Doubly robust estimation and causal inference in longitudinal studies with dropout and truncation by death: Supplementary material

MICHELLE SHARDELL*1, GREGORY E HICKS2, LUIGI FERRUCCI3
1Department of Epidemiology and Public Health, University of Maryland 660 West Redwood Street Baltimore, Maryland 21201, U.S.A.
2Department of Physical Therapy, University of Delaware 303 McKinly Lab Newark, Delaware 19716, U.S.A.
3National Institute on Aging, 3001 S Hanover Street Baltimore, Maryland 21225, U.S.A.

mshardel@epi.umaryland.edu

A. Generalized Data Structure and Assumptions

Throughout this supplement, we consider a more general data structure and set of assumptions. Namely, we include $W(j)$, a set of variables that may be affected by $\bar{A}(j)$ and may be common causes of both $Y(k)$ and $Z(k)$ for $k = j + 1, \ldots, J + 1$. We henceforth assume that data are generated in the order $L(0), A(0), W(0), R(0), Z(1), Y(1), L(1), A(1), W(1), R(1), \ldots, L(J), A(J), W(J), R(J), Z(J + 1), Y(J + 1), R(J + 1)$.

For a single follow-up visit, the inclusion of $W(0)$ leads to the scenario considered by TGSW. Under the DAG in Supplementary Figure 1A, no conditioning set is sufficient to remove bias, because conditioning on $W(0)$ would adjust out a pathway of scientific interest, namely $A(0) \rightarrow W(0) \rightarrow Y(1)$. Note that Supplementary Figure 1A satisfies the two core assumptions from the manuscript, where MAR among survivors means that

*To whom correspondence should be addressed.
\[ Pr\{R(1) = 1 \mid L(0), A(0), W(0), Y(1), Z(1) = 1\} = Pr\{R(1) = 1 \mid L(0), A(0), W(0), Z(1) = 1\}. \]

Assumptions encoded in Supplementary Figure 1A may be relaxed, yet satisfy the two core assumptions. Figure 1B has \( U(0) \), an unmeasured factor, with pathway \( W(0) \leftarrow U(0) \rightarrow Y(1) \) that is independent of \( Z(1) \) given \( \{L(0), A(0), W(0)\} \). Conditioning on \( \{L(0), A(0), W(0), Z(1) = 1\} \) does not open a \( U(0) - R(1) \) path, thus omitting \( U(0) \) does not lead to bias (it factors from the g-formula), but it impacts meaning (see TGSW). Although \( U(0) \) is plausible, it will henceforth be excluded for ease of exposition.

**B. G-FORMULA, \( g\{\bar{a}(j)\} \), FOR \( j = 0, \ldots, J \)**

In this Section, we generalize the g-formulas presented in (2.1) and (2.2) in the main paper to an arbitrary follow-up visit that includes \( W \). Now, let \( \bar{V}(j) = \{\bar{Y}(j), \bar{L}(j), \bar{W}(j)\} \) and \( \bar{v}(j) = \{\bar{y}(j), \bar{l}(j), \bar{w}(j)\} \). For general \( j = 0, \ldots, J \), the g-formula is

\[
g\{\bar{a}(j)\} = \int E\{Y(j + 1) \mid \bar{A}(j) = \bar{a}(j), Z(j + 1) = 1, \bar{V}(j) = \bar{v}(j)\} \times dF_{\bar{V}(j)}[\bar{v}(j)], \quad \text{(b.1)}
\]
Supplementary material: Double robustness with deaths

where

$$dF_{V(j)\{\bar{\epsilon}(j)\}} = dF_{W(j)\{\bar{\alpha}(j),Z(j)=1,Y(j)=y(j),L(j)=l(j),V(j-1)=\bar{v}(j-1)\{w(j)\}}$$

$$\times dF_{L(j)\{\bar{\alpha}(j-1)=\bar{a}(j-1),Z(j)=1,Y(j)=y(j),V(j-1)=\bar{v}(j-1)\{l(j)\}}$$

$$\times dF_{Y(j)|Z(j)=1,\bar{\alpha}(j-1)=\bar{a}(j-1),V(j-1)=\bar{v}(j-1)\{y(j)\}}$$

$$\times dF_{W(j-1)|\bar{\alpha}(j-1)=\bar{a}(j-1),Z(j-1)=1,Y(j-1)=y(j-1),L(j-1)=l(j-1),V(j-2)=\bar{v}(j-2)\{w(j-1)\}}$$

$$\times dF_{L(j-1)|\bar{\alpha}(j-2)=\bar{a}(j-2),Z(j-1)=1,Y(j-1)=y(j-1),V(j-2)=\bar{v}(j-2)\{l(j-1)\}}$$

$$\times dF_{Y(j-1)|Z(j-1)=1,\bar{\alpha}(j-2)=\bar{a}(j-2),V(j-2)=\bar{v}(j-2)\{y(j-1)\}}$$

$$\vdots$$

$$\times dF_{W(1)|\bar{\alpha}(1)=\bar{a}(1),Z(1)=1,Y(1)=y(1),L(1)=l(1),V(0)=\bar{v}(0)\{w(1)\}}$$

$$\times dF_{L(1)|A(0)=a(0),Z(1)=1,Y(1)=y(1),V(0)=\bar{v}(0)\{l(1)\}}$$

$$\times dF_{Y(1)|Z(1)=1,A(0)=a(0),V(0)=\bar{v}(0)\{y(1)\}}$$

$$\times dF_{W(0)|A(0)=a(0),L(0)=l(0)\{w(0)\}} \times dF_{L(0)\{l(0)\}}, \quad (b.2)$$

and \(F_{W(j)},\bar{\alpha}(j),Z(j)=1,Y(j)=y(j),L(j)=l(j),V(j-1)=\bar{v}(j-1)\{w(j)\})\) is a conditional distribution of \(W(j)\) at \(w(j)\), with other distributions similarly defined. By MAR, \((b.1)\) equals the \(g\)-formula with first term conditioned on \(R(j+1) = 1\) and the distributions of \(W(k), L(k),\) and \(Y(k)\) conditioned on \(R(k) = 1\), for \(k = 1, \ldots, j\) in \((b.2)\).

