ONLINE APPENDIX

**Applications to Causal Inquiries Contrasting Specific Regimes**

Here we describe how to estimate metrics among those following regimes of interest $\bar{a}^\dagger$ as they evolve. Imagine a trial that only randomized exposure $A(t)$ for persons who were previously randomized to stay on a regime of interest. We would only expect and care about exchangeability $Y_{\bar{a}^\dagger} \coprod A(t) | \bar{A}(t-1)^\dagger$ and statistical exogeneity $\bar{C}(t) \coprod A(t) | \bar{A}(t-1)^\dagger$ among exposure histories $\bar{A}(t-1)^\dagger$ compatible with regimes of interest.

To apply Diagnostics 1 and 3 to regimes as they evolve, one must censor persons when their exposure *history* at time $t$ i.e. $\bar{A}(t-1)$ becomes incompatible with all regimes of interest. Similarly, for Diagnostic 2, persons would be censored when their exposure *history* at time $t-k$ is incompatible, regardless of the exposure trajectory after time $t-k$. Note that so far we have considered so-called static exposure regimes defined only by exposure history. This censoring approach could be extended to consider so-called dynamic regimes where exposures depend on covariate history.[1,2] Though, the exchangeability conditions for dynamic exposure regimes differ from those used here.

This implementation aligns closely with a strategy to estimate causal effects of exposure regimes via censoring rules to reduce modeling assumptions.[2] When models are used to "borrow" information from other regimes—sometimes of no scientific interest—and thereby improve the efficiency of effect estimates, those "borrowed" regimes should also be diagnosed (e.g. diagnosing all observed regimes as they evolve, as in the main text simulated example).

**Bias Metrics for Measured Confounding**

The covariate balance metrics in the main text ignore data on the covariate-outcome relationship. Suppose that $C(t)$ is necessary to attain exchangeability $Y_{\bar{a}} \amalg A(t) | \bar{A}(t-1), \bar{C}(t)$. Following the notation in Table 2 of the main text, from VanderWeele and Arah 2010[3] it follows that a crude estimate for the additive effect of exposure $A(\tau)$ on outcome $Y$ among those with exposure history $H(\tau)$ would, upon adjustment for a binary covariate $C(\tau_c)$, be reduced by the amount:

$$(E[C(\tau_c)|A(\tau) = a', H(\tau)] - E[C(\tau_c)|A(\tau) = a'', H(\tau)])$$
$$\times \; (E[Y|C(\tau_c) = 1, A(\tau) = a'', H(\tau)] - E[Y|C(\tau_c) = 0, A(\tau) = a'', H(\tau)])$$

Thus, for binary variables and dummy indicators, we can incorporate empirical data on the covariate-outcome relationship by multiplying the balance of covariates at time $\tau_c$ across exposure at time $\tau$, among a particular level of exposure history at time $\tau$,

$$E[W(\tau) \times I(A(\tau) = a') \times C(\tau_c)|H(\tau)] - E[W(\tau) \times I(A(\tau) = a'') \times C(\tau_c)|H(\tau)],$$

by the difference in mean outcome across covariate levels at time $\tau_c$ (among the referent group at time $\tau$ with the same level of exposure history):

$$E[W(\tau) \times I(C(\tau_c) = 1) \times Y|A(\tau) = a'', H(\tau)] - E[W(\tau) \times I(C(\tau_c) = 0) \times Y|A(\tau) = a'', H(\tau)]$$

where $W(\tau)$ represents equals one for Diagnostic 1 or a weight for Diagnostic 3. The bias metric will equal zero when, conditional on exposure history, the covariate $C(\tau_c)$ is not associated with the exposure $A(\tau)$ or the outcome $Y$ in the weighted population. The expression given here is based on a simplifying assumption that $A(\tau)$ does not statistically interact with $C(\tau_c)$ for $Y$ (on the additive scale) but more general bias expressions can be derived.[3,4] Future work might examine these generalizations and accommodations for censoring in greater depth.

**Extensions to Multivariate Time-Varying Exposures**

To diagnose confounding in studies of multivariate exposures (e.g. joint effects, direct effects and interaction) one can assess each exposure while conditioning on their joint exposure history. We develop this result for the case of two distinct time-varying exposures where (i) at each time $t$, exposure $A(t)$ affects covariates $C(t)$ which affects exposure $Z(t)$, and (ii) covariates $C(t)$ affect subsequent exposures $A(t + k)$ and $Z(t + k)$ and also the outcome $Y$.

*Diagnostic 1: time-varying confounding in the study population*

In trials that randomly assign two distinct exposures $A(t)$ and $Z(t)$ within levels of their joint exposure history, we expect exchangeability for each exposure $Y_{\overline{az}} \coprod A(t) | \bar{A}(t-1), \bar{Z}(t-1)$ and $Y_{\overline{az}} \coprod Z(t) | \bar{Z}(t-1), \bar{A}(t)$ and, as a result, statistical exogeneity for each exposure with respect to their joint history i.e. $\bar{C}(t) \coprod A(t) | \bar{A}(t-1), \bar{Z}(t-1)$ and $\bar{C}(t) \coprod Z(t) | \bar{A}(t), \bar{Z}(t-1)$.[5,6] In observational studies where there is confounding, causal inference regarding $\overline{az}$ requires exchangeability for both exposures within levels of their joint exposure history plus some set of covariate history e.g. $Y_{\overline{az}} \coprod A(t) | \bar{A}(t-1), \bar{Z}(t-1), \bar{C}(t)$ and $Y_{\overline{az}} \coprod Z(t) | \bar{Z}(t-1), \bar{A}(t), \bar{C}(t)$.[34] Confounding is reflected by departures from joint statistical exogeneity. In **eAppendix Table 1** we propose analogues of Diagnostic 1 to describe confounding for distinct exposures $A(t)$ and $Z(t)$. Note that when $A$ and $Z$ are point exposures and $C$ and $L$ are covariates (with ordering $C, A, L, Z$), this version of Diagnostic 1 assesses the balance of covariates $C$ across $A$, and the balance of covariates $C, L$ across $Z$ within strata of $A$. Applied studies have used this point exposure metric to describe measured mediator-outcome confounding in mediation analyses.[7,8]

