Supplementary material for "Multiply robust inference for statistical interactions"

Stijn Vansteelandt
Department of Applied Mathematics and Computer Sciences
Ghent University, 281 (S9) Krijgslaan, 9000 Ghent, Belgium

Tyler J. Vanderweele
Department of Health Studies, University of Chicago
5841 South Maryland Avenue, MC 2007, Chicago, IL 60637

Eric J. Tchetgen Tchetgen
Departments of Biostatistics and Epidemiology
Harvard School of Public Health, Boston, MA 02115, U.S.A.

and James M. Robins
Departments of Biostatistics and Epidemiology
Harvard School of Public Health, Boston, MA 02115, U.S.A.

For notational convenience, we will ignore boldface notation for vectors and matrices in these Supplementary Materials.

A.1 Proof of Theorem 1

Let P denote the law of (Y, A, X). Suppose first that $q_2(X, A_2)$ and $q_1(X, A_1)$ are known functions. With this additional restriction, model $\mathcal{A}$ is the semipara-

metric regression model

$$Y - m(X, A; \beta^*) = \epsilon, \text{ when } g(x) = x \text{ and}$$
$$Y \exp\{-m(X, A; \beta^*)\} = 1 + \epsilon, \text{when } g(x) = \exp(x)$$

where $m(X, A; \beta) = q_3(A, X; \beta) + q_2(X, A_2) + q_1(X, A_1) + h(X)$ with $h(X)$ an unknown function and $m(X, A; \beta) - h(X)$ a known function, ϵ has an unknown distribution satisfying $E(\epsilon|A, X) = 0$ and the joint law of (A, X) is unknown. The nuisance tangent space for this model is $\Lambda^{SR}_{nuis} = \Lambda^{SR}_{1,nuis} + \Lambda^{SR}_{2,nuis} + \Lambda^{SR}_{3,nuis}$ where $\Lambda^{SR}_{1,nuis} = \{a(A, X) : E[a(A, X)] = 0\} \cap L_2(P)$ is the closed linear span of all scores for parametric submodels for the joint law of (A, X), $\Lambda^{SR}_{2,nuis} = \{a(\epsilon, A, X) : E[a(\epsilon, A, X)|A, X] = 0, E[\epsilon a(\epsilon, A, X)|A, X] = 0\} \cap L_2(P)$ is the closed linear span of all scores for parametric submodels for the joint conditional law of ϵ, given (A, X), that satisfy $E(\epsilon|A, X) = 0$, and $\Lambda^{SR}_{3,nuis} = \left\{\epsilon \text{Var}(\epsilon|A, X)^{-1} b(X)\right\} \cap L_2(P)$ is the closed linear span of all scores for parametric submodels for the unknown function $h(X)$. Note that $\Lambda^{SR}_{1,nuis}, \Lambda^{SR}_{2,nuis}, \Lambda^{SR}_{3,nuis}$ are mutually orthogonal.

Denote $\sigma^2(A, X) \equiv \text{Var}(\epsilon|A, X)$. The orthocomplement $\Lambda^{SR,\perp}_{nuis}$ to Λ^{SR}_{nuis} in the Hilbert space $L^0_2(P)$ (with covariance inner product) of functions in $L_2(P)$ with mean zero is

$$\begin{aligned}
\Lambda^{SR,\perp}_{nuis} &= \{\epsilon[d(A, X) - E\{d(A, X)|X\}]\} \cap L^0_2(P) \\
&= \left\{\epsilon\sigma^{-2}(A, X) J(d)\right\} \cap L^0_2(P) \text{ where} \\
J(d) &= d(A, X) - E\left[\sigma^{-2}(A, X)|X\right]^{-1} E\left[\sigma^{-2}(A, X) d(A, X)|X\right]
\end{aligned}$$

since (i) $\epsilon[d(A, X) - E\{d(A, X)|X\}]$ is orthogonal (uncorrelated) to $\Lambda^{SR}_{1,nuis}, \Lambda^{SR}_{2,nuis}$ and $\Lambda^{SR}_{3,nuis}$ in $L^0_2(P)$; and (ii) the projection of $\epsilon\sigma^{-2}(A, X) d(A, X)$ on $\Lambda^{SR,\perp}_{nuis}$ is $\epsilon\sigma^{-2}(A, X)\left[d(A, X) - E\{\sigma^{-2}(A, X)|X\}^{-1} E\{\sigma^{-2}(A, X) d(A, X)|X\}\right]$ and thus $\Lambda^{SR,\perp}_{nuis} + \Lambda^{SR}_{nuis} = L^0_2(P)$. See for example Chamberlain (1987).

Consider now again the original model $\mathcal{A}$ with $q_2(X, A_2)$ and $q_1(X, A_1)$ unrestricted. Consider one-dimensional submodels $q_1(X, A_1; t) = q_1(X, A_1) + tk_1(X, A_1)$, $q_2(X, A_2; t) = q_2(X, A_2) + tk_2(X, A_2)$. Then the score $S_t(\epsilon, A, X)$ for t at the truth $t = 0$ satisfies $E\{S_t(\epsilon, A, X)|A, X\} = 0$ and $E\{\epsilon S_t(\epsilon, A, X)|A, X\} = k_1(X, A_1) + k_2(X, A_2)$ for both $g(x) = x$ and $g(x) = \exp(x)$. Its projection on $\Lambda_{nuis}^{SR,\perp}$ is $\sigma^{-2}(A, X) J(k_1 + k_2)\epsilon$, which can be seen because
$E[\{S_t - \sigma^{-2}(A, X) J(k_1 + k_2)\epsilon\}\sigma^{-2}(A, X) J(d)\epsilon] = 0$. Thus Λ_{nuis} for model $\mathcal{A}$ is

$$\Lambda_{nuis}^{SR} + \left\{\sigma^{-2}(A, X) J(k_1 + k_2)\epsilon; k_1 = k_1(X, A_1), k_2 = k_2(X, A_2) \text{ arbitrary } \right\}$$

It follows that the orthocomplement of the nuisance tangent space in model $\mathcal{A}$ equals the set of functions $[d(A, X) - E\{d(A, X)|X\}]\epsilon$, where $d(A, X)$ is such that for all $k_1(X, A_1), k_2(X, A_2)$:

$$\begin{aligned}
0 &= E\left[[d(A, X) - E\{d(A, X)|X\}]\epsilon\sigma^{-2}(A, X) J(k_1 + k_2)\epsilon\right] \\
&= E[[d(A, X) - E\{d(A, X)|X\}] J(k_1 + k_2)] \\
&= E[d(A, X)\{J(k_1 + k_2) - E[J(k_1 + k_2)|X]\}] \\
&= E[d(A, X)\{k_1(X, A_1) - E[k_1(X, A_1)|X] + k_2(X, A_2) - E[k_2(X, A_2)|X]\}]
\end{aligned}$$

Let first $k_2(A, X) = 0$. Then $d(A, X)$ must be such that for all $k_1(X, A_1)$

$$\begin{aligned}
0 &= E[d(A, X)\{k_1(X, A_1) - E[k_1(X, A_1)|X]\}] \\
\Leftrightarrow 0 &= E[\{d(A, X) - E[d(A, X)|X]\}|A_1, X]
\end{aligned}$$

Reversing the roles of k_1 and k_2, we conclude that the orthocomplement of the nuisance tangent space in model $\mathcal{A}$ is

$$\Lambda_{nuis}^{\perp} = \{\epsilon d(A, X); E\{d(A, X)|A_1, X\} = E\{d(A, X)|A_2, X\} = 0\}$$

A.2 Lemmas 1, 2 and 3

Lemma 1. When $g(x) = \exp(x)$ and A_1 and A_2 are dichotomous, the set of unbiased estimating functions $S(\beta) = s(Y, A, X; \beta)$ for β^* under model $\mathcal{A} \cap \mathcal{M}_a$ with $f(A|X)$ known that have nonzero expected derivative w.r.t. β and finite variance (and thus power against local alternatives) is empty.

Proof. By Theorem 1, equation (14) and the fact that β is a functional of the conditional law $f(Y|A, X)$, the orthocomplement to the nuisance tangent space under model $\mathcal{A} \cap \mathcal{M}_a$ at (β^*, q_2, q_1, h) is

$$\{U(d, v; \beta^*, q_2, q_1, h) = d(X)\Delta(A, X)\{Y \exp[-q_3(\beta^*, X)A_1A_2 - q_2(X)A_2 - q_1(X)A_1 - h(X)] - 1\} + v(A, X); E\{v(A, X)|X\} = 0, d(X) \in R^p\}$$

Suppose an unbiased estimating function $S(\beta)$ existed under model $\mathcal{A} \cap \mathcal{M}_a$ with nonzero expected derivative w.r.t. β. Then $S(\beta^*)$ is an element of the orthocomplement to the nuisance tangent space for β^* at (β^*, q_2, q_1, h) and must equal $U(d(\beta^*), v(\beta^*); \beta^*, q_2, q_1, h)$ w.p.1 for some functions $d(X; \beta^*)$ and $v(A, X; \beta^*)$ with $E\{v(A, X; \beta^*)|X\} = 0$. But, by the unbiasedness of $S(\beta)$, $E_{\beta^*, q_2^*, q_1^*, h^*}\{S(\beta^*)\} = E_{\beta^*, q_2^*, q_1^*, h^*}\{U(d(\beta^*), v(\beta^*); \beta^*, q_2, q_1, h)\} = 0$ for all (q_2^*, q_1^*, h^*). Hence, taking conditional expectations, we conclude that $d(X; \beta^*)$ must satisfy

$$E[d(X; \beta^*)\Delta(A_i, X_i) \exp[\{q_2^*(X) - q_2(X)\}A_2 + \{q_1^*(X) - q_1(X)\}A_1 + h^*(X) - h(X)]] = 0 \quad (1)$$

for all $q_2^*(X), q_1^*(X), h^*(X), q_1(X), q_2(X), h(X)$. We shall now prove that $d(X; \beta^*) = 0$ w.p.1. Equation (1) implies that under the actual data generating process, for

all functions $m_2(X), m_1(X), m_0(X)$,

$$\begin{aligned} 0 &= E\left[d(X;\beta^*)\Delta(A,X)\exp\left\{m_2(X)A_2 + m_1(X)A_1 + m_0(X)\right\}\right] \\ &= E\left[d(X;\beta^*)\left\{\exp\left\{m_2(X) + m_1(X) + m_0(X)\right\} - \exp\left\{m_1(X) + m_0(X)\right\}\right.\right. \\ &\quad \left.\left. - \exp\left\{m_2(X) + m_0(X)\right\} + \exp\left\{m_0(X)\right\}\right\}\right] \end{aligned}$$

We conclude that $d(X;\beta^*) = 0$ w.p.1. Thus $S(\beta) = v(A,X;\beta)$. Hence, for all β, the derivative $\partial E\left\{S(\beta)\right\}/\partial\beta = 0$, proving the lemma.

Lemma 1 implies that, regardless of the the distribution of X, no first order unbiased estimating function $S_i(\beta) = s(Y_i, A_i, X_i; \beta)$ exists. However, in Lemma 2a we prove that, when $g(x) = \exp(x)$ and A_1 and A_2 are dichotomous, there exists a second order unbiased estimating function $S_{2,ij}(\beta)$ depending on two subjects' data with nonzero expected derivative whenever $P(X_i = X_j) > 0$ and $f(A|X)$ is known. Lemmas 2b and 2c extend this result to model $\mathcal{A}$ in which $f(A|X)$ is unknown. However, our interest in this paper is in covariate vectors X with continuously distributed components which implies $P(X_i = X_j) = 0$. In Lemma 3 we argue that no higher-order unbiased estimating function exists for β^* in model $\mathcal{A} \cap \mathcal{M}_a$ with $f(A|X)$ known (much less in model A) when $g(x) = \exp(x)$ and X has continuous components.

Lemma 2a. When $g(x) = \exp(x)$ and A_1 and A_2 are dichotomous, $E\left\{S_{2,ij}(\beta^*)\right\} = 0$ and $E\left\{\partial S_{2,ij}(\beta^*)/\partial\beta\right\} \neq 0$ under model $\mathcal{A} \cap \mathcal{M}_a$ with $f(A|X)$ known, provided $P(X_i = X_j) > 0$ and $P(A = (a,b)|X) > 0$ w.p.1 for all a, b in $\{0,1\}$, where

$$S_{2,ij}(\beta) = \frac{Y_i Y_j I(X_i = X_j)}{f(A_i|X_i) f(A_j|X_j)} (1 - A_{1i}) A_{1j} \left\{\exp\left(-q_3(X_j;\beta)\right) A_{2j} - A_{2i}\right\}$$

with $q_3(X_j;\beta) \equiv q_3((1,1), X_j; \beta)$.

Proof. The result follows immediately from the identities

$$\begin{aligned} & (1-A_{1i})\,A_{1j}\left\{\exp\left(-q_3(X_j;\beta)\right)A_{2j}-A_{2i}\right\} \\ = \;& \exp\left(-q_3(X_j;\beta)\right)(1-A_{1i})\,(1-A_{2i})\,A_{1j}A_{2j}-(1-A_{1i})\,A_{2i}A_{1j}\,(1-A_{2j}) \end{aligned}$$

and

$$\begin{aligned} & E\left(Y_iY_j|A_{1i}=A_{2i}=0,A_{1j}=A_{2j}=1,X_j,X_i=X_j\right)\exp\left(-q_3(X_j;\beta^*)\right) \\ = \;& E\left[Y_iY_j|A_{1i}=A_{2j}=0,A_{1j}=A_{2i}=1,X_j,X_i=X_j\right]. \end{aligned}$$

The proofs of the following Lemmas are similar.

Lemma 2b. Suppose A_1 and A_2 are dichotomous and conditionally independent given X under the true density $f\left(A|X\right)$, which we assume unknown. Let $f^*\left(A|X\right) = f^*\left(A_1|X\right)f^*\left(A_2|X\right)$ be an arbitrary (user supplied) positive density with A_1 and A_2 conditionally independent given X. Define $S^*_{2,ij}\left(\beta\right)$ to be $S_{2,ij}\left(\beta\right)$ as defined above but with $f\left(A|X\right)$ everywhere replaced by the user-supplied density $f^*\left(A|X\right)$. Then, provided $P\left(X_i=X_j\right)>0$, $E\left\{S^*_{2,ij}\left(\beta^*\right)\right\}=0$ and $E\left\{\partial S^*_{2,ij}\left(\beta^*\right)/\partial\beta\right\}\neq 0$ under model $\mathcal{A}$ with $g(x)=\exp(x)$.

Lemma 2c. Define

$$\begin{aligned} & S^\dagger_{4,ijlm}\left(\beta\right)=I\left(X_i=X_j=X_l=X_m\right)\times \\ & \left\{\exp\left(-q_3(X_j;\beta)\right)(1-A_{1i})\,(1-A_{2i})\,A_{1j}A_{2j}-(1-A_{1l})\,A_{2l}A_{1m}\,(1-A_{2m})\right\}, \end{aligned}$$

a function of 4 subjects' data. Then, provided $P\left(X_i=X_j=X_l=X_m\right)>0$, $E\left\{S^\dagger_{4,ijlm}\left(\beta^*\right)\right\}=0$ and $E\left\{\partial S^\dagger_{4,ijlm}\left(\beta^*\right)/\partial\beta\right\}\neq 0$ under model $\mathcal{A}$ with $g(x)=\exp(x)$ and A_1 and A_2 dichotomous.

Lemma 3. When $g(x)=\exp(x)$, A_1 and A_2 are dichotomous, and X has a continuous component whose marginal density is absolutely continuous with

respect to Lesbegue measure, no unbiased estimating function $S_{m,i_1,\ldots,i_m}(\beta)$ of m subjects' data with finite variance and nonzero expected derivative w.r.t. β exists in model $\mathcal{A} \cap \mathcal{M}_a$ with $f(A|X)$ known.

Outline of Proof. The proof depends on methods developed in Robins et al. (2008). Let β^* be the true value of β. By Theorem 6.5 in Robins et al. (2008), we know if $S_{m,i_1,\ldots,i_m}(\beta)$ exists then the mth order U-statistic $S_m(\beta^*)$with (possibly nonsymmetric) kernel $S_{m,i_1,\ldots,i_m}(\beta^*)$ must be orthogonal to the kth order testing nuisance tangent space for all $k \geq m$ in the model $\mathcal{A} \cap \mathcal{M}_a$ with $f(A|X)$ known. [The kth order testing nuisance tangent space is defined in the statement of their Theorem 6.5]. By Lemma 1, we know that if $S_{m,i_1,\ldots,i_m}(\beta)$ exists, then $m > 1$. Thus it suffices to show that, for continuous X, for each $m > 1$, there is no function $S_{m,i_1,\ldots,i_m}(\beta)$ with nonzero expected β-derivative at β^* that is contained in the orthocomplement to the kth order testing tangent space for all $k \geq m$. The proof proceeds by showing that the existence of such a function would imply that the set L_2 of square integrable functions in R^2 contains a Dirac kernel $K(x, x')$ [i.e. a function $K(x, x')$ satisfying $\int K(x, x') g(x') dx' = g(x)$ for all $g(\cdot)$ in L_2 and all x contained in a set with positive Lesbegue measure]. Since it is known a Dirac kernel does not exist in L_2, we arrive at a contradiction.

We only give the proof for the case $m = 2$. The proof for $m > 2$ is similar but the details are tedious.

It suffices to consider the smaller model in which the marginal distribution of (A, X) is known. Without loss of generality assume X is one dimensional with a density w.r.t. to Lesbegue measure. Define $e = Y - \mu, \mu = \exp\left\{\beta^* A_1 A_2 + \sum_{j=0}^{2} A_j q_j(X)\right\}$, where we have defined A_0 to be the constant 1. Let $f(Y, A, X)$ denote the density of an observation $O = (Y, A, X)$ generated under the unknown $\theta = (\beta^*, q_2(\cdot), q_1(\cdot), q_0(\cdot), f(e|A, X))$. To obtain the first and

second order nuisance tangent spaces at θ, it is sufficient to consider parametric submodels

$$f\left(Y_i, A_i, X_i; \left(\tau_s, \tau_t\right), \left(\omega_1, \omega_2\right)\right) = f\left(e_i\left(\tau_s, \tau_t\right) | A_i, X_i; \left(\omega_1, \omega_2\right)\right) f\left(A_i, X_i\right)$$

indexed by parameters $\left(\left(\tau_s, \tau_t\right), \omega_1, \omega_2\right)$ with $(s,t) \in \{0,1,2\}^2$, $e_i\left(\tau_s, \tau_t\right) = Y_i - \mu_i\left(\tau_s, \tau_t\right), \mu\left(\tau_s, \tau_t\right) = \exp\left\{\beta^* A_1 A_2 + \sum_{j=0}^{2} A_j q_j\left(X\right) + \tau_s A_s g_s\left(X\right) + \tau_t A_t g_t\left(X\right)\right\}$ with $g_s\left(X\right), g_t\left(X\right)$ arbitrary, $f\left(e|A_i, X_i; (0,0)\right) = f\left(e|A_i, X_i\right), \int u dF\left(u|A_i, X_i; \left(\omega_1, \omega_2\right)\right) = 0$, but the model $f\left(e_i|A_i, X_i; \omega_1, \omega_2\right)$ otherwise unrestricted.

It is sufficient to consider 2 subjects. Then in the model $f\left(Y, A, X; \left(\tau_s, \tau_t\right), \left(\omega_1, \omega_2\right)\right)$ the first order score for τ_s at $\tau = \left(\tau_s, \tau_t\right) = 0, \omega = \left(\omega_1, \omega_2\right) = 0$ for the two subjects is

$$\sum_{i=1}^{2} S_{s,i}\left(g_s\right) = \sum_{i=1}^{2} \frac{f_{\tau_s}\left(Y_i, A_i, X_i; (0,0)\right)}{f\left(Y_i, A_i, X_i\right)} = \sum_{i=1}^{2} g_s\left(X_i\right) A_{si} \varsigma\left(e_i, A_i, X_i\right) \mu_i,$$

where $\varsigma\left(e, A, X\right) = f_e\left(e|A, X\right) / f\left(e|A, X\right)$ and where, for an arbitrary parameter τ, $f_\tau \equiv \partial f / \partial \tau$. The first order tangent space for $q_s\left(\cdot\right)$ is $\Lambda_{1,s}^{nu} = \left\{\sum_{i=1}^{2} S_{s,i}\left(g\right); g \text{ unrestricted}\right\}$. Throughout we silently understand that each set of random variables has been intersected with $L_2\left(F\right)$. Similarly, the first order tangent space for $f\left(e|A, X\right)$ is

$$\Lambda_{1,e}^{nu} = \left\{\sum_{i=1}^{2} R_i; R = r\left(e, A, X\right) \text{restricted by } E\left(eR|A, X\right) = \mathrm{E}\left(R|A, X\right) = 0\right\}.$$

The first order nuisance tangent space for β^* is $\Lambda_1^{nu} = \cup_{s=1}^{3} \Lambda_{1,s}^{nu} \cup \Lambda_{1,e}^{nu}$. The 1st order efficient score for β^* is $S_{1,eff} = \Pi\left[S_{\beta^*} | \Lambda_1^{nu}\right] = \Pi\left[\sum_{i=1}^{2} \varsigma\left(e_i, A_i, X_i\right) \mu_i A_{1i} A_{2i} | \Lambda_1^{nu}\right] = \sum_{i=1}^{2} E\left[\Delta\left(A, X\right) \mu^{-2} | X_i\right] \Delta\left(A_i, X_i\right) e_i / \mu_i$ where

$$\Delta\left(A, X\right) = \left\{f\left(A|X\right)\right\}^{-1} \left[I\left\{A_1 = A_2\right\} - I\left\{A_1 \neq A_2\right\}\right].$$

In the model $f\left(Y, A, X; \left(\tau_s, \tau_t\right), \left(\omega_1, \omega_2\right)\right)$ the second order score for $\left(\tau_s, \tau_t\right)$ at

$\tau = \omega = 0$ for the 2 subjects is, by definition,

$$\left\{ \frac{\partial^2}{\partial \tau_s \partial \tau_t} \Pi_{i=1}^2 f\left(Y_i, A_i, X_i; (0,0), (0,0)\right) \right\} / \Pi_{i=1}^2 f\left(Y_i, A_i, X_i\right) = \sum_{i=1}^{2} S_{st,i}\left(g_s, g_t\right) + \sum_{i \neq j} S_{s,i}\left(g_s\right) S_{t,j}\left(g_t\right),$$

where

$$\begin{aligned} S_{st,i}\left(g_s, g_t\right) &= f_{\tau_s,\tau_s}\left(Y_i, A_i, X_i; (0,0), (0,0)\right) / f\left(Y_i, A_i, X_i\right) \\ &= \left\{ \varsigma\left(e_i, A_i, X_i\right) \mu_i + \varsigma^*\left(e_i, A_i, X_i\right) \mu_i^2 \right\} g_s\left(X_i\right) A_{si} g_t\left(X_i\right) A_{ti}, \\ \varsigma^*\left(e, A, X\right) &= f_{ee}\left(e|A, X\right) / f\left(e|A, X\right), \end{aligned}$$

and $\sum_{i \neq j} a_i b_j = a_1 b_2 + a_2 b_1$.

The second order tangent space for $\left(q_2\left(\cdot\right), q_1\left(\cdot\right), q_0(\cdot)\right)$ is thus, by definition,

$$\Lambda_{2,q}^{nu} = \cup_{(s,t) \in \{0,1,2\}^2} \left\{ \sum_{i=1}^{2} S_{st,i}\left(g_1, g_2\right) + \sum_{i \neq j} S_{s,i}\left(g_1\right) S_{t,j}\left(g_2\right); g_1, g_2 \text{ unrestricted} \right\}.$$

Similiarly, there is a second order tangent space $\Lambda_{2,e}^{nu}$ for $f\left(e|A, X\right)$ based on all possible second order score for (ω_1, ω_2) at $\tau = \omega = 0$ and a second order mixed score tangent space $\Lambda_{2,eq}^{nu}$based on all possible scores $\{\frac{\partial^2}{\partial \tau_s \partial w_t} \Pi_{i=1}^2 f\left(Y_i, A_i, X_i; (0,0), (0,0)\right)\}$ $/\Pi_{i=1}^2 f\left(Y_i, A_i, X_i\right)$. The second order testing nuisance tangent space Λ_2^{nu} is, by definition, $\Lambda_1^{nu} \cup \Lambda_{2,q}^{nu} \cup \Lambda_{2,e}^{nu} \cup \Lambda_{2,eq}^{nu}$ which can be written $\Lambda_2^{nu} = \Pi\left[\Lambda_{2,q}^{nu} | \Lambda_1^{nu,\perp}\right] \cup \Lambda_1^{nu} \cup \Lambda_{2,e}^{nu} \cup \Lambda_{2,eq}^{nu}$, where

$$\begin{aligned} \Pi\left[\Lambda_{2,q}^{nu} | \Lambda_1^{nu,\perp}\right] &= \left\{ \sum_{i=1}^{2} S_{1,eff,i} g_1\left(X_i\right) g_2\left(X_i\right) \right. \\ &\quad \left. + \sum_{i \neq j} g_1\left(X_i\right) A_{1i} e_i \mu_i e_j \mu_j A_{2j} g_2\left(X_j\right); g_1, g_2 \text{ unrestricted} \right\} \cup \qquad (2) \\ &\left[\cup_{\{s,t\} \in \{0,1,2\}^2 \setminus \{(1,2),(2,1)\}} \left\{ \sum_{i \neq j} g_1\left(X_i\right) A_{si} e_i \mu_i e_j \mu_j A_{tj} g_2\left(X_j\right); g_1, g_2 \text{ unrestricted} \right\} \right], \end{aligned}$$

since $\Pi\left[S_{st,i}\left(g_s,g_t\right)|\Lambda_1^{nu,\perp}\right]=0$ if $\{s,t\}\in\{0,1,2\}^2\setminus\{(1,2),(2,1)\}$, $\Pi\left[S_{st,i}\left(g_s,g_t\right)|\Lambda_1^{nu,\perp}\right]=S_{1,eff,i}g_s\left(X_i\right)g_t\left(X_i\right)$ if $\{s,t\}\in\{(1,2),(2,1)\}$. In these calculations we used the fact that $\mathrm{E}[e\varsigma^*\left(e,A,X\right)|A,X]=\int ef_{ee}\left(e|A,X\right)de=ef_e\left(e|A,X\right)|_{-\infty}^{\infty}-\int f_e\left(e|A,X\right)de=0-0=0$.

The orthocomplement to the 2th order testing tangent space for β^* is defined to be $\Lambda_2^{nu,\perp}$.

Now suppose $S_{2,ij}\left(\beta\right)$ satisfying the conditions of the Lemma existed. We will derive a contradiction. The U-statistic $S_2\left(\beta^*\right)$ with kernel $S_{2,ij}\left(\beta^*\right)$ has a unique Hoeffding representation with $S_2\left(\beta^*\right)=\sum_i Q_{1i}+\sum_{i\neq j}Q_{2ij}$ satisfying $E\left[Q_{2ij}|O_i\right]=E\left[Q_{2ij}|O_j\right]=0$. By $S_{2,ij}\left(\beta\right)$ having nonzero expected β-derivative at β^*, we can use the extended information equality of Theorem 2.2 of Robins et al. (2008) to conclude that $E\left[S_{2,ij}\left(\beta^*\right)S_{1,eff,i}\right]=E\left[Q_{1i}\left(\beta^*\right)S_{1,eff,i}\right]\neq 0$.

Now by the set of equation (2) contained in Λ_2^{nu},we conclude that for all functions g_1,g_2

$$\begin{aligned}0&=E\left[Q_{1i}S_{1,eff,i}g_1\left(X_i\right)g_2\left(X_i\right)\right]+E\left[Q_{2ij}g_1\left(X_i\right)A_{1i}e_i\mu_ie_j\mu_jA_{2j}g_2\left(X_j\right)\right]\\&=E\left[b_1\left(X_i\right)g_1\left(X_i\right)g_2\left(X_i\right)\right]+E\left[b_2\left(X_i,X_j\right)g_1\left(X_i\right)g_2\left(X_j\right)\right]\\&=E\left[\left\{b_1\left(X_i\right)g_2\left(X_i\right)-E\left[b_2\left(X_i,X_j\right)g_2\left(X_j\right)|X_i\right]\right\}g_1\left(X_i\right)\right]\end{aligned}$$

where $b_1\left(X_i\right)=E\left[Q_{1i}S_{1,eff,i}|X_i\right]$ and $b_2\left(X_i,X_j\right)=E\left[Q_{2ij}A_{1i}e_i\mu_ie_j\mu_jA_{2j}|X_i,X_j\right]$. Since $g_1\left(X_i\right)$ is unrestricted, we conclude that

$$b_1\left(X_i\right)g_2\left(X_i\right)-E\left[b_2\left(X_i,X_j\right)g_2\left(X_j\right)|X_i\right]=0$$

almost surely. Since we know that $E\left[Q_{1i}\left(\beta^*\right)S_{1,eff,i}\right]\neq 0$, we have that the set $\mathcal{B}\left(\delta\right)=\{x;|b_1\left(x\right)|>\delta\}$ has positive Lesbegue measure for some $\delta>0$ and on this set $g_2\left(x\right)=\int K\left(x,X\right)g_2\left(X\right)dX$ for all $g_2\left(X\right)$ in L_2, where $K\left(x,X\right)=I\left(x\in\mathcal{B}\left(\delta\right)\right)f\left(x\right)\left\{b_1\left(x\right)\right\}^{-1}b_2\left(x,X\right)$. Further $K\left(x,X\right)$is an element of L_2 w.r.t.

Lesbegue measure on R^2. We conclude that $K(x, X)$ is a Dirac kernel for x in a set of positive Lesbegue measure that is contained in L_2, a contradiction.

A.3 Proof of Theorems 2 and 3

We first prove Theorem 2 by showing that

$$E\{U(\beta^*, \tilde{\gamma}(\beta^*), \tilde{\alpha})\} = 0 \tag{3}$$

under model $\mathcal{B}^{\text{id}}_{\text{cip}}$ where again we use $\tilde{}$ to denote probability limits. First, $\tilde{\gamma}(\beta^*) = \gamma^*$ under model $\mathcal{A}_{\text{cip}} \cap \mathcal{M}_y$ because $G_i(\beta^*, \gamma^*, \alpha)$ has mean zero for each α under this model. Equality (3) now follows because $E(\epsilon(\beta^*, \gamma^*)|A, X) = 0$. Second, $\tilde{\alpha} = \alpha^*$ under model $\mathcal{A}_{\text{cip}} \cap \mathcal{M}_a$. Equality (3) now follows because

$$\begin{aligned} E\{U(\beta^*, \gamma, \alpha^*)\} &= E\left[\left[d(A, X) - E\left\{d(A, X)\,|A_1, X; \alpha_2^*\right\} - E\left\{d(A, X)\,|A_2, X; \alpha_1^*\right\}\right.\right. \\ &\quad \left.\left. + E\left\{d(A, X)\,|X; \alpha^*\right\}\right]\left\{\Delta q_2(X, A_2; \gamma) + \Delta q_1(X, A_1; \gamma) + \Delta h(X; \gamma)\right\}\right] \end{aligned}$$

for functions $\Delta q_j(X, A_j; \gamma), j = 1, 2$ and $\Delta h(X; \gamma)$, which is zero for each γ when $A_1 \amalg A_2|X$. Third, $\tilde{\alpha}_1 = \alpha_1^*$ under model $\mathcal{A}_{\text{cip}} \cap \mathcal{M}_{ya1}$ and $\tilde{\gamma}_1(\beta^*) = \gamma_1^*$ by the fact that $G_{i1}(\beta^*, (\gamma_0, \gamma_1^*, \gamma_2), \alpha_1^*)$ has mean zero for each (γ_0, γ_2) under this model. Equality (3) now follows because

$$\begin{aligned} E\{U(\beta^*, (\gamma_0, \gamma_1^*, \gamma_2), (\alpha_1^*, \alpha_2))\} &= E\left[\left[d(A, X) - E\left\{d(A, X)\,|A_1, X; \alpha_2\right\}\right.\right. \\ &\quad - E\left\{d(A, X)\,|A_2, X; \alpha_1^*\right\} + E\left\{d(A, X)\,|X; (\alpha_1^*, \alpha_2)\right\}\left.\right] \\ &\quad \left.\times\left\{\Delta q_2(X, A_2; \gamma_2) + \Delta h(X; \gamma_0)\right\}\right] \end{aligned}$$

for functions $\Delta q_2(X, A_2; \gamma)$ and $\Delta h(X; \gamma)$, which is zero for each $(\gamma_0, \gamma_2, \alpha_2)$ by the fact that

$$\begin{aligned} &E\left\{d(A, X)\,|A_1, X; \alpha_2\right\} - E\left\{d(A, X)\,|X; (\alpha_1^*, \alpha_2)\right\} \\ &= E\left\{d(A, X)\,|A_1, X; \alpha_2\right\} - E\left[E\left\{d(A, X)\,|A_1, X; \alpha_2\right\}|X; \alpha_1^*\right\} \end{aligned}$$

and because, using $A_1 \amalg A_2 | X$,

$$E\left\{\Delta q_2\left(X, A_2; \gamma_2\right) | A_1, X\right\} = E\left\{\Delta q_2\left(X, A_2; \gamma_2\right) | X\right\}$$

Using similar arguments, one can show that equality (3) holds under model $\mathcal{A}_{\text{cip}} \cap \mathcal{M}_{ya2}$ and under model $\mathcal{B}_{\text{cip}}^{\text{exp}}$.

Assuming that the regularity conditions of Theorem 1A in Robins, Mark and Newey (1992) hold for $U_i(\beta, \gamma, \alpha)$, $G_i(\beta, \gamma, \alpha)$ and $A_i(\alpha)$, it now follows by standard Taylor expansion arguments that

$$\begin{aligned}
0 \;=\; & n^{-1/2} \sum_{i=1}^{n} U_i\left(\beta^*, \tilde{\gamma}(\beta^*), \tilde{\alpha}\right) + \left[E\left\{ \frac{\partial}{\partial \beta} U_i\left(\beta, \tilde{\gamma}(\beta^*), \tilde{\alpha}\right) \right\}_{|\beta=\beta^*} - E\left\{ \frac{\partial}{\partial \gamma} U_i(\beta^*, \gamma, \tilde{\alpha}) \right\}_{|\gamma=\tilde{\gamma}(\beta^*)} \right. \\
& \left. \times E^{-1}\left\{ \frac{\partial}{\partial \gamma} G_i(\beta^*, \gamma, \tilde{\alpha}) \right\}_{|\gamma=\tilde{\gamma}(\beta^*)} E\left\{ \frac{\partial}{\partial \beta} G_i\left(\beta, \tilde{\gamma}(\beta^*), \tilde{\alpha}\right) \right\}_{|\beta=\beta^*} \right] \sqrt{n}(\widehat{\beta} - \beta^*) \\
& - E\left\{ \frac{\partial}{\partial \gamma} U_i(\beta^*, \gamma, \tilde{\alpha}) \right\}_{|\gamma=\tilde{\gamma}(\beta^*)} E^{-1}\left\{ \frac{\partial}{\partial \gamma} G_i(\beta^*, \gamma, \tilde{\alpha}) \right\}_{|\gamma=\tilde{\gamma}(\beta^*)} G_i(\beta^*, \tilde{\gamma}(\beta^*), \tilde{\alpha}) \\
& - \left[E\left\{ \frac{\partial}{\partial \alpha} U_i(\beta^*, \tilde{\gamma}(\beta^*), \alpha) \right\}_{|\alpha=\tilde{\alpha}} - E\left\{ \frac{\partial}{\partial \gamma} U_i(\beta^*, \gamma, \tilde{\alpha}) \right\}_{|\gamma=\tilde{\gamma}(\beta^*)} E^{-1}\left\{ \frac{\partial}{\partial \gamma} G_i(\beta^*, \gamma, \tilde{\alpha}) \right\}_{|\gamma=\tilde{\gamma}(\beta^*)} \right. \\
& \left. \times E\left\{ \frac{\partial}{\partial \alpha} G_i(\beta^*, \tilde{\gamma}(\beta^*), \alpha) \right\}_{|\alpha=\tilde{\alpha}} \right] E^{-1}\left\{ \frac{\partial}{\partial \alpha} A_i(\alpha) