(\bar{l}_{J-1},\bar{a}_{J-2})$ and whose $Y_{J-1}{=}C_{J-1}{=}0$. Set $q{=}2$.

(b) If $f^{\text{int}}(a_{J-1}|Y_{J-1}{=}C_{J-1}{=}0,\bar{l}_{J-1},\bar{a}_{J-2}){=}f(a_{J-1}|Y_{J-1}{=}C_{J-1}{=}0,\bar{l}_{J-1},\bar{a}_{J-2})$, the observed treatment distribution, for some $(\bar{l}_{J-1},\bar{a}_{J-2})$: then using the fitted model $\text{E}(Y_J|Y_{J-1}{=}C_J{=}0,\bar{L}_{J-1},\bar{A}_{J-1};\hat{\theta}_{J,J-1})$, predict $\hat{h}^{\text{int}}_{J,J-1}$ by setting A_{J-1} to the observed treatment value at time $J{-}1$ for those with this observed history of $(\bar{l}_{J-1},\bar{a}_{J-2})$ and whose $Y_{J-1}{=}C_{J-1}{=}0$

(c) Otherwise obtain predicted values $\hat{h}^{\text{int}}_{J,J-1}$ by estimating

$$\sum_{\forall a_{J-1}} \text{E}(Y_J|Y_{J-1}{=}C_J{=}0,\bar{L}_{J-1},\bar{A}_{J-2},A_{J-1}{=}a_{J-1};\hat{\theta}_{J,J-1})f^{\text{int}}(a_{J-1}|Y_{J-1}{=}C_{J-1}{=}0,\bar{L}_{J-1},\bar{A}_{J-2})$$

for those who do not meet the conditions in (a) and (b) and whose $Y_{J-1}{=}C_{J-1}{=}0$.

(d) Set $q{=}2$.

3. Let $k{=}J{-}q$: Let $\hat{Q}^{\text{int}}_{J,k+1}{=}\hat{h}^{\text{int}}_{J,k+1}$ for those whose $Y_{k+1}{=}0$ and let $\hat{Q}^{\text{int}}_{J,k+1}{=}1$ for those whose $Y_{k+1}{=}1$. Regress the predicted values $\hat{Q}^{\text{int}}_{J,k+1}$ from the previous step on $\bar{L}_k$ and $\bar{A}_k$ in those whose $Y_k{=}C_{k+1}{=}0$ with parameter $\theta_{J,k}$.

4. (a) If $f^{\text{int}}(a_k|Y_k{=}C_k{=}0,\bar{l}_k,\bar{a}_{k-1})$ deterministically equals 1 for some $(\bar{l}_k,\bar{a}_{k-1})$: then using the fitted model

$$\text{E}(\hat{Q}^{\text{int}}_{J,k+1}|Y_k{=}C_{k+1}{=}0,\bar{L}_k,\bar{A}_k;\hat{\theta}_{J,k}) \tag{D.1}$$

predict $\hat{h}^{\text{int}}_{J,k}$ by fixing $A_k{=}a_k$ according to the random treatment strategy for of those with this observed history of $(\bar{l}_k,\bar{a}_{k-1})$ and whose $Y_k{=}C_k{=}0$.

(b) If $f^{\text{int}}(a_k|Y_k{=}C_k{=}0,\bar{l}_k,\bar{a}_{k-1}){=}f(a_k|Y_k{=}C_k{=}0,\bar{l}_k,\bar{a}_{k-1})$ for some $(\bar{l}_k,\bar{a}_{k-1})$: then using Expression (D.1), predict $\hat{h}^{\text{int}}_{J,k}$ by setting A_k to the observed treatment value at time k for those with this observed history of $(\bar{l}_k,\bar{a}_{k-1})$ and whose $Y_k{=}C_k{=}0$.

(c) Otherwise predict $\hat{h}^{\text{int}}_{J,k}$ by estimating

$$\sum_{\forall a_k} f^{\text{int}}(a_k|Y_k{=}C_k{=}0,\bar{L}_k,\bar{A}_{k-1})\text{E}(\hat{Q}^{\text{int}}_{J,k+1}|Y_k{=}C_{k+1}{=}0,\bar{L}_k,\bar{A}_k;\hat{\theta}_{J,k})$$

for those who do not meet the conditions in (a) and (b) and whose $Y_k{=}C_k{=}0$.