*Diagnostic 2: exposure-covariate feedback in the study population*

Assuming $Y_{\overline{az}} \perp\!\!\!\perp A(t) \mid \bar{A}(t-1), \bar{Z}(t-1), \bar{C}(t)$ and $Y_{\overline{az}} \perp\!\!\!\perp Z(t) \mid \bar{Z}(t-1), \bar{A}(t), \bar{C}(t)$, any covariate $C(t)$ associated with a prior exposure in $\bar{A}(t)$ or $\bar{Z}(t-1)$ given its past contributes to exposure-covariate feedback for the average joint effect of $\bar{A}$ and $\bar{Z}$ on $Y$. In **eAppendix Table 1**, we present analogues of Diagnostics 2a and 2b for exposure-covariate feedback in censored data. The models for exposures $A(t-k)$ and $Z(t-k)$ and also censoring $S(t-k)$ used to develop the weights (or propensity score strata) must condition on joint exposure history; the model for exposure $A(t-k)$ should also include covariates that block backdoor paths between $A(t-k)$ and $Y$; the model for exposure $Z(t-k)$ should also include covariates that block backdoor paths between $Z(t-k)$ and $Y$. When the models and weights (or strata) are correctly specified, the diagnostic will produce two populations: one statistically exogenous for $A(t-k)$, and the other statistically exogenous for $Z(t-k)$, isolating the associations between covariates and prior exposures that motivate g-methods: ones that are causal or arise through unmeasured causes. Note that when $A$ and $Z$ are point exposures and $C$ and $L$ are covariates (with ordering $C, A, L, Z$), this version of Diagnostic 2 would check for balance of $L$ across $A$ after weighting the population by the inverse probability of exposure $A$ or within levels of propensity score strata for $A$.


*Diagnostic 3: residual time-varying confounding in the weighted population*

Assuming $Y_{\overline{az}} \perp\!\!\!\perp A(t) \mid \bar{A}(t-1), \bar{Z}(t-1), \bar{C}(t)$ and $Y_{\overline{az}} \perp\!\!\!\perp Z(t) \mid \bar{Z}(t-1), \bar{A}(t), \bar{C}(t)$, inverse probability weights can be used to adjust for confounding of causal contrasts between regimes involving two distinct exposures $A(t)$ and $Z(t)$.[6] The weight at each time $t$ is the product of the cumulative inverse probability weights for $A(t)$ and $Z(t)$. The weights aim to create a pseudo-population where, at each time $t$, $A(t)$ and $Z(t)$ are both statistically exogenous with respect to

prior covariates within levels of joint exposure history.[9] But the weighted population will suffer residual imbalance if, for either exposure, weights or models are misspecified or positivity is violated. In **eAppendix Table 1**, we propose analogues for Diagnostic 3 to assess departures from joint statistical exogeneity for $A(t)$ and $Z(t)$ in the weighted population. Note that when $A$ and $Z$ are point exposures and $C$ and $L$ are covariates (with ordering $C, A, L, Z$), this version of Diagnostic 3 assesses the weighted balance of covariates $C$ across $A$, and the weighted balance of covariates $C, L$ across $Z$ within strata of $A$.

**Extension to a Special Case of the Parametric G-formula**

Diagnostic 3 can be reframed to examine residual confounding for a special case of the parametric g-formula, where models for longitudinal propensity score histories $\bar{e}_{a(t)}$ replace those for covariate histories $\bar{C}(t)$. Achy-Brou et al[10] proved that if $Y_{\bar{a}} \coprod A(t) | \bar{A}(t-1), \bar{C}(t)$ holds then $Y_{\bar{a}} \coprod A(t) | \bar{A}(t-1), \bar{e}_{a(t)}$ also holds, implying statistical exogeneity within joint levels of exposure and propensity score history i.e. $\bar{C}(t) \coprod A(t) | \bar{A}(t-1), \bar{e}_{a(t)}$. The curse of dimensionality might prohibit checking such statistical exogeneity directly, but a model might provide progress if $\bar{e}_{a(t)}$ are coarsened into strata. At a minimum, one can check the unweighted balance of all prior covariates across $A(t)$ within levels of exposure history $\bar{A}(t-1)$ and time-specific propensity score $e_{a(t)}$ strata, as proposed by Shinohara,[11] since this is a prerequisite.

| eAppendix Table 1. Time-specific balance metrics for confounding of multivariate time-varying exposures $A(t)$ and $Z(t)$ by time-varying covariates $C(t)$ among the uncensored $S(t) = 0$ | |
|---|---|
| **Diagnostic 1: Time-varying confounding i.e. $C(t-k)$ across levels of $A(t)$ and $Z(t)$** <br> (for all $t \in \{0, ..., t\}$ and chosen $k \in \{0, ..., t\}$ where $1 \le k \le t'$ for $A$ and $0 \le k \le t'$ for $Z$) | |
| Balance metrics | Definitions |
| $E[C(t-k)\|A(t) = a', \bar{A}(t-1), \bar{Z}(t-1), \bar{S}(t) = 0]$ <br> $- E[C(t-k)\|A(t) = a'', \bar{A}(t-1), \bar{Z}(t-1), \bar{S}(t) = 0]$ | N/A |
| $E[C(t-k)\|Z(t) = z', \bar{Z}(t-1), \bar{A}(t), \bar{S}(t) = 0]$ <br> $- E[C(t-k)\|Z(t) = z'', \bar{Z}(t-1), \bar{A}(t), \bar{S}(t) = 0]$ | N/A |
| **Diagnostic 2: Exposure-covariate feedback i.e. $C(t)$ across levels of $A(t-k)$ and $Z(t-k)$** <br> (for all $t \in \{0, ..., t\}$ and all $k \in \{0, ..., t\}$ where $0 \le k \le t'$ for $A$ and $1 \le k \le t'$ for $Z$) | |
| Balance metrics | |
| **a. by inverse probability weighting** | Weight $W_{as(t-k)} = W_{a(t-k)} \times W_{s(t)}$ and $W_{zs(t-k)} = W_{z(t-k)} \times W_{s(t)}$ |
| $E\left[W_{a's(t-k)} \times I(A(t-k) = a') \times C(t)\|\bar{A}(t-k-1), \bar{Z}(t-k-1), \bar{S}(t) = 0\right]$ <br> $- E\left[W_{a''s(t-k)} \times I(A(t-k) = a'') \times C(t)\|\bar{A}(t-k-1), \bar{Z}(t-k-1), \bar{S}(t) = 0\right]$ | $W_{a(t-k)} = \dfrac{P[A(t-k) = a\|\bar{A}(t-k-1), \bar{Z}(t-k-1), \bar{S}(t-k) = 0]}{P[A(t-k) = a\|\bar{A}(t-k-1), \bar{Z}(t-k-1), \bar{C}(t-k-1), \bar{S}(t-k) = 0]}$ <br><br> $W_{z(t-k)} = \dfrac{P[Z(t-k) = z\|\bar{Z}(t-k-1), \bar{A}(t-k), \bar{S}(t-k) = 0]}{P[Z(t-k) = z\|\bar{Z}(t-k-1), \bar{A}(t-k), \bar{C}(t-k), \bar{S}(t-k) = 0]}$ |
| $E\left[W_{z's(t-k)} \times I(Z(t-k) = z') \times C(t)\|\bar{Z}(t-k-1), \bar{A}(t-k), \bar{S}(t) = 0\right]$ <br> $- E\left[W_{z''s(t-k)} \times I(Z(t-k) = z'') \times C(t)\|\bar{Z}(t-k-1), \bar{A}(t-k), \bar{S}(t) = 0\right]$ | $W_{s(t)}$ <br> $= \prod_{k=0}^{k=t} \dfrac{P[S(t-k) = 0\|\bar{S}(t-k-1) = 0]}{P[S(t-k) = 0\|\bar{S}(t-k-1) = 0, \bar{Z}(t-k-1), \bar{A}(t-k-1), \bar{C}(t-k-1)]}$ |
| **b. by propensity score stratification**[b] | Propensity scores $e_{a(t-k)}$ and $e_{z(t-k)}$ and Weight $W_{s(t)}$ |
| $E\left[W_{a's(t-k)} \times I(A(t-k) = a') \times C(t)\|e_{a'(t-k)}, \bar{S}(t) = 0\right]$ <br> $- E\left[W_{a''s(t-k)} \times I(A(t-k) = a'') \times C(t)\|e_{a'(t-k)}, \bar{S}(t) = 0\right]$ | $e_{a'(t-k)}$ <br> $= P[A(t-k) = a'\|\bar{A}(t-k-1), \bar{Z}(t-k-1), \bar{C}(t-k-1), \bar{S}(t-k) = 0]$ <br><br> $e_{z'(t-k)}$ <br> $= P[Z(t-k) = z'\|\bar{Z}(t-k-1), \bar{A}(t-k), \bar{C}(t-k), \bar{S}(t-k) = 0]$ |
| $E\left[W_{z's(t-k)} \times I(Z(t-k) = z') \times C(t)\|e_{z'(t-k)}, \bar{S}(t) = 0\right]$ <br> $- E\left[W_{z''s(t-k)} \times I(Z(t-k) = z'') \times C(t)\|e_{z'(t-k)}, \bar{S}(t) = 0\right]$ | $W_{s(t)}$ <br> $= \prod_{k=0}^{k=t} \dfrac{P[S(t-k) = 0\|\bar{S}(t-k-1) = 0]}{P[S(t-k) = 0\|\bar{S}(t-k-1) = 0, \bar{Z}(t-k-1), \bar{A}(t-k-1), \bar{C}(t-k-1)]}$ |
| **Diagnostic 3: Residual time-varying confounding i.e. $C(t-k)$ across levels of $A(t)$ and $Z(t)$** <br> (for all $t \in \{0, ..., t\}$ and chosen $k \in \{0, ..., t\}$ where $1 \le k \le t'$ for $A$ and $0 \le k \le t'$ for $Z$) | |
| Balance metrics | Weight $W_{az(t)} = W_{a(t)} \times W_{z(t)}$ |
| $E\left[W_{a'z(t)} \times I(A(t) = a') \times C(t-k)\|\bar{A}(t-1), \bar{Z}(t-1), \bar{S}(t) = 0\right]$ <br> $- E\left[W_{a''z(t)} \times I(A(t) = a'') \times C(t-k)\|\bar{A}(t-1), \bar{Z}(t-1), \bar{S}(t) = 0\right]$ | $W_{a(t)}$ <br> $= \prod_{k=0}^{k=t} \dfrac{P[A(t-k) = a\|\bar{A}(t-k-1), \bar{Z}(t-k-1), \bar{S}(t-k) = 0]}{P[A(t-k) = a\|\bar{A}(t-k-1), \bar{Z}(t-k-1), \bar{C}(t-k-1), \bar{S}(t-k) = 0]}$ |
| $E\left[W_{az's(t)} \times I(Z(t) = z') \times C(t-k)\|\bar{Z}(t-1), \bar{A}(t), \bar{S}(t) = 0\right]$ <br> $- E\left[W_{az''s(t)} \times I(Z(t) = z'') \times C(t-k)\|\bar{Z}(t-1), \bar{A}(t), \bar{S}(t) = 0\right]$ | $W_{z(t)}$ <br> $= \prod_{k=0}^{k=t} \dfrac{P[Z(t-k) = z\|\bar{Z}(t-k-1), \bar{A}(t-k), \bar{S}(t-k) = 0]}{P[Z(t-k) = z\|\bar{Z}(t-k-1), \bar{A}(t-k), \bar{C}(t-k), \bar{S}(t-k) = 0]}$ |
| [a]These balance metrics are on the mean difference scale. They can be reported on the standardized mean difference scale by dividing by the (unweighted) pooled standard deviation conditional on joint exposure history (for Diagnostics 1, 2a and 3) or propensity-score strata (for Diagnostic 2b). Doing so places metrics for binary and continuous covariates on the same scale. <br> [b]For categorical exposures one needs to jointly condition on the predicted probabilities for each non-referent exposure level | |

**Details of Example Simulations**

We simulated a baseline-randomized trial (n=100,000) of a time-varying exposure $A(t)$ subject to confounding over follow-up by a vector of covariates $C(t) = (L(t), M(t), N(t), O(t), P(t))$, and also a censoring indicator $S(t)$ with no censoring at baseline (all realized at times $t = (0,1,2)$ in the order $S(t), C(t), A(t)$ at each time $t$). Each covariate in $C(t)$ was specified as a function of its prior realizations and an unmeasured variable $U$, and also exposure $A(t)$ in the case of covariates $L(t)$ and $M(t)$. For times $t > 0$ exposure $A(t)$ depended on entire exposure and covariate histories $\bar{A}(t-1)$ and $\bar{C}(t)$ and included the interactions $M(t) \times N(t)$ and $O(t) \times P(t)$. For time $t = 0$ there was no censoring but for times $t > 0$ censoring $S(t)$ depended on full exposure and covariate histories $\bar{A}(t-1)$ and $\bar{C}(t-1)$. We did not simulate an outcome $Y$ because each diagnostic in the main text ignores data on the outcome. If we had, though, each variable in $\bar{C}(t)$ would be a common cause of exposures $A(t)$ and the outcome $Y$; and each variable in $\bar{A}(t-1)$ and $\bar{C}(t-1)$ would be a common cause of censoring indicators $S(t)$ and the outcome $Y$.

Realizations for all variables $V(t)$ were simulated as random bernoulli draws with probability $p_v(t)$. For $A(0)$, $S(0)$, and also $U$, the probability was specified directly. For all other variables we used linear combinations on the log odds scale to calculate $p_v(t)$. This setup allowed for encoding of time-varying confounding through immediate and distant covariates, temporally indexed by $t - k$. The logits for each $p_v(t)$ are listed below:

For U: ~bernouli(0.4)

$$\text{logit } p_l(t) = \beta_{0l(t)} + \beta_u^l u + \sum_{k=1}^{k=t} L(t-k)\beta_l^l(t-k) + \sum_{k=1}^{k=t} A(t-k)\beta_a^l(t-k)$$

For L(0): $\beta_0$=-1.20; $\beta_U$=0.88;
For L(1): $\beta_0$=-1.20; $\beta_U$=0.88; $\beta_{L0}$=0.56 ; $\beta_{A0}$=1.10

For L(2): $\beta_0$=-1.20; $\beta_U$=0.88; $\beta_{L0}$=0.14 ; $\beta_{A0}$=.41 ; $\beta_{L1}$=.56 ; $\beta_{A1}$= 1.10

$$\text{logit } p_m(t) = \beta_{0m(t)} + \beta_u^m u + \sum_{k=1}^{k=t} M(t-k)\beta_m^m(t-k) + \sum_{k=1}^{k=t} A(t-k)\beta_a^m(t-k)$$

For M(0): $\beta_0$=-.29; $\beta_U$=-1.61;

For M(1): $\beta_0$=-.29; $\beta_U$=-1.61; $\beta_{M0}$=.69; $\beta_{A0}$=-.43

For M(2): $\beta_0$=-.29; $\beta_U$=-1.61; $\beta_{M0}$=.59; $\beta_{A0}$=-.29 ; $\beta_{M1}$=.69; $\beta_{A1}$=-.43


$$\text{logit } p_n(t) = \beta_{0n(t)} + \beta_u^n u + \sum_{k=1}^{k=t} N(t-k)\beta_n^n(t-k)$$

For N(0): $\beta_0$=-.69; $\beta_U$=0.69;

For N(1): $\beta_0$=-.69; $\beta_U$=0.69; $\beta_{N0}$=.26;

For N(2): $\beta_0$=-.69; $\beta_U$=0.69; $\beta_{N0}$=.18; $\beta_{N1}$=.26;


$$\text{logit } p_o(t) = \beta_{0o(t)} + \beta_u^o u + \sum_{k=1}^{k=t} O(t-k)\beta_o^o(t-k)$$