C. CAUSAL INTERPRETATION OF \(g\{\bar{\alpha}(j)\} - g\{\bar{\alpha}'(j)\}\) FOR \(j = 0, \ldots, J\)

Consider the structural equation models of Section 2.2 in the main paper. We augment the models to include \(W\). The models are now

\[
L(j) = \begin{cases} 
  g_{L(j)}\{\bar{A}(j-1), \bar{V}(j-1), Y(j), \epsilon_{L(j)}\} & \text{if } Z(j) = 1 \\
  \text{undefined} & \text{if } Z(j) = 0 
\end{cases} 
\]

\[
A(j) = \begin{cases} 
  g_{A(j)}\{\bar{A}(j-1), \bar{V}(j-1), Y(j), L(j), \epsilon_{A(j)}\} & \text{if } Z(j) = 1 \\
  \text{undefined} & \text{if } Z(j) = 0 
\end{cases} 
\]
\[ W(j) = \begin{cases} \{g W(j) \{ \bar{A}(j), \bar{V}(j-1), Y(j), L(j), \epsilon W(j) \} \} & \text{if } Z(j) = 1 \\ \text{undefined} & \text{if } Z(j) = 0 \end{cases} \] (c.3)

\[ Z(j + 1) = \begin{cases} \{g Z(j+1) \{ \bar{A}(j), \bar{V}(j), \epsilon Z(j+1) \} \} & \text{if } Z(j) = 1 \\ 0 & \text{if } Z(j) = 0 \end{cases} \] (c.4)

\[ Y(j + 1) = \begin{cases} \{g Y(j+1) \{ \bar{A}(j), \bar{V}(j), \epsilon Y(j+1) \} \} & \text{if } Z(j + 1) = 1 \\ \text{undefined} & \text{if } Z(j + 1) = 0 \end{cases} \] (c.5)

To derive the causal contrasts identified by \( g\{\bar{a}(j)\} - g\{\bar{a}'(j)\} \) for \( j = 0, \ldots, J \), we need counterfactuals of components of \( \bar{V}(j) \) that may mediate the effect of \( \bar{A}(j) \) on both \( Y(k) \) and \( Z(k) \) for \( k \in \{j + 1, \ldots, J + 1\} \). Let \( \bar{V}^*(j) \) denote these components of \( \bar{V}(j) \), where

\[ \bar{V}^*(j) = \{W(j), L(1), \ldots, L(j), Y(1), \ldots, Y(j)\}. \]

Similarly, let \( \bar{v}^*(j) = \{\bar{w}(j), l(1), \ldots, l(j), y(1), \ldots, y(j)\} \). Replacing observables with counterfactuals in (c.1)-(c.5) in the main paper, setting \( \{\bar{A}(j), \bar{V}^*(j)\} \) equal to \( \{\bar{a}(j), \bar{v}^*(j)\} \) implies

\[ Y_{\bar{a}(j), \bar{a}^*(j)}(j + 1) \perp Z_{\bar{a}(j), \bar{v}^*(j)}(j + 1) \mid \{Z_{\bar{a}(j), \bar{v}^*(j)}(2), L(0)\} \]

for \( \bar{a}(j) \neq \bar{a}'(j) \) or \( \{\bar{v}(j)\} \neq \{\bar{v}'(j)\} \). (c.6)

The PS controlled direct effect, within the stratum \( L(0) = l(0) \) setting \( \bar{V}^*(j) \) equal to \( \bar{v}^*(j) \) is therefore

\[ E\{Y_{\bar{a}(j), \bar{v}^*(j)}(j + 1) - Y_{\bar{a}'(j), \bar{v}^*(j)}(j + 1) \mid Z_{\bar{a}(j), \bar{v}^*(j)}(j + 1) = Z_{\bar{a}'(j), \bar{v}^*(j)}(j + 1) = 1, L(0) = l(0)\} \]

\[ = E[g_{Y(j+1)}\{\bar{a}(j), \bar{v}(j), \epsilon Y(j+1)\} - g_{Y(j+1)}\{\bar{a}'(j), \bar{v}(j), \epsilon Y(j+1)\}] \] (c.7)

Additional structural assumptions that are analogous to (c.6) are encoded by the structural equations in (c.1)-(c.5). For example setting \( \{\bar{A}(j - 1), \bar{V}^*(j - 1), Y(j)\} \) equal to \( \{\bar{a}(j - 1), \bar{v}^*(j - 1), y(j)\} \) in (c.1), implies

\[ L_{\bar{a}(j-1), \bar{v}^*(j-1), y(j)}(j) \perp Z_{\bar{a}'(j-1), \bar{v}^*(j-1), y'(j)}(j) \mid \{Z_{\bar{a}(j-1), \bar{v}^*(j-1), y(j)}(j), L(0)\} \]

for \( \bar{a}(j - 1) \neq \bar{a}'(j - 1) \) or \( \{\bar{v}^*(j - 1), y(j)\} \neq \{\bar{v}'^*(j - 1), y'(j)\} \). (c.8)
Similarly, setting \( \{\bar{A}(j), \bar{V}^*(j-1), Y(j), L(j)\} \) equal to \( \{\bar{a}(j), \bar{v}^*(j-1), y(j), l(j)\} \) in (c.3) implies

\[
W_{\bar{a}(j), \bar{v}^*(j-1), y(j), l(j)}(j) \perp Z_{a^*(j), \bar{v}^*(j-1), y(j), l^*(j)}(j) | \{Z_{\bar{a}(j), \bar{v}^*(j-1), y(j), l(j)}(j), L(0)\}
\]

for \( \bar{a}(j) \neq \tilde{a}'(j) \) or \( \{\bar{v}^*(j-1), y(j), l(j)\} \neq \{\bar{v}^*(j-1), y^*(j), l^*(j)\} \). (c.9)

Under (c.1)-(c.5), and the implied assumption (c.6),

\[
E\{Y(j+1) | \bar{A}(j) = \bar{a}(j), Z(j+1) = 1, \bar{V}(j) = \bar{v}(j)\} = E[g_{Y(j+1)}(\bar{a}(j), \bar{v}(j), \epsilon_{Y(j+1)})]
\]

\[
= \int g_{Y(j+1)}(\bar{a}(j), \bar{v}(j), \epsilon) dF_{\epsilon_{Y(j+1)}}(\epsilon) = E\{Y_{\bar{a}(j), \bar{v}^*(j)}(j+1) | Z_{\bar{a}(j), \bar{v}^*(j)}(j+1) = 1, L(0) = l(0)\}
\]

\[
= E\{Y_{\bar{a}(j), \bar{v}^*(j)}(j+1) | Z_{\bar{a}(j), \bar{v}^*(j)}(j+1) = Z_{\tilde{a}'(j), \bar{v}^*(j)}(j+1) = 1, L(0) = l(0)\}
\]

\[
= E\{Y_{\bar{a}(j), \bar{v}^*(j)}(j+1) | Z_{\bar{a}(j), \bar{v}^*(j)}(j+1) = Z_{\tilde{a}'(j), \bar{v}^*(j)}(j+1) = 1, L(0) = l(0)\}
\]

where \( F_{\epsilon_{Y(j+1)}}(\epsilon) \) is the distribution of \( \epsilon_{Y(j+1)} \) evaluated at \( \epsilon \). Define
\[ dF_{\bar{a}(j)}(j) \{ \bar{\theta}(j) \} \]
\[ = dF_{\bar{a}(j), \bar{L}(j), y(j), \bar{w}(j-1)(j)} Z_{\bar{a}(j), \bar{L}(j), y(j), \bar{w}(j-1)(j)} = 1, \bar{L}(0) = t(0) \{ w(j) \} \]
\[ \times dF_{\bar{L}(j-1), y(j-1)(j)} Z_{\bar{a}(j-1), \bar{L}(j-1), y(j-1), \bar{w}(j-1)(j)} = 1, \bar{L}(0) = t(0) \{ l(j-1) \} \]
\[ \times dF_{\bar{Y}(j-1), \bar{a}(j-1)(j)} Z_{\bar{a}(j-1), \bar{Y}(j-1), \bar{a}(j-1), \bar{w}(j-1)(j-1)} = 1, \bar{L}(0) = t(0) \{ y(j-1) \} \]
\[ \times dF_{\bar{W}(j-2), \bar{a}(j-2)(j-1)} Z_{\bar{a}(j-2), \bar{W}(j-2), \bar{a}(j-2)(j-1)} = 1, \bar{L}(0) = t(0) \{ w(j-2) \} \]
\[ \times dF_{\bar{L}(j-2), \bar{a}(j-2)(j-1)} Z_{\bar{a}(j-2), \bar{L}(j-2), \bar{a}(j-2), \bar{w}(j-2)(j-1)} = 1, \bar{L}(0) = t(0) \{ l(j-2) \} \]
\[ \times dF_{\bar{Y}(j-2), \bar{a}(j-2)(j-2)} Z_{\bar{a}(j-2), \bar{Y}(j-2), \bar{a}(j-2), \bar{w}(j-2)(j-2)} = 1, \bar{L}(0) = t(0) \{ y(j-2) \} \]
\[ \vdots \]
\[ \times dF_{\bar{W}(1), \bar{a}(1)(1)} Z_{\bar{a}(1), \bar{W}(1), \bar{a}(1)(1)} = 1, \bar{L}(0) = t(0) \{ w(1) \} \]
\[ \times dF_{\bar{L}(0), \bar{a}(0)(1)} Z_{\bar{a}(0), \bar{L}(0), \bar{a}(0), \bar{w}(0)(1)} = 1, \bar{L}(0) = t(0) \{ l(1) \} \]
\[ \times dF_{\bar{Y}(0), \bar{a}(0)(1)} Z_{\bar{a}(0), \bar{Y}(0), \bar{a}(0), \bar{w}(0)(1)} = 1, \bar{L}(0) = t(0) \{ y(1) \} \]
\[ \times dF_{\bar{W}(0), \bar{a}(0)(1)} Z_{\bar{a}(0), \bar{W}(0), \bar{a}(0), \bar{w}(0)(1)} = 1, \bar{L}(0) = t(0) \{ w(0) \} \times dF_{\bar{L}(0), \bar{a}(0)(1)} = t(0) \{ l(0) \}, \quad (c.11) \]

where the use of counterfactuals follows from \((c.1)-(c.5)\) and the independence assumptions \((c.6)-(c.9)\). Replacing the expectation in \((b.1)\) with \((c.10)\) and replacing \((b.2)\) with \((c.11)\) shows that the g-formula \((b.1)\), and hence IPW and AIPW, identifies the causal contrast...
Supplementary material: Double robustness with deaths

\[ g\{\bar{a}(j)\} - g\{\bar{a}'(j)\} = \int E\{Y_{\bar{a}(j), a^*}(j+1) \mid Z_{\bar{a}(j), a^*}(j + 1) = 1, L(0) = l(0)\} \times dF_{V_{a'(j)}(j)}\{\bar{v}(j)\} \]

\[ - \int E\{Y_{\bar{a}'(j), a^*}(j + 1) \mid Z_{\bar{a}'(j), a^*}(j + 1) = 1, L(0) = l(0)\} \times dF_{V_{a'(j)}(j)}\{\bar{v}(j)\} \]

\[ = \int E\{Y_{\bar{a}(j), a^*}(j + 1) \mid Z_{\bar{a}(j), a^*}(j + 1) = 1, L(0) = l(0)\} \times dF_{V_{a'(j)}(j)}\{\bar{v}(j)\} \]

\[ - \int E\{Y_{\bar{a}'(j), a^*}(j + 1) \mid Z_{\bar{a}'(j), a^*}(j + 1) = 1, L(0) = l(0)\} \times dF_{V_{a'(j)}(j)}\{\bar{v}(j)\} \]

\[ + \int E\{Y_{\bar{a}(j), a^*}(j + 1) \mid Z_{\bar{a}(j), a^*}(j + 1) = 1, L(0) = l(0)\} \times \left[ dF_{V_{a'(j)}(j)}\{\bar{v}(j)\} - dF_{V_{a'(j)}(j)}\{\bar{v}(j)\} \right] \]

\[ + \int E\{Y_{\bar{a}(j), a^*}(j + 1) \mid Z_{\bar{a}(j), a^*}(j + 1) = 1, L(0) = l(0)\} \times \left[ dF_{V_{a'(j)}(j)}\{\bar{v}(j)\} - dF_{V_{a'(j)}(j)}\{\bar{v}(j)\} \right] \]

where \( \bar{v}'(j) \) is a fixed reference value of \( V^*(1) \). Rearranging the terms in (c.12), noting that the last term equals 0, and applying (c.10) leads to

\[ g\{\bar{a}(j)\} - g\{\bar{a}'(j)\} = \int E\{Y_{\bar{a}(j), a^*}(j + 1) - Y_{\bar{a}'(j), a^*}(j + 1) \mid Z_{\bar{a}(j), a^*}(j + 1) = Z_{\bar{a}'(j), a^*}(j + 1) = 1, L(0) = l(0)\} \times dF_{V_{a'(j)}(j)}\{\bar{v}(j)\} \]

\[ + \int E\{Y_{\bar{a}(j), a^*}(j + 1) - Y_{\bar{a}'(j), a^*}(j + 1) \mid Z_{\bar{a}(j), a^*}(j + 1) = Z_{\bar{a}'(j), a^*}(j + 1) = 1, L(0) = l(0)\} \times \left[ dF_{V_{a'(j)}(j)}\{\bar{v}(j)\} - dF_{V_{a'(j)}(j)}\{\bar{v}(j)\} \right] \]

\[ = \text{Direct Effect}(j) + \text{Indirect Effect}(j). \quad (c.13) \]

Direct Effect(\( j \)) in expression (c.13) is (c.7) averaged over \( W_{\hat{a}'(k), l(k), y(k), a^*(k-1)}(k) \), given

\[ \{ \bar{Z}_{\hat{a}(k), l(k), y(k), a^*(k-1)}(k) = Z_{\hat{a}'(k), l(k), y(k), a^*(k-1)}(k) = 1 \}; \bar{L}_{\hat{a}'(k-1), y(k), a^*(k-1)}(k), \text{ given } \{ \bar{Z}_{\hat{a}(k-1), y(k), a^*(k-1)}(k) = \bar{Z}_{\hat{a}'(k-1), y(k), a^*(k-1)}(k) = 1 \}; \bar{Y}_{\hat{a}'(k-1), a^*(k-1)}(k), \text{ given } \{ \bar{Z}_{\hat{a}(k-1), a^*(k-1)}(k) = \bar{Z}_{\hat{a}'(k-1), a^*(k-1)}(k) = 1 \}, \text{ for } k = 1, \ldots, j; \text{ and } W_{\hat{a}'(0)}(0) \text{ and } L(0). \] Indirect Effect(\( j \)) is the integrated indirect effect
the nuisance tangent space for $A$ let $F$ to $a$ and $W$ is $E$ $f$

Namely, Direct Effect(0) is the PS controlled direct effect of $A$ for outcome $W(t)$ and $W_{a′(0)}(0)$; Indirect Effect(0) is the difference in average effects of $W(0)$ on $Y(1)$ setting $A(0)$ to $a(0)$ versus $a′(0)$, comparing the distribution of $W_{a(0)}(0)$ to that of $W_{a′(0)}(0)$.

D. AIPW Estimation

Let $\mathcal{F}(k) = \{\bar{A}(k-1), \bar{W}(k-1), \bar{R}(k) = 1, \bar{Z}(k) = 1, \bar{Y}(k), \bar{L}(k)\}$, $k = 0, \ldots, J$, be pre-A($k$) history. Let $\Lambda_{R(k+1)} = \{(R(k+1)/p_{R(k+1)} - 1)\phi_{R(k+1)}\{Z(k+1) = 1, A(k), \mathcal{F}(k)\}\}$ be the nuisance tangent space for $R(k+1)$, $k = 0, \ldots, J$. Also, let $\Lambda_{Z(k+1)} = \{(Z(k+1)/p_{Z(k+1)} - 1)\phi_{Z(k+1)}\{A(k), \mathcal{F}(k)\}\}$ be the nuisance tangent space for $Z(k+1)$, and $\Lambda_{A(k)} = \{\{A(k) - f_{A(k)}\}\phi_{A(k)}\{\mathcal{F}(k)\}\}$ be the nuisance tangent space for $A(k)$. Optimal $\phi_{R(k+1)}\{Z(k+1) = 1, A(k), \mathcal{F}(k)\}$ is $E\{U_{IPW} | Z(k+1) = 1, A(k), \mathcal{F}(k)\}$, optimal $\phi_{Z(k+1)}\{A(k), \mathcal{F}(k)\}$ is $E\{U_{IPW} | A(k), \mathcal{F}(k)\}$, and optimal $\phi_{A(k)}\{\mathcal{F}(k)\}$ is $E\{A(k) - f_{A(k)}\}U_{IPW} | \mathcal{F}(k)\}$ (Robins, 1999). This factoring of mechanisms implies that the nuisance tangent space at $k$ is $\Lambda_{k} = \Lambda_{R(k+1)} \oplus \Lambda_{Z(k+1)} \oplus \Lambda_{A(k)}$.

Thus, the projection is $\Pi\{U_{IPW} | \Lambda_{R(k+1)}\} + \Pi\{U_{IPW} | \Lambda_{Z(k+1)}\} + \Pi\{U_{IPW} | \Lambda_{A(k)}\}$ where

$$\Pi\{U_{IPW} | \Lambda_{R(k+1)}\} = E\{U_{IPW} | R(k+1)Z(k+1) = 1, A(k), \mathcal{F}(k)\} - E\{U_{IPW} | Z(k+1) = 1, A(k), \mathcal{F}(k)\}, \quad (d.1)$$

$$\Pi\{U_{IPW} | \Lambda_{Z(k+1)}\} = E\{U_{IPW} | Z(k+1) = 1, A(k), \mathcal{F}(k)\} - E\{U_{IPW} | A(k), \mathcal{F}(k)\}, \quad (d.2)$$

$$\Pi\{U_{IPW} | \Lambda_{A(k)}\} = E\{U_{IPW} | A(k), \mathcal{F}(k)\} - E[U_{IPW} | \mathcal{F}(k)]. \quad (d.3)$$
Dropout, mortality, and exposure mechanisms can be factored into time-specific terms due to the two core assumptions. Thus, $\Lambda = \Lambda_0 \oplus \Lambda_1 \oplus \cdots \oplus \Lambda_J$. The projection equals

$$\Pi \{ U_{IPW} \mid \Lambda \} = \sum_{k=0}^{J} \left[ \Pi \{ U_{IPW} \mid \Lambda_{R(k+1)} \} + \Pi \{ U_{IPW} \mid \Lambda_{Z(k+1)} \} + \Pi \{ U_{IPW} \mid \Lambda_{A(k)} \} \right],$$

and the resulting estimating equation for $\beta$ is

$$U_{AIPW} = \sum_{i=1}^{N} U_{i,AIPW} = \sum_{i=1}^{N} U_{i,IPW} - \Pi \{ U_{IPW} \mid \Lambda \}. \quad (d.4)$$

Regressing $U_{i,IPW}$ on the past estimates (d.1)-(d.3), $\hat{\beta}_{AIPW}$ solves (d.4) and is doubly robust; that is, (d.4) has mean 0 if outcome regressions or $SW$ are correct. The variance of $\hat{\beta}_{AIPW}$ follows from theory in van der Laan and Robins (2003) and Tsiatis (2006). Technically, the goal of the outcome regressions is to average $U_{i,IPW}$ over the possible histories of $A$, $Z$, and $R$. One approach is to literally perform linear regressions $U_{i,IPW}$ on past variables; however, for categorical $A$, one can take expected values over each possible value of $A$ (and $Z$ and $R$) at each time point. This approach for binary $A$ and $R$ is demonstrated in Bryan, Yu and van der Laan (2004).

### d.1 Variance of $\hat{\beta}_{AIPW}$

Under standard regularity conditions, $\hat{\beta}_{AIPW}$ is a consistent asymptotically normal, regular asymptotically linear estimator for $\beta$ (van der Laan and Robins, 2003; Tsiatis, 2006) with influence function

$$- D^{-1} U_{AIPW} + \Pi \{ D^{-1} U_{AIPW} \mid \Lambda_{SW} \}, \quad (d.5)$$

where $\Lambda_{SW} \subset \Lambda$ is the tangent space for the dropout, mortality, and exposure assignment mechanisms under assumed models, and $D = E \{ \frac{\partial}{\partial \beta} U_{AIPW} \}$. By treating $SW$ as known rather than estimated, the projection in (d.5) need not be taken and can be ignored, and the variance for
\( \hat{\beta}_{AIPW} \) can be conservatively estimated as

\[
\hat{D}^{-1} \left\{ N^{-1} \sum_{i=1}^{N} U_{i,AIPW} U_{i,AIPW}^T \right\} \left( \hat{D}^{-1} \right)^T,
\]

where \( \hat{D} = N^{-1} \sum_{i=1}^{N} \frac{1}{\beta} U_{i,AIPW} \). This variance estimator takes advantage of the counterintuitive property that the asymptotic variance is larger when SW are known than when SW are estimated and plugged into \( U_{i,AIPW} \) (van der Laan and Robins, 2003, Tsiatis, 2006).

d.2 Proof of double robustness

For ease of exposition, we will consider unstabilized weights (SW without the numerator terms), focus on the second follow-up visit, and suppress the subscript \( i \). Therefore, the second item of the vector (2.23) in the main paper (i.e., the term corresponding to the second follow-up visit) can be written as

\[
U_{2,IPW} = \left[ \prod_{k=0}^{1} \left\{ \frac{A(k)}{f_{A(k)}} + \frac{1 - A(k)}{1 - f_{A(k)}} \left\{ Z(k + 1) \frac{R(k + 1)}{f_{R(k+1)}} \right\} \delta' \frac{g{\{ \bar{A}(1), \beta \}}}{\delta \beta} \right\} \right].
\]

By applying this notation to the projections (d.1)–(d.3), we can rewrite the second item of the vector (d.4) (i.e., the term corresponding to the second follow-up visit) as
Supplementary material: Double robustness with deaths

\[ U_{2,\text{AIPW}} = \left[ \prod_{k=0}^{1} \left\{ \frac{A(k)}{f_{A(k)}} + \frac{1 - A(k)}{1 - f_{A(k)}} \left( \frac{Z(k + 1) R(k + 1)}{f_{Z(k+1)} f_{R(k+1)}} \right) \delta g(\bar{A}(1), \beta) \right\} \frac{\delta}{\delta \beta} [Y(2) - g(\bar{A}(1), \beta)] \right. \]  
\[ \left. - \left[ \prod_{k=0}^{1} \left\{ \frac{A(k)}{f_{A(k)}} + \frac{1 - A(k)}{1 - f_{A(k)}} \left( \frac{Z(k + 1) R(k + 1)}{f_{Z(k+1)} f_{R(k+1)}} \right) \delta g(\bar{A}(1), \beta) \right\} \frac{\delta}{\delta \beta} \times [E\{Y(2) | R(2)Z(2) = 1, W(1), A(1), \mathcal{F}(1)\} - g(\bar{A}(1), \beta)] \right. \]  
\[ \left. + \left\{ \frac{A(0)}{f_{A(0)}} + \frac{1 - A(0)}{1 - f_{A(0)}} \left( \frac{Z(1) R(1)}{f_{Z(1)} f_{R(1)}} \right) \left( \sum_{a(1) \in \{0, 1\}} \delta g(A(0), a(1), \beta) \right) \frac{\delta}{\delta \beta} \times [E\{Y(2) | A(1) = a(1), \mathcal{F}(1)\} - g(A(0), a(1), \beta)] \right. \]  
\[ \left. - \left\{ \frac{A(0)}{f_{A(0)}} + \frac{1 - A(0)}{1 - f_{A(0)}} \left( \frac{Z(1) R(1)}{f_{Z(1)} f_{R(1)}} \right) \left( \sum_{a(1) \in \{0, 1\}} \delta g(A(0), a(1), \beta) \right) \frac{\delta}{\delta \beta} \right. \]  
\[ \left. \times [E\{Y(2) | A(1) = a(1), R(1)Z(1) = 1, W(0), A(0), \mathcal{F}(0)\} - g(A(0), a(1), \beta)] \right. \]  
\[ \left. + \left\{ \frac{A(0)}{f_{A(0)}} + \frac{1 - A(0)}{1 - f_{A(0)}} \left( \sum_{a(1) \in \{0, 1\}} \delta g(A(0), a(1), \beta) \right) \frac{\delta}{\delta \beta} \times [E\{Y(2) | A(1) = a(1), W(0), A(0), \mathcal{F}(0)\} - g(A(0), a(1), \beta)] \right. \]  
\[ \left. - \left\{ \frac{A(0)}{f_{A(0)}} + \frac{1 - A(0)}{1 - f_{A(0)}} \left( \sum_{a(1) \in \{0, 1\}} \delta g(A(0), a(1), \beta) \right) \frac{\delta}{\delta \beta} \right. \]  
\[ \left. \times [E\{Y(2) | A(1) = a(1), A(0), \mathcal{F}(0)\} - g(A(0), a(1), \beta)] \right. \]  
\[ \left. + \sum_{\bar{a}(1) \in \{0, 1\}^2} \delta g(\bar{a}(1), \beta) \frac{\delta}{\delta \beta} [E\{Y(2) | \bar{A}(1) = \bar{a}(1), \mathcal{F}(0)\} - g(\bar{a}(1), \beta)]. \right] \]
The double robustness property follows from the telescoping sum that results from taking iterated expectations, and because of the two core assumptions described in Section 2.1 of the main paper. If $f_{A(k)}$, $f_{Z(k+1)}$, or $f_{Z(k)+1}$, $k = 0$ or 1, are incorrect, but the outcome regressions are all correct, then the expectation of (d.6) over the conditional distribution of $Y(2)$ cancels with (d.7). The expectation of (d.8) over the conditional distribution of $W(1)$ cancels with (d.9). The expectation of (d.10) over the conditional distribution of $L(1)$ cancels with (d.11). The expectation of (d.12) over the conditional distribution of $W(0)$ cancels with (d.13). Finally, the expectation of (d.14) over the distribution of $L(0)$ equals 0. Therefore, $U_{2,AIPW}$ is unbiased when the weights are not all correct, but the outcome regressions are all correct.

If the outcome regressions are not all correct, but $f_{A(k)}$, $f_{Z(k+1)}$, and $f_{Z(k)+1}$, $k = 0, 1$, are all correct then the expectation of (d.6) equals 0. The expectation of (d.7) over the conditional distribution of \{R(2), Z(2)\} cancels with (d.8). The expectation of (d.9) over the conditional distribution of $A(1)$ cancels with (d.10). The expectation of (d.11) over the conditional distribution of $\{R(1), Z(1)\}$ cancels with (d.12). Finally, the expectation of (d.13) over the conditional distribution of $A(0)$ cancels with (d.14). Therefore, $U_{2,AIPW}$ is unbiased when the outcome regressions are not all correct, but the weights are all correct.

When the weights and the outcome regressions are both all correct, then the respective expectations of (d.6) and (d.14) both equal 0, and the intermediate terms form a telescoping sum that cancel with repeated applications of the iterated expectation.

Therefore, $\hat{\beta}_{AIPW}$ is doubly robust.

References


Robins, J.M. (1999). Marginal structural models versus structural nested models as tools for


Supplementary Table 1: Simulation scenarios. True models are data-generating models; analysis performed under both true and false models.

<table>
<thead>
<tr>
<th>True</th>
<th>$\logit{Pr(A(0) = 1 \mid L(0))} = \alpha \cdot [1, X, Y(0), XY(0)], \alpha = [0, 0.5, -0.75, 0.25].$</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>$\logit{Pr(Z_a(1) = 1 \mid L(0), a(0) = 0)} = \gamma \cdot [1, X, Y(0), XY(0)],$</td>
</tr>
<tr>
<td></td>
<td>$\gamma = [0.5, 0.2, -0.2, 0.1].$</td>
</tr>
<tr>
<td></td>
<td>$\logit{Pr(Z_a(1) = 1 \mid L(0), a(0) = 1)} = \gamma \cdot [1, X, Y(0), XY(0)],$</td>
</tr>
<tr>
<td></td>
<td>$\gamma = [1.5, 0.3, -0.35, 0.2].$</td>
</tr>
<tr>
<td></td>
<td>$m_1(L(0), a(0) = 0, Z(1) = 1) = \delta \cdot [1, X, Y(0)], \delta = [0, 0.2].$</td>
</tr>
<tr>
<td></td>
<td>$m_1(L(0), a(0) = 1, Z(1) = 1) = \delta \cdot [1, X, Y(0)], \delta = [2, -1.3].$</td>
</tr>
<tr>
<td></td>
<td>$\logit{Pr(R(1) = 1 \mid L(0), A(0), Z(1) = 1)} = \lambda \cdot [1, A(0), X, Y(0)],$</td>
</tr>
<tr>
<td></td>
<td>$\lambda = [1, 0.2, 0.02, 0.05].$</td>
</tr>
<tr>
<td></td>
<td>$\logit{Pr(A(1) = 1 \mid \bar{L}(1), a(0) = 0, \bar{Z}(1) = 1)} = \alpha \cdot [1, X, Y(1), Y(0)Y(1)],$</td>
</tr>
<tr>
<td></td>
<td>$\alpha = [0, 0.2, 0.2, 0].$</td>
</tr>
<tr>
<td></td>
<td>$\logit{Pr(A(1) = 1 \mid \bar{L}(1), a(0) = 1, \bar{Z}(1) = 1)} = \alpha \cdot [1, X, Y(1), Y(0)Y(1)],$</td>
</tr>
<tr>
<td></td>
<td>$\alpha = [0.5, 0, 0.05, 0.1].$</td>
</tr>
<tr>
<td></td>
<td>$\logit{Pr(Z_a(2) = 1 \mid \bar{L}(1), \bar{a}(1) = (0, 0), \bar{Z}(1) = 1} = \gamma \cdot [1, Y(0), Y(1), Y(0)Y(1)],$</td>
</tr>
<tr>
<td></td>
<td>$\gamma = [0.4, -0.2, 0.2, 0].$</td>
</tr>
<tr>
<td></td>
<td>$\logit{Pr(Z_a(2) = 1 \mid \bar{L}(1), \bar{a}(1) = (0, 1), \bar{Z}(1) = 1} = \gamma \cdot [1, Y(0), Y(1), Y(0)Y(1)],$</td>
</tr>
<tr>
<td></td>
<td>$\gamma = [0.6, -0.4, 0.4, 0.1].$</td>
</tr>
<tr>
<td></td>
<td>$\logit{Pr(Z_a(2) = 1 \mid \bar{L}(1), \bar{a}(1) = (1, 0), \bar{Z}(1) = 1} = \gamma \cdot [1, Y(0), Y(1), Y(0)Y(1)],$</td>
</tr>
<tr>
<td></td>
<td>$\gamma = [0.6, -0.4, 0.4, 0.1].$</td>
</tr>
<tr>
<td></td>
<td>$\logit{Pr(Z_a(2) = 1 \mid \bar{L}(1), \bar{a}(1) = (1, 1), \bar{Z}(1) = 1} = \gamma \cdot [1, Y(0), Y(1), Y(0)Y(1)],$</td>
</tr>
<tr>
<td></td>
<td>$\gamma = [0.8, -0.6, 0.6, 0.2].$</td>
</tr>
<tr>
<td></td>
<td>$m_2(\bar{L}(1), \bar{a}(1) = (0, 0), \bar{Z}(2) = 1) = \delta \cdot [X, Y(0), Y(1), XY(0), XY(1), Y(0)Y(1)],$</td>
</tr>
<tr>
<td></td>
<td>$\delta = [0, 1, 0, 0.0, -1].$</td>
</tr>
<tr>
<td></td>
<td>$m_2(\bar{L}(1), \bar{a}(1) = (0, 1), \bar{Z}(2) = 1) = \delta \cdot [X, Y(0), Y(1), XY(0), XY(1), Y(0)Y(1)],$</td>
</tr>
<tr>
<td></td>
<td>$\delta = [0.5, 0, -1, 0.5, -2, 0].$</td>
</tr>
<tr>
<td></td>
<td>$m_2(\bar{L}(1), \bar{a}(1) = (1, 0), \bar{Z}(2) = 1) = \delta \cdot [X, Y(0), Y(1), XY(0), XY(1), Y(0)Y(1)],$</td>
</tr>
<tr>
<td></td>
<td>$\delta = [0.5, 0.5, 0, 1.0, 0].$</td>
</tr>
<tr>
<td></td>
<td>$m_2(\bar{L}(1), \bar{a}(1) = (1, 1), \bar{Z}(2) = 1) = \delta \cdot [X, Y(0), Y(1), XY(0), XY(1), Y(0)Y(1)],$</td>
</tr>
<tr>
<td></td>
<td>$\delta = [0.5, 0.5, 0, -0.5, 1.1].$</td>
</tr>
<tr>
<td></td>
<td>$\logit{Pr(R(2) = 1 \mid \bar{L}(1), \bar{A}(1), \bar{Z}(2) = 1, R(1) = 1} = \lambda \cdot [1, A(0), A(1), X, Y(0), Y(1)],$</td>
</tr>
<tr>
<td></td>
<td>$\lambda = [1, 0.1, 0.1, 0.05, 0.05, -0.05].$</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>False</th>
<th>$\logit{Pr(A(0) = 1 \mid L(0))} = \alpha \cdot [1, X].$</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>$\logit{Pr(Z_a(1) = 1 \mid L(0), a(0) = 0)} = \gamma \cdot [1, X].$</td>
</tr>
<tr>
<td></td>
<td>$\logit{Pr(Z_a(1) = 1 \mid L(0), a(0) = 1)} = \gamma \cdot [1, X].$</td>
</tr>
<tr>
<td></td>
<td>$m_1(L(0), a(0) = 0, Z(1) = 1) = \delta \cdot [1, X].$</td>
</tr>
<tr>
<td></td>
<td>$m_1(L(0), a(0) = 1, Z(1) = 1) = \delta \cdot [1, X].$</td>
</tr>
<tr>
<td></td>
<td>$\logit{Pr(R(1) = 1 \mid L(0), A(0), Z(1) = 1} = \lambda \cdot [1, A(0), X].$</td>
</tr>
<tr>
<td></td>
<td>$\logit{Pr(A(1) = 1 \mid L(1), a(0) = 0, Z(1) = 1} = \alpha \cdot [1, X].$</td>
</tr>
<tr>
<td></td>
<td>$\logit{Pr(A(1) = 1 \mid L(1), a(0) = 1, Z(1) = 1} = \alpha \cdot [1, X].$</td>
</tr>
<tr>
<td></td>
<td>$\logit{Pr(Z_a(2) = 1 \mid \bar{L}(1), \bar{a}(1) = (0, 0), \bar{Z}(1) = 1} = \gamma \cdot [1, X].$</td>
</tr>
<tr>
<td></td>
<td>$\logit{Pr(Z_a(2) = 1 \mid \bar{L}(1), \bar{a}(1) = (0, 1), \bar{Z}(1) = 1} = \gamma \cdot [1, X].$</td>
</tr>
<tr>
<td></td>
<td>$\logit{Pr(Z_a(2) = 1 \mid \bar{L}(1), \bar{a}(1) = (1, 0), \bar{Z}(1) = 1} = \gamma \cdot [1, X].$</td>
</tr>
<tr>
<td></td>
<td>$\logit{Pr(Z_a(2) = 1 \mid \bar{L}(1), \bar{a}(1) = (1, 1), \bar{Z}(1) = 1} = \gamma \cdot [1, X].$</td>
</tr>
<tr>
<td></td>
<td>$m_2(\bar{L}(1), \bar{a}(1) = (0, 0), \bar{Z}(2) = 1) = \delta \cdot [1, X].$</td>
</tr>
<tr>
<td></td>
<td>$m_2(\bar{L}(1), \bar{a}(1) = (0, 1), \bar{Z}(2) = 1) = \delta \cdot [1, X].$</td>
</tr>
<tr>
<td></td>
<td>$m_2(\bar{L}(1), \bar{a}(1) = (1, 0), \bar{Z}(2) = 1) = \delta \cdot [1, X].$</td>
</tr>
<tr>
<td></td>
<td>$m_2(\bar{L}(1), \bar{a}(1) = (1, 1), \bar{Z}(2) = 1) = \delta \cdot [1, X].$</td>
</tr>
</tbody>
</table>
Supplementary Fig. 1: Directed acyclic graphs of the hypothesized pathway linking $A(0)$ to $Y(1)$. In both figures, $L(0)$ are $A(0) - Y(1)$ confounders, $Z(1)$ is vital status, $W(0)$ are endogenous common causes of $Z(1)$ and $Y(1)$, and $R(1)$ is observation status. The absence of an arrow from $Y(1)$ to $R(1)$ encodes missing at random. Bold lines represent causal pathways through which $A(0)$ can affect $Y(1)$. Dashed lines represent relationships between $Z(1)$ and subsequent variables that are undefined. Under these DAGs, a conventional completers-only analysis would need to condition on $\{L(0), W(0)\}$ to remove selection bias, but doing so would adjust out causal pathways of interest, $A(0) \rightarrow W(0) \rightarrow Y(0)$ in A and B and also $A(0) \rightarrow U(0) \rightarrow W(0) \rightarrow Y(0)$ in B. In B, $U(0)$ is an unmeasured factor like that considered in TGSW whose omission from analysis does not bias results.