5. If $q{<}J$, then let $q{=}q{+}1$ and return to Step 3.

6. Average over $\hat{h}^{\text{int}}_{J,0}$.

We can also show how Expression (2.d) correspond to the classical ICE:

$$\mathrm{E}_{f_{L_0}}\left(\mathrm{E}_{f_{Y_1}}\left\{Y_1+\mathrm{E}_{f_{L_1}}\left(\mathrm{E}_{f_{Y_2}}\{Y_2(1-Y_1)|Y_1,C_2=0,\bar{L}_1,\bar{A}_1=\bar{A}^g_1\}|Y_1,C_1=0,L_0,A_0=A^g_0\right)\middle|C_1=0,L_0,A_0=A^g_0\right\}\middle|\right)$$

$$=\mathrm{E}_{f_{L_0}}\left\{\mathrm{E}_{f_{Y_1}}\left[Y_1+(1-Y_1)\mathrm{E}_{f_{L_1}}\left\{\underbrace{P(Y_2=1|Y_1=C_2=0,\bar{L}_1,\bar{A}_1=\bar{A}^g_1)}_{h^g_{2,1}}|Y_1=C_1=0,L_0,A_0=A^g_0\right\}\middle|C_1=0,L_0,A_0=A^g_0\right]\right\}$$

$$=\mathrm{E}_{f_{L_0}}\left\{\mathrm{E}_{Y_1,L_1}\left[\underbrace{Y_1+(1-Y_1)h^g_{2,1}}_{Q^g_{2,1}}\middle|C_1=0,L_0,A_0=A^g_0\right]\right\}=\mathrm{E}_{f_{L_0}}\left\{\underbrace{\mathrm{E}_{Y_1,L_1}\left(Q^g_{2,1}\middle|C_1=0,L_0,A_0=A^g_0\right)}_{h^g_{2,0}}\right\}=\mathrm{E}_{f_{L_0}}(h^g_{2,0})$$

Note that regressing $Y_1+(1-Y_1)\hat{h}^g_{2,1}$ on (L_0,A_0) with a particular functional form is the same as regressing $P^{\text{sat}}(Y_1=1|L_0,A_0)+P^{\text{sat}}(Y_1=0|L_0,A_0)\hat{E}(\hat{h}^g_{2,1}|Y_1=0,L_0,A_0)$ on (L_0,A_0) with the same functional form, where $P^{\text{sat}}(Y_1=1|L_0,A_0)$ is the nonparametric estimate of the hazard at time 1 given L_0 and A_0, and the regression of $\hat{h}^g_{2,1}$ on (L_0,A_0) (in order to estimate $\hat{E}(\hat{h}^g_{2,1}|L_0,A_0)$) is also of the same functional form.

D.3 Heuristic reason why classical ICE estimators may be less efficient than hazard-extended ICE estimators

We utilize the idea of Rao-Blackwellization often used in Monte Carlo computation. Suppose that we have two variables X and Y. Both Y and $\mathrm{E}(Y|X)$ are unbiased estimators of $\mathrm{E}(Y)$, but it can be shown that $\mathrm{Var}(Y)\geq\mathrm{Var}\{\mathrm{E}(Y|X)\}$ and so it would be preferable to use $\mathrm{E}(Y|X)$. Indeed, we can reduce the variance of an estimator if we can replace a random sample with its expectation (see also Schafer, 1997; Robins, 1986; Liu, 2008).

Suppose that we observe (L,A,Y,Z) and we are interested in $\mu=\mathrm{E}(Z)$. Then an unbiased estimator of μ is $\mathbb{P}_n\{\mathrm{E}(Z|Y=0,A,L)(1-Y)+Y\}$. Let $\tilde{\mu}=\mathrm{E}(Z|Y=0,A,L)(1-Y)+Y$. We now show that the variance for $\tilde{\mu}$ can be improved if we exploit further knowledge about the distribution of Y. If we now define $\mu^*=\mathrm{E}(\tilde{\mu}|A,L)=\mathrm{E}(Z|Y=0,A,L)\mathrm{P}(Y=0|A,L)+\mathrm{P}(Y=1|A,L)$, then $\mathbb{P}_n(\mu^*)$ is another unbiased estimator of $\mathrm{E}(Z)$. Moreover,

$$\begin{aligned}\mathrm{E}\{(\mu^*-\mu)^2\}&=\mathrm{E}[\{\mathrm{E}(\tilde{\mu}|A,L)-\mu\}^2]\\&=\mathrm{E}[\{\mathrm{E}(\tilde{\mu}-\mu|A,L)\}^2]\\&\leq\mathrm{E}[\mathrm{E}\{(\tilde{\mu}-\mu)^2|A,L\}]=\mathrm{E}\{(\tilde{\mu}-\mu)^2\}\end{aligned}$$

The inequality follows by Jensen's inequality. This implies that the unbiased estimate obtained from μ^* will be at least as efficient than the unbiased estimate obtained from $\tilde{\mu}$. By the same argument we can see that Rao-Blackwellization can also be useful in our hazard-extended ICE procedure for the g-formula. We would expect that it may be more efficient if we can exploit knowledge about the hazard of past outcomes Y_j (for all $j\leq J$) or if we have a good approximation of the hazard, than implicitly integrating over Y_j nonparametrically (see paragraph directly above Section D.3).

E. Simulation description and additional results

In the simulation study described in Sections 4.3 and 5.4, $L_0 \sim \text{Ber}(0.5)$. For $j \geq 1$:

$$\begin{aligned}
C_j &\sim \text{Ber}\{\text{expit}(-3-A_{j-1}+0.75L_{j-1})\}, && \text{if } Y_{j-1}=0,\\
C_j &=1 \text{ if } C_{j-1}=1, && \text{otherwise } C_j=\emptyset\\
Y_j &\sim \text{Ber}\{\text{expit}(-2-2A_{j-1}+L_{j-1})\}, && \text{if } Y_{j-1}=0 \text{ and } C_j=0,\\
Y_j &=1 \text{ if } Y_{j-1}=1, && \text{otherwise } Y_j=\emptyset\\
L_j &\sim \text{Ber}\{\text{expit}(-2-2A_{j-1})\}, && \text{if } Y_j=0 \text{ and } L_{j-1}=0,\\
L_j &=1 \text{ if } Y_j=0 \text{ and } L_{j-1}=1, && \text{otherwise } L_j=\emptyset\\
A_j &\sim \text{Ber}\{\text{expit}(-1.5+L_j)\}, && \text{if } Y_j=0 \text{ and } A_{j-1}=0,\\
A_j &=1 \text{ if } Y_j=0 \text{ and } A_{j-1}=1, && \text{otherwise } A_j=\emptyset
\end{aligned}$$

The stratified ICE estimator for our random treatment strategy requires one to estimate $\sum_{g\in\mathcal{G}} wt(g)E(Y_J^g)$, where $wt(g)=q_0^g$. Since we only have one time-varying covariate L_j, we can work out $q_{J-1}^g(\bar{l}_{J-2})$ and $q_j^g(\bar{l}_{j-1})$ for each $J=1,...,5$.

When $J=1$, $q_0^g=f(a_0^g|l_{0,1})f(a_0^g|l_{0,2})$, where $l_{0,1}=0$ and $l_{0,2}=1$. Note that $f(a_0^g|l_0)\equiv f^{\text{int}}(a_0|l_0)$. There are two deterministic treatment strategies $\{g_1,g_2\}\in\mathcal{G}$ with non-zero weights, and they are $a_0^{g_1}=0$ and $a_0^{g_2}=l_0$. When $a_0^{g_1}=0$, for example, $q_0^{g_1}=f^{\text{int}}(0|l_{0,2})$, which in turn equals the observed treatment distribution $\text{P}(A_0=0|L_0=1)$. When $a_0^{g_2}=L_0$, $q_0^{g_2}=f^{\text{int}}(l_{0,2}|l_{0,2})=\text{P}(A_0=1|L_0=1)$. Clearly, $wt(g_1)+wt(g_2)=1$.

The number of deterministic treatment strategies corresponding to the random treatment strategy increases with the number of follow-ups. When $J=2$, there are four deterministic treatment strategies $\{g_1,g_2,g_3,g_4\}\in\mathcal{G}$ with non-zero weights, and they are $(a_0^{g_1},a_1^{g_1})=(l_0,l_1)$, $(a_0^{g_2},a_1^{g_2})=(0,l_1)$, $(a_0^{g_3},a_1^{g_3})=(l_0,l_0)$, $(a_0^{g_4},a_1^{g_4})=(0,l_0)$.

When $J=3$, there are eight deterministic treatment strategies $\{g_1,...,g_7,g_8\}\in\mathcal{G}$ with non-zero weights, and they are $(a_0^{g_1},a_1^{g_1},a_2^{g_1})=(l_0,l_1,l_2)$, $(a_0^{g_2},a_1^{g_2},a_2^{g_2})=(0,l_1,l_2)$, $(a_0^{g_3},a_1^{g_3},a_2^{g_3})=(l_0,l_0,l_1)$, $(a_0^{g_4},a_1^{g_4},a_2^{g_4})=(l_0,l_1,l_1)$, $(a_0^{g_5},a_1^{g_5},a_2^{g_5})=(0,l_0,l_1)$, $(a_0^{g_6},a_1^{g_6},a_2^{g_6})=(0,l_1,l_1)$, $(a_0^{g_7},a_1^{g_7},a_2^{g_7})=(0,l_0,l_2)$, $(a_0^{g_8},a_1^{g_8},a_2^{g_8})=(l_0,l_0,l_2)$.

When $J=4$, there are sixteen deterministic treatment strategies $\{g_1,...,g_{15},g_{16}\}\in\mathcal{G}$ with non-zero weights, and they are $(a_0^{g_1},a_1^{g_1},a_2^{g_1},a_3^{g_1})=(l_0,l_1,l_2,l_3)$, $(a_0^{g_2},a_1^{g_2},a_2^{g_2},a_3^{g_2})=(0,l_1,l_2,l_3)$, $(a_0^{g_3},a_1^{g_3},a_2^{g_3},a_3^{g_3})=(0,l_1,l_2,l_2)$, $(a_0^{g_4},a_1^{g_4},a_2^{g_4},a_3^{g_4})=(l_0,l_0,l_2,l_2)$, $(a_0^{g_5},a_1^{g_5},a_2^{g_5},a_3^{g_5})=(l_0,l_0,l_2,l_3)$, $(a_0^{g_6},a_1^{g_6},a_2^{g_6},a_3^{g_6})=(l_0,l_0,l_1,l_2)$, $(a_0^{g_7},a_1^{g_7},a_2^{g_7},a_3^{g_7})=(l_0,l_0,l_1,l_3)$, $(a_0^{g_8},a_1^{g_8},a_2^{g_8},a_3^{g_8})=(l_0,l_1,l_1,l_2)$, $(a_0^{g_9},a_1^{g_8},a_2^{g_8},a_3^{g_8})=(l_0,l_1,l_1,l_3)$, $(a_0^{g_{10}},a_1^{g_{10}},a_2^{g_{10}},a_3^{g_{10}})=(0,l_0,l_1,l_2)$, $(a_0^{g_{11}},a_1^{g_{11}},a_2^{g_{11}},a_3^{g_{11}})=(0,l_0,l_1,l_3)$, $(a_0^{g_{12}},a_1^{g_{12}},a_2^{g_{12}},a_3^{g_{12}})=(0,l_1,l_1,l_2)$, $(a_0^{g_{13}},a_1^{g_{13}},a_2^{g_{13}},a_3^{g_{13}})=(0,l_1,l_1,l_3)$, $(a_0^{g_{14}},a_1^{g_{14}},a_2^{g_{14}},a_3^{g_{14}})=(0,l_0,l_2,l_2)$, $(a_0^{g_{15}},a_1^{g_{15}},a_2^{g_{15}},a_3^{g_{15}})=(0,l_0,l_2,l_3)$, $(a_0^{g_{16}},a_1^{g_{16}},a_2^{g_{16}},a_3^{g_{16}})=(l_0,l_1,l_2,l_2)$.

When $J=5$, there are thirty-two deterministic treatment strategies $\{g_1,...,g_{31},g_{32}\}\in\mathcal{G}$ with non-zero weights, and they are: $(a_0^{g_1},a_1^{g_1},a_2^{g_1},a_3^{g_1},a_4^{g_1})=(l_0,l_1,l_2,l_3,l_4)$, $(a_0^{g_2},a_1^{g_2},a_2^{g_2},a_3^{g_2},a_4^{g_2})=(0,l_1,l_2,l_3,l_4)$, $(a_0^{g_3},a_1^{g_3},a_2^{g_3},a_3^{g_3},a_4^{g_3})=(0,l_1,l_2,l_2,l_3)$, $(a_0^{g_4},a_1^{g_4},a_2^{g_4},a_3^{g_4},a_4^{g_4})=(0,l_1,l_2,l_2,l_4)$, $(a_0^{g_5},a_1^{g_5},a_2^{g_5},a_3^{g_5},a_4^{g_5})=(0,l_1,l_2,l_3,l_3)$, $(a_0^{g_6},a_1^{g_6},a_2^{g_6},a_3^{g_6},a_4^{g_6})=(l_0,l_0,l_2,l_2,l_3)$, $(a_0^{g_7},a_1^{g_7},a_2^{g_7},a_3^{g_7},a_4^{g_7})=(l_0,l_0,l_2,l_2,l_4)$, $(a_0^{g_8},a_1^{g_8},a_2^{g_8},a_3^{g_8},a_4^{g_8})=(l_0,l_0,l_2,l_3,l_3)$,

$(a_0^{g_9},a_1^{g_8},a_2^{g_8},a_3^{g_8},a_4^{g_9})=(l_0,l_0,l_2,l_3,l_4)$, $(a_0^{g_{10}},a_1^{g_{10}},a_2^{g_{10}},a_3^{g_{10}},a_4^{g_{10}})=(l_0,l_0,l_1,l_2,l_3)$,
$(a_0^{g_{11}},a_1^{g_{11}},a_2^{g_{11}},a_3^{g_{11}},a_4^{g_{11}})=(l_0,l_0,l_1,l_2,l_4)$, $(a_0^{g_{12}},a_1^{g_{12}},a_2^{g_{12}},a_3^{g_{12}},a_4^{g_{12}})=(l_0,l_0,l_1,l_3,l_3)$,
$(a_0^{g_{13}},a_1^{g_{13}},a_2^{g_{13}},a_3^{g_{13}},a_4^{g_{13}})=(l_0,l_0,l_1,l_3,l_4)$, $(a_0^{g_{14}},a_1^{g_{14}},a_2^{g_{14}},a_3^{g_{14}},a_4^{g_{14}})=(l_0,l_1,l_1,l_2,l_3)$,
$(a_0^{g_{15}},a_1^{g_{15}},a_2^{g_{15}},a_3^{g_{15}},a_4^{g_{15}})=(l_0,l_1,l_1,l_2,l_4)$, $(a_0^{g_{16}},a_1^{g_{16}},a_2^{g_{16}},a_3^{g_{16}},a_4^{g_{16}})=(l_0,l_1,l_1,l_3,l_3)$,
$(a_0^{g_{17}},a_1^{g_{17}},a_2^{g_{17}},a_3^{g_{17}},a_4^{g_{17}})=(l_0,l_1,l_1,l_3,l_4)$, $(a_0^{g_{18}},a_1^{g_{18}},a_2^{g_{18}},a_3^{g_{18}},a_4^{g_{18}})=(0,l_0,l_1,l_2,l_3)$,
$(a_0^{g_{19}},a_1^{g_{19}},a_2^{g_{19}},a_3^{g_{19}},a_4^{g_{19}})=(0,l_0,l_1,l_2,l_4)$, $(a_0^{g_{20}},a_1^{g_{20}},a_2^{g_{20}},a_3^{g_{20}},a_4^{g_{20}})=(0,l_0,l_1,l_3,l_3)$,
$(a_0^{g_{21}},a_1^{g_{21}},a_2^{g_{21}},a_3^{g_{21}},a_4^{g_{21}})=(0,l_0,l_1,l_3,l_4)$, $(a_0^{g_{22}},a_1^{g_{22}},a_2^{g_{22}},a_3^{g_{22}},a_4^{g_{22}})=(0,l_1,l_1,l_2,l_3)$,
$(a_0^{g_{23}},a_1^{g_{23}},a_2^{g_{23}},a_3^{g_{23}},a_4^{g_{23}})=(0,l_1,l_1,l_2,l_4)$, $(a_0^{g_{24}},a_1^{g_{24}},a_2^{g_{24}},a_3^{g_{24}},a_4^{g_{24}})=(0,l_1,l_1,l_3,l_3)$,
$(a_0^{g_{25}},a_1^{g_{25}},a_2^{g_{25}},a_3^{g_{25}},a_4^{g_{25}})=(0,l_1,l_1,l_3,l_4)$, $(a_0^{g_{26}},a_1^{g_{26}},a_2^{g_{26}},a_3^{g_{26}},a_4^{g_{26}})=(0,l_0,l_2,l_2,l_3)$,
$(a_0^{g_{27}},a_1^{g_{27}},a_2^{g_{27}},a_3^{g_{27}},a_4^{g_{27}})=(0,l_0,l_2,l_2,l_4)$, $(a_0^{g_{28}},a_1^{g_{28}},a_2^{g_{28}},a_3^{g_{28}},a_4^{g_{28}})=(0,l_0,l_2,l_3,l_3)$,
$(a_0^{g_{29}},a_1^{g_{29}},a_2^{g_{29}},a_3^{g_{29}},a_4^{g_{29}})=(0,l_0,l_2,l_3,l_4)$, $(a_0^{g_{30}},a_1^{g_{30}},a_2^{g_{30}},a_3^{g_{30}},a_4^{g_{30}})=(l_0,l_1,l_2,l_2,l_3)$,
$(a_0^{g_{31}},a_1^{g_{31}},a_2^{g_{31}},a_3^{g_{31}},a_4^{g_{31}})=(l_0,l_1,l_2,l_2,l_4)$, $(a_0^{g_{32}},a_1^{g_{32}},a_2^{g_{32}},a_3^{g_{32}},a_4^{g_{32}})=(l_0,l_1,l_2,l_3,l_3)$.

[Table 1 about here.]

[Table 2 about here.]

E.1 Additional example of a random treatment strategy

In this section, we consider a simple simulation study for one time point. We generate (L_0,A_0,Y_1), where $L_0=(L_{10},L_{20})$, $L_{10}\sim\text{Ber}(0.5)$, $L_{20}\sim\text{Ber}(0.5)$, $A_0\sim\text{Ber}\{\text{expit}(1-2L_{10}+L_{20}+L_{10}L_{20})\}$ and $Y_1\sim\text{Ber}\{\text{expit}(-1+A_0+2L_{10}-2L_{20}+L_{10}L_{20})\}$. We consider a delayed initiation of antiretroviral therapy until $L_{10}=0$ (e.g. L_{10} is an indicator of a high CD4 count), but allowing for 1-year grace period from the time that a subject first drops to a low CD4 count.

Now that we are considering two time-varying covariates, the cardinality of $|\mathcal{L}_0|$ is four. Therefore, $q_0^g=f(a_0^g|l_{10,1},l_{20,1})f(a_0^g|l_{10,2},l_{20,2})f(a_0^g|l_{10,3},l_{20,3})f(a_0^g|l_{10,4},l_{20,4})$, where $(l_{10,1},l_{20,1})=(0,0)$, $(l_{10,2},l_{20,2})=(0,1)$, $(l_{10,3},l_{20,3})=(1,0)$ and $(l_{10,4},l_{20,4})=(1,1)$. There are four deterministic treatment strategies $\{g_1,g_2,g_3,g_4\}\in\mathcal{G}$ with non-zero weights, and they are $a_0^{g_1}=(1-l_{10})l_{20}$, $a_0^{g_2}=(1-l_{10})(1-l_{20})$, $a_0^{g_3}=1-l_{10}$, $a_0^{g_4}=0$. The bias and standard error from this simulation are 0.0001 and 0.0205, respectively.

We have shown previously that the number of deterministic treatment strategies in $\mathcal{G}$ with non-zero weights corresponding to a random treatment strategy increases with the number of follow-up times. It is easy to see here that when we increase the number of time-varying covariates, it will become more difficult to list every deterministic treatment strategy that is consistent with the random treatment strategy.

E.2 Simulation study with a continuous time-varying confounder

In this section, we consider a simulation study for $J=5$. We generate $(L_0,A_0,C_1,Y_1,...,C_5,Y_5)$, where $L_j=(L_{1j},L_{2j})$, $L_{10}\sim\text{Ber}(0.5)$, $L_{20}\sim\text{Normal}(2,1)$, $A_0\sim\text{Ber}\{\text{expit}(-1+L_{10}-0.25L_{20})\}$, $C_1\sim\text{Ber}\{\text{expit}(-3-A_0+0.75L_{10}-0.5L_{20})\}$ and $Y_1\sim\text{Ber}\{\text{expit}(-1-2A_0+2L_{10}-0.5L_{20})\}$. In

addition for $j \geq 1$:

$$
\begin{aligned}
C_j &\sim \text{Ber}\{\text{expit}(-3-A_{j-1}+0.75L_{1,j-1}-0.5L_{2,j-1})\}, && \text{if } Y_{j-1}=0,\\
C_j &=1 \text{ if } C_{j-1}=1, && \text{otherwise } C_j=\emptyset\\
Y_j &\sim \text{Ber}\{\text{expit}(-1-2A_{j-1}+2L_{1,j-1}-0.5L_{2,j-1})\}, && \text{if } Y_{j-1}=0 \text{ and } C_j=0,\\
Y_j &=1 \text{ if } Y_{j-1}=1, && \text{otherwise } Y_j=\emptyset\\
L_{1j} &\sim \text{Ber}\{\text{expit}(-2-2A_{j-1}-0.5L_{2,j-1})\}, && \text{if } Y_j=0 \text{ and } L_{j-1}=0,\\
L_{1j} &=1 \text{ if } Y_j=0 \text{ and } L_{1,j-1}=1, && \text{otherwise } L_{1j}=\emptyset\\
L_{2j} &\sim \text{Normal}(2+A_{j-1}-L_{1,j-1}+0.5L_{2,j-1},1), && \text{if } Y_j=0, \text{ otherwise } L_{2j}=\emptyset\\
A_j &\sim \text{Ber}\{\text{expit}(-1+L_{1j}-0.25L_{2j})\}, && \text{if } Y_j=0 \text{ and } A_{j-1}=0,\\
A_j &=1 \text{ if } Y_j=0 \text{ and } A_{j-1}=1, && \text{otherwise } A_j=\emptyset
\end{aligned}
$$

We consider deterministic and random treatment strategies: (1) a delayed initiation of treatment until the first time $L_{2j}<2.75$ (e.g. transformed CD4 count) or when $L_{1j}=1$ (e.g. AIDS), whichever happens first, (2) a delayed initiation of treatment that also allows for 1-year grace period from the time that this threshold is met.

The true parametric models for the g-formula estimator based on NICE were the ones used for the simulation because we used a data generating mechanism without unmeasured common causes of covariates and outcomes. Similarly, the correct treatment model for IP weighting was the one used in the data generation process. The true risks of death were calculated by using these models to generate a Monte Carlo sample of size 10^7. We use the true outcome regression model in the nonstratified ICE to model the hazard at each time point. For all other models in ICE, we include L_{1j}, L_{2j} and W_j and their pairwise interactions.

[Table 3 about here.]

[Table 4 about here.]

F. Computation considerations for ICE and NICE

One way to determine the computational efficiency under each method is time complexity, which describes how the runtime of a function grows as the size of our input grows. In a GLM, the time complexity is given by $O(p^3+Np^2)$, where N is the number of rows (person-time) and p is the number of covariates in our GLM (H2O, 2017).

In our HIV-CAUSAL Collaboration analysis, there are more than 2,500,000 person-month observations and 55,826 individuals. Hence, the time complexity of a pooled regression model fitted using the person-time observations will increase the computational cost greatly.

Generally in NICE, we fit separate models for the time-varying confounders, treatment and outcome by pooling the person-time observations across time. For example, in HIV-CAUSAL, each of models for the outcome, time-varying confounders and treatment uses over 2,500,000 observations. If there are a lot of time-varying confounders, then ICE can be more computationally efficient than NICE. This is because in ICE, we do not fit regression models for the time-varying confounders or treatment. Moreover, the regression models in Step 3 of the algorithm for ICE generally require a much smaller data set. For example, in HIV-CAUSAL, each of the regression models uses 55,826 observations or less.

A computational disadvantage of the ICE is that at the end of the algorithm for ICE, we are only able to estimate the average causal effect at time J. Whereas at the end of the algorithm for NICE, we are able to estimate the average causal effect for all time points through time J (i.e. time points 1,2,...,J). Furthermore, the ICE estimator requires a separate set of regression models for each treatment strategy, whereas NICE only requires a one-time estimation of the conditional densities.

References

Cain, L. E., Robins, J. M., Lanoy, E., Logan, R., Costagliola, D. and Hernán, M. A. (2010). When to start treatment? a systematic approach to the comparison of dynamic regimes using observational data. *The International Journal of Biostatistics* **6**, 1–24.

Cain, L. E., Saag, M. S., Petersen, M., May, M. T., Ingle, S. M., Logan, R., Robins, J. M., Abgrall, S., Shepherd, B. E., Deeks, S. G. *and others*. (2016). Using observational data to emulate a randomized trial of dynamic treatment-switching strategies: an application to antiretroviral therapy. *International Journal of Epidemiology* **45**, 2038–2049.

Caniglia, E. C., Robins, J. M., Cain, L. E., Sabin, C., Logan, R., Abgrall, S., Mugavero, M. J., Hernández-Díaz, S, Meyer, L., Seng, R. *and others*. (2019). Emulating a trial of joint dynamic strategies: An application to monitoring and treatment of hiv-positive individuals. *Statistics in Medicine* **38**, 2428–2446.

H2O. (2017). Generalized linear model (glm). `https://h2o-release.s3.amazonaws.com/h2o/rel-ueno/2/docs-website/h2o-docs/data-science/glm.html`.

Lendle, S. D., Schwab, J., Petersen, M. L. and van der Laan, M. J. (2017). ltmle: An R package implementing targeted minimum loss-based estimation for longitudinal data. *Journal of Statistical Software* **81**, 1–21.

Liu, J. S. (2008). *Monte Carlo strategies in scientific computing*. Springer Science & Business Media.

Neugebauer, R., Schmittdiel, J. A. and van der Laan, M. J. (2016). A case study of the impact of data-adaptive versus model-based estimation of the propensity scores on causal inferences from three inverse probability weighting estimators. *The International Journal of Biostatistics* **12**, 131–155.

Robins, J. M. (1986). A new approach to causal inference in mortality studies with a sustained exposure period: application to control of the healthy worker survivor effect. *Mathematical Modelling* **7**, 1393–1512.

Schafer, J.L. (1997). *Analysis of Incomplete Multivariate Data*, 1 edition. London: Chapman & Hall.

Tsiatis, A.A. (2006). *Semiparametric Theory and Missing Data*, 1 edition. New York, NY: Springer.

Young, J. G., Cain, L. E., Robins, J. M., O'Reilly, E. J. and Hernán, M. A. (2011). Comparative effectiveness of dynamic treatment regimes: an application of the parametric g-formula. *Statistics in Biosciences* **3**, 119–143.

Young, J. G., Logan, R. W., Robins, J. M. and Hernán, M. A. (2019). Inverse probability weighted estimation of risk under representative interventions in observational studies. *Journal of the American Statistical Association* **114**, 938–947.

Table 1: Simulation study for deterministic dynamic treatment strategies. NICE, Strat. ICE_{HE} (hazard-extended stratified iterative conditional expectation), Pooled ICE_{HE} (pooled hazard-extended iterative conditional expectation), and Pooled ICE-classic (classical pooled iterative conditional expectation). True counterfactual risk of death at each time point under the static treatment strategy is (0.0327,0.0644,0.0950,0.1247,0.1535). True counterfactual risk of death at each time point under the dynamic treatment strategy is (0.0834,0.1547,0.2163,0.2702,0.3179). **Bias and SE are multiplied by 100.**

Always treat												
j	NICE	Pooled ICE_{HE}	Pooled ICE (classic)	Strat. ICE_{HE}	IPW_{HT}	IPW_{Haz}	NICE	Pooled ICE_{HE}	Pooled ICE (classic)	Strat. ICE_{HE}	IPW_{HT}	IPW_{Haz}
	Bias						SE					
1	0.0034	0.0031	0.0031	-0.0235	-0.0024	-0.0256	0.3519	0.3512	0.3512	1.0393	1.0592	1.0395
2	-0.0040	-0.0037	-0.0297	-0.0376	0.0177	-0.0349	0.6790	0.6782	1.1583	1.4336	1.5290	1.4490
3	-0.0047	-0.0026	-0.0365	-0.0081	0.0753	-0.0038	0.9828	0.9821	1.4334	1.7114	1.8781	1.7256
4	-0.0099	-0.0053	-0.0178	-0.0233	0.0822	-0.0202	1.2648	1.2640	1.6294	1.9730	2.2265	2.0178
5	-0.0214	-0.0145	-0.0152	-0.0660	0.0577	-0.0672	1.5263	1.5252	1.7452	2.1888	2.5260	2.2588
Treat if AIDS												
j	NICE	Pooled ICE_{HE}	Pooled ICE (classic)	Strat. ICE_{HE}	IPW_{HT}	IPW_{Haz}	NICE	Pooled ICE_{HE}	Pooled ICE (classic)	Strat. ICE_{HE}	IPW_{HT}	IPW_{Haz}
	Bias						SE					
1	0.0186	0.0213	0.0213	0.0189	0.0173	0.0091	0.5540	0.5544	0.5544	1.1899	1.1934	1.2026
2	0.0307	0.0327	0.0299	0.0206	0.0402	0.0051	0.9549	0.9576	1.3802	1.6165	1.7229	1.6241
3	0.0398	0.0439	0.0330	-0.0028	0.0754	-0.0241	1.2551	1.2598	1.6939	1.9129	2.0539	1.9520
4	0.0347	0.0427	0.0160	-0.0055	0.0734	-0.0292	1.4909	1.4967	1.8998	2.1970	2.4218	2.2447
5	0.0232	0.0334	0.0136	-0.0577	0.0656	-0.0759	1.6856	1.6871	2.0594	2.3825	2.6950	2.4614

Table 2: Simulation study results for random treatment strategy with a grace period of 1 years. NICE, Strat. ICE_{HE} (stratified hazard-extended iterative conditional expectation), Pooled ICE_{HE} (pooled hazard-extended iterative conditional expectation), and Pooled ICE-classic (classical pooled iterative conditional expectation). True risk of death at each time point is (0.1523,0.2275,0.2915,0.3462,0.3936). **Bias and SE are multiplied by 100.**

j	NICE	Pooled ICE_{HE}	Pooled ICE (classic)	Strat. ICE_{HE}	IPW_{HT}	IPW_{Haz}	NICE	Pooled ICE_{HE}	Pooled ICE (classic)	Strat. ICE_{HE}	IPW_{HT}	IPW_{Haz}
	Bias						SE					
1	0.0371	0.0122	0.0122	-0.0044	0.0001	0.0001	0.8437	0.8347	0.8347	1.2662	1.2785	1.2688
2	0.0542	0.0334	0.0169	-0.0076	-0.0045	-0.0013	1.1436	1.1327	1.4452	1.5679	1.6192	1.5702
3	0.0446	0.0215	-0.0066	-0.0136	-0.0074	-0.0034	1.3752	1.3700	1.6554	1.8447	1.9451	1.8545
4	0.0422	0.0267	-0.0088	-0.0015	-0.0006	0.0065	1.5473	1.5519	1.8217	2.0279	2.1559	2.0388
5	0.0301	0.0204	-0.0033	-0.0123	-0.0071	0.0056	1.6808	1.6897	1.9520	2.1639	2.3351	2.1872

Table 3: Extra simulation study for deterministic dynamic treatment strategies. NICE, Strat. ICE_{HE} (stratified hazard-extended iterative conditional expectation), Pooled ICE_{HE} (pooled hazard-extended iterative conditional expectation), and Pooled ICE-classic (classical pooled iterative conditional expectation). True counterfactual risk of death at each time point under the dynamic treatment strategy is (0.0811,0.1257,0.1582,0.1852,0.2092). **Bias and SE are multiplied by 100.**

j	NICE	Pooled ICE_{HE}	Pooled ICE (classic)	Strat. ICE_{HE}	Strat. ICE (classic)	IPW_{HT}	IPW_{Haz}	NICE	Pooled ICE_{HE}	Pooled ICE (classic)	Strat. ICE_{HE}	Strat. ICE (classic)	IPW_{HT}	IPW_{Haz}
	Bias							SE						
1	0.0092	0.0138	0.0138	-0.0122	-0.0122	0.0162	-0.0312	0.7594	0.7617	0.7617	1.4579	1.4579	1.6016	1.5014
2	-0.0014	0.0054	-0.0257	-0.0055	-0.0039	0.0323	-0.0428	1.0963	1.1091	1.6301	1.7611	1.7617	2.0354	1.8406
3	-0.0122	-0.0061	-0.0249	0.0032	-0.0003	0.0562	-0.0388	1.3309	1.3503	1.7955	1.9374	1.9399	2.3371	2.0871
4	-0.0204	-0.0211	-0.0226	-0.0093	-0.0164	0.0523	-0.0611	1.5243	1.5465	1.9321	2.0524	2.0533	2.5591	2.2399
5	-0.0290	-0.0397	-0.0662	-0.0109	-0.0271	0.0404	-0.0923	1.6944	1.7141	1.9929	2.1357	2.1375	2.7625	2.3770

Table 4: Extra simulation study for random dynamic treatment strategies. NICE, Strat. ICE_{HE} (stratified hazard-extended iterative conditional expectation), Pooled ICE_{HE} (pooled hazard-extended iterative conditional expectation), and Pooled ICE-classic (classical pooled iterative conditional expectation). True risk of death at each time point under the dynamic treatment strategy is (0.2323,0.2782,0.3075,0.3304,0.3501). **Bias and SE are multiplied by 100.**

j	NICE	Pooled ICE_{HE}	Pooled ICE (classic)	IPW_{HT}	IPW_{Haz}	NICE	Pooled ICE_{HE}	Pooled ICE (classic)	IPW_{HT}	IPW_{Haz}
	Bias					SE				
1	-0.0443	-0.0071	-0.0071	-0.0213	-0.0155	1.1776	1.1048	1.1048	1.3799	1.3787
2	-0.0479	-0.0074	-0.0200	-0.0418	-0.0337	1.2932	1.2382	1.4580	1.6018	1.5930
3	-0.0519	-0.0068	-0.0301	-0.0504	-0.0404	1.3823	1.3468	1.5888	1.7651	1.7566
4	-0.0546	-0.0083	-0.0145	-0.0979	-0.0881	1.4647	1.4423	1.6871	1.8622	1.8550
5	-0.0578	-0.0123	-0.0322	-0.1010	-0.0919	1.5451	1.5299	1.7356	1.9181	1.9040