For O(0): $\beta_0$=-.51; $\beta_U$=1.39;

For O(1): $\beta_0$=-.51; $\beta_U$=1.39; $\beta_{O0}$=.10 ;

For O(2): $\beta_0$=-.51; $\beta_U$=1.39; $\beta_{O0}$=.05; $\beta_{O1}$=.10;


$$\text{logit } p_p(t) = \beta_{0p(t)} + \beta_u^p u + \sum_{k=1}^{k=t} P(t-k)\beta_p^p(t-k)$$

For P(0): $\beta_0$=-.92; $\beta_U$=-1.61;

For P(1): $\beta_0$=-.92; $\beta_U$=-1.61; $\beta_{P0}$=.18;

For P(2): $\beta_0$=-.92; $\beta_U$=-1.61; $\beta_{P0}$=.10; $\beta_{P1}$=.18;


For A(0): ~bernouli(0.4)

$$\text{logit } p_a(t) = \beta_{0a(t)} + \sum_{k=1}^{k=t} A(t-k)\beta_a^a(t-k) + \sum_{k=0}^{k=t} L(t-k)\beta_l^a(t-k) + \sum_{k=0}^{k=t} M(t-k)\beta_m^a(t-k) + \sum_{k=0}^{k=t} N(t-k)\beta_n^a(t-k) + \sum_{k=0}^{k=t} O(t-k)\beta_o^a(t-k) + \sum_{k=0}^{k=t} P(t-k)\beta_p^a(t-k) + \beta_{m\times n}^a(t) \times (M(t) \times N(t)) + \beta_{o\times p}^a(t) \times (O(t) \times P(t))$$

For A(1): $\beta_0$=-.92; $\beta_{A0}$=.26; $\beta_{L0}$= .14; $\beta_{M0}$= -.36; $\beta_{N0}$=.18; $\beta_{O0}$=.10; $\beta_{P0}$=.18;

$\beta_{L1}$= .83; $\beta_{M1}$= -.92; $\beta_{N1}$=.26; $\beta_{O1}$=.18; $\beta_{P1}$=.10; $\beta_{M1*N1}$=-.24; $\beta_{O1*P1}$=.02


For A(2): $\beta_0$=-1.20; $\beta_{A0}$=.34; $\beta_{L0}$= .05; $\beta_{M0}$= -.22; $\beta_{N0}$=.10; $\beta_{O0}$=.05; $\beta_{P0}$=.10;

$\beta_{A1}$=.59; $\beta_{L1}$= .14; $\beta_{M1}$= -.36; $\beta_{N1}$=.18; $\beta_{O1}$=.10; $\beta_{P1}$=.05;

$\beta_{L2}$= .83; $\beta_{M2}$= -.92; $\beta_{N2}$=.26; $\beta_{O2}$=.18; $\beta_{P2}$=.18; $\beta_{M1*N1}$=-.24; $\beta_{O1*P1}$=.03

$$\text{logit } p_s(t) = \beta_{0s(t)} + \sum_{k=1}^{k=t} A(t-k)\beta_a^s(t-k) + \sum_{k=1}^{k=t} L(t-k)\beta_l^s(t-k) + \sum_{k=1}^{k=t} M(t-k)\beta_m^s(t-k) + \sum_{k=1}^{k=t} N(t-k)\beta_n^s(t-k) + \sum_{k=1}^{k=t} O(t-k)\beta_o^s(t-k) + \sum_{k=1}^{k=t} P(t-k)\beta_p^s(t-k)$$

For S(1): $\beta_0$=-1.90; $\beta_{A0}$=1.69; $\beta_{L0}$=.10; $\beta_{M0}$=.10; $\beta_{N0}$=.10; $\beta_{O0}$=-1.61; $\beta_{P0}$=-1.61;
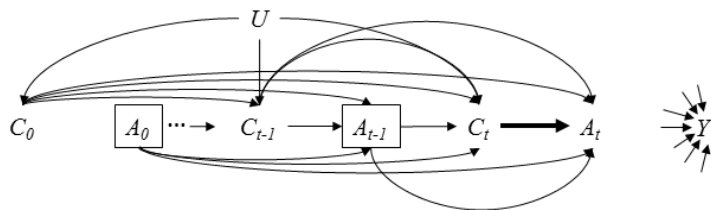

For S(2): $\beta_0$=-1.90; $\beta_{A0}$=1.39; $\beta_{L0}$=.10; $\beta_{M0}$=.10; $\beta_{N0}$=.10; $\beta_{O0}$=-1.20; $\beta_{P0}$=-1.20;

$\beta_{A1}$=1.61; $\beta_{L1}$=.10; $\beta_{M1}$=.10; $\beta_{N1}$=.10; $\beta_{O1}$=-1.61; $\beta_{P1}$=-1.61;
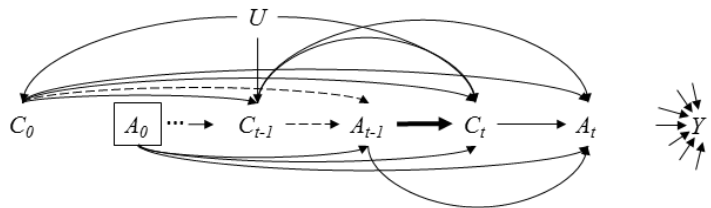
To obtain censored data used in **Figure 6**, **eFigure 4**, and **eFigure 5** we repeated the simulation but censored records once $S(t) = 1$ was realized, and to obtain data with positivity violations used in Figure 5c, 5d, and 5e we set $p_a(t) = 1/10{,}000{,}000$ whenever $L(t) = 0$ and $O(t) = 1$.
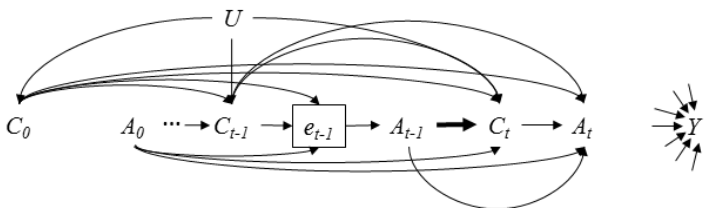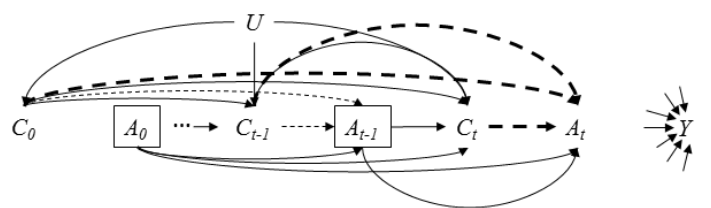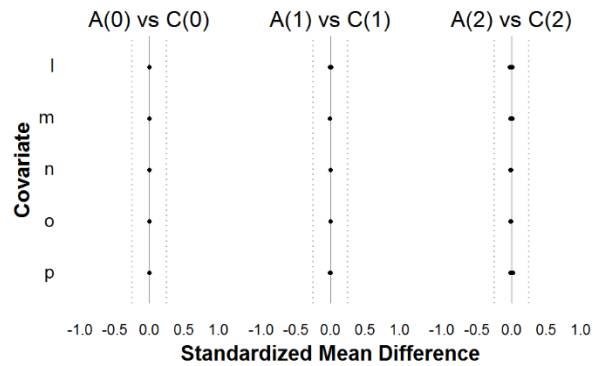
**eFigures**

(a)



(b)



(c)



(d)



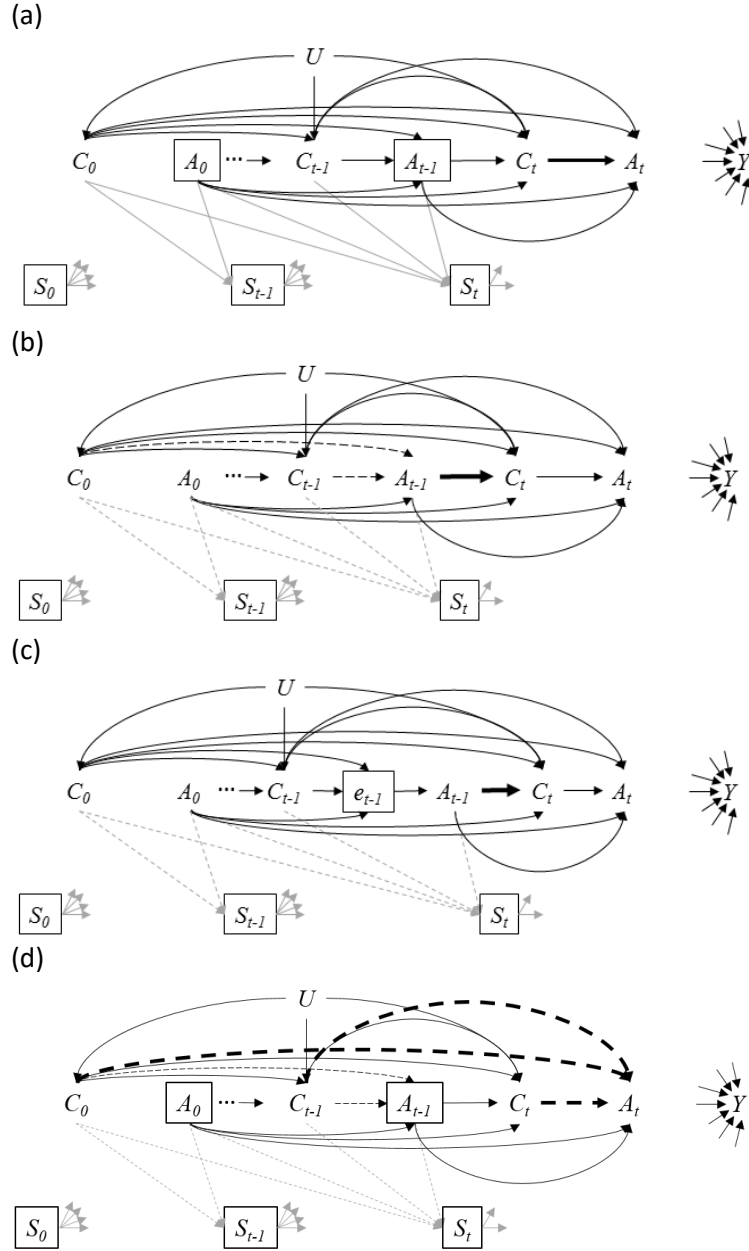**eFigure 1**. Directed Acyclic Graphs for each diagnostic in the absence of censoring. (a) Diagnostic 1 for assessing $A(t)$'s association with $C(t)$. (b) Diagnostic 2a by inverse probability weighting to assess the association between $C(t)$ and $A(t-1)$ through their causal relationship and through unmeasured common causes (c) Diagnostic 2b by propensity score stratification to assess the association between $C(t)$ and $A(t-1)$ through their causal relationship and through unmeasured common causes. (d) Diagnostic 3 to assess residual associations between $C(t)$ and $A(t)$ after applying cumulative inverse probability weights for $A(t)$. Boxes represent evaluating the diagnostic within levels of those variables. Bold arrows represent associations captured by the diagnostic and (for diagnostics that apply weights) dashed arrows represent associations that the weights are meant to remove. Note that in (a) only the causal path between $C(t)$ and $A(t)$ is bolded but many other non-causal paths also contribute to Diagnostic 1.

**eFigure 2**. The diagonal of a full trellised plot for Diagnostic 3 (residual confounding) indexed by exposure measurement times. Each subplot evaluates the weighted balance of covariates $C$ ($i.e. L, M, N, O, P$) across levels of exposure $A = 1$ versus $A = 0$, both measured at time $t$. Each dot represents a pattern of exposure history through time $t - 1$. The interpretation is similar to the plot for Diagnostic 1. In our simulated example there is no measured residual confounding of $A(t)$ by proximal covariates and, although not shown, this also was true for more distant covariates. Note that here, all observed exposure regimes were examined.

(a)



(b)



(c)



(d)



**eFigure 3**. Directed Acyclic Graphs for each diagnostic in the presence of censoring. (a) Diagnostic 1 for assessing $A(t)$'s association with $C(t)$. (b) Diagnostic 2a by inverse probability weighting to assess the association between $C(t)$ and $A(t-1)$ through their causal relationship and through unmeasured common causes (c) Diagnostic 2b by propensity score stratification to assess the association between $C(t)$ and $A(t-1)$ through their causal relationship and through unmeasured common causes (d) Diagnostic 3 to assess residual associations between $C(t)$ and $A(t)$ after applying joint cumulative inverse probability weights for $A(t)$ and $S(t)$. Boxes represent evaluating the diagnostic within levels of those variables. Bold arrows represent associations captured by the diagnostic and (for diagnostics that apply weights) dashed arrows represent associations that the weights are meant to remove. Note that in (a) several other non-bolded paths contribute to Diagnostic 1 for $C(t)$ and $A(t)$.
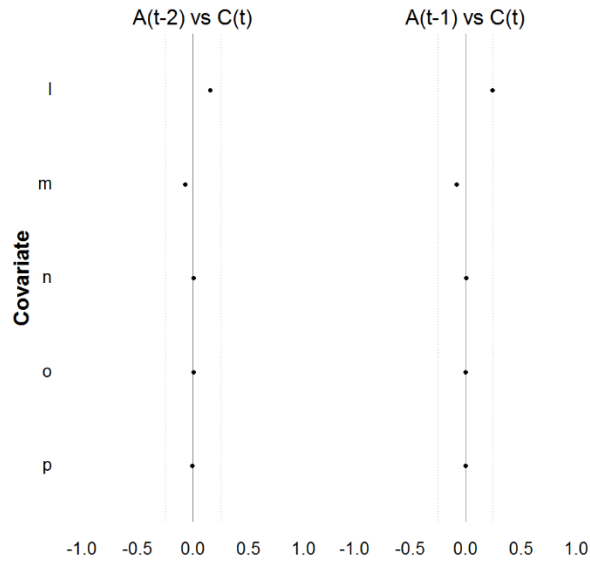
**eFigure 4**. A trellised plot for Diagnostic 2b among the uncensored, averaged over propensity score strata and time. The panels are indexed by distance between exposure and covariate times (columns). Each dot represents the average balance of covariates, measured at time $t$, across levels of prior exposure, measured at times $t - k$. The right panel reports the average over balances for $C(2)$ vs. $A(1)$ and also $C(1)$ vs. $A(0)$, while the left is for $C(2)$ vs. $A(0)$. In our example, the average balance across pools of person-time appear similar, and reflect the same patterns described in **Figure 4**. These plots suggest that, on-average, covariates $L(t)$ and $M(t)$ contribute to exposure-covariate feedback. Note that here, all observed exposure regimes among the uncensored were examined.
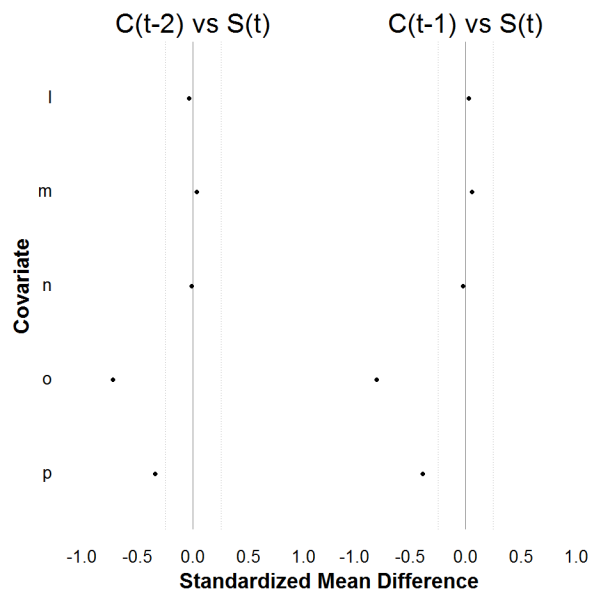
**eFigure 5**. A trellised plot for Diagnostic 1 examining statistical exogeneity for censoring $S(t)$, after averaging over exposure history and also time. Implicitly, the exposure history of interest here is among the uncensored i.e. $\bar{A}(t-1) = \bar{a}(t-1), \bar{S}(t-1) = 0$. The panels are indexed by distance between censoring and covariate measurement times (columns). The right panel reports summary metrics that assess the average balance between censoring and proximal covariates at one unit of distance (adjusting for exposure history): $S(1)$ vs. $C(0)$ and $S(2)$ vs. $C(1)$. The left panel reports the analogous balance metric at two units of distance: $S(2)$ vs. $C(0)$. Note that balance measures comparing $S(0)$ vs $C(0)$, $S(1)$ vs $C(1)$, and $S(2)$ vs $C(2)$ do not contribute to this plot because, in the data-generating model, censoring $S(t)$ precedes covariate measurement $C(t)$ at every time $t$. Note also that here, all exposure regimes observed among the uncensored were examined.
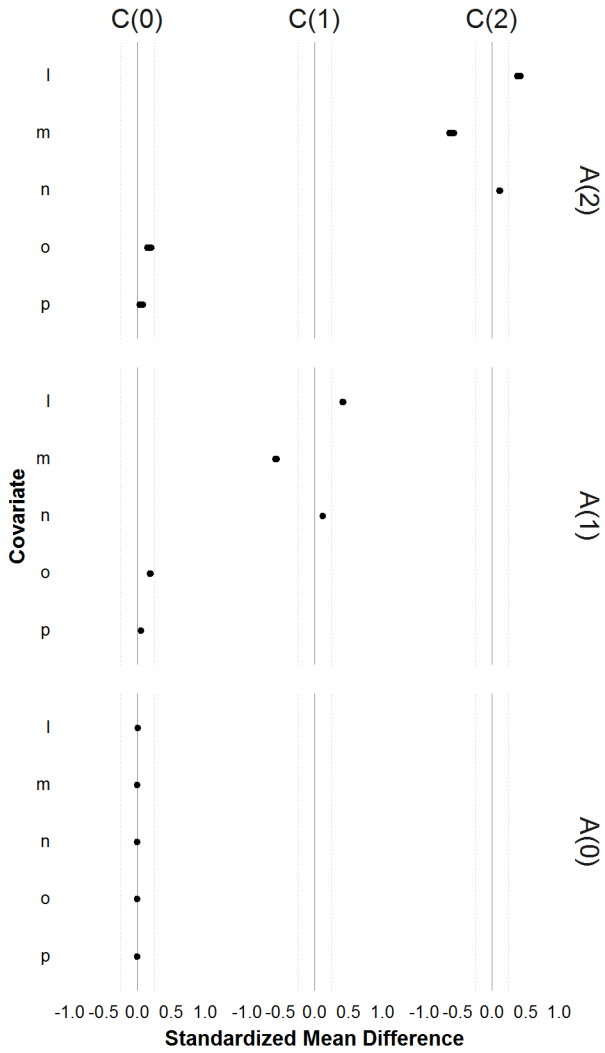
**eFigure 6**. A modified trellised plot for Diagnostic 1 under the assumption that baseline covariates $O(0)$ and $P(0)$ and the most recent values of $L(t)$, $M(t)$, and $N(t)$ are sufficient to adjust for confounding of each exposure $A(t)$. Even though this assumption is incorrect in our simulated example (it may be tenable in real settings), we apply it here to demonstrate the plots when investigators assume that a subset of the chosen covariate history is sufficient to adjust for confounding. Such assumptions are commonly invoked in analyses of time-to-event and repeated measures outcomes.

1.  Robins JM, Hernán MA. Estimation of the causal effects of time-varying exposures. In: Fitzmaurice G, Davidian M, Verbeke G, Molenberghs G, eds. *Longitudinal Data Analysis*. New York, NY: Chapman and Hall/CRC Press; 2009:553-599.

2.  Hernán MA, Lanoy E, Costagliola D, Robins JM. Comparison of dynamic treatment regimes via inverse probability weighting. *Basic & clinical pharmacology & toxicology.* Mar 2006;98(3):237-242.

3.  VanderWeele TJ, Arah OA. Bias formulas for sensitivity analysis of unmeasured confounding for general outcomes, treatments, and confounders. *Epidemiology.* 2011;22(1):42-52.

4.  VanderWeele TJ. Unmeasured confounding and hazard scales: sensitivity analysis for total, direct, and indirect effects. *European journal of epidemiology.* Feb 2013;28(2):113-117.

5.  Robins JM. Association, causation, and marginal structural models. *Synthese.* 1999;121:151-179.

6.  VanderWeele TJ. Marginal structural models for the estimation of direct and indirect effects. *Epidemiology.* 2009;20(1):6-13.

7.  Jackson JW, VanderWeele TJ, Viswanathan A, Blacker D, Schneeweiss S. The explanatory role of stroke as a mediator of the mortality risk difference between older adults who initiate first- versus second-generation antipsychotic drugs. *American Journal of Epidemiology.* 2014;180(8):547-852.

8.  Jackson JW, VanderWeele TJ, Blacker D, Schneeweiss S. Mediators of first versus second-generation antipsychotic-related mortality in older adults. *Epidemiology.* 2015;26(5):700-709.

9.  Hernán MA, Brumback B, Robins JM. Marginal structural models to estimate the joint causal effect of nonrandomized treatments. *Journal of the American Statistical Association.* 2001;96(454):440-448.

10. Achy-Brou AC, Frangakis CE, Griswold M. Estimating treatment effects of longitudinal designs using regression models on propensity scores. *Biometrics.* 2010;66(3):824-833.

11. Shinohara RT, Narayan AK, Hong K, et al. Estmiating parsimonious models of longitudinal causal effects using regressions on propensity scores. *Statistics in Medicine.* 2013;30(22):3829-3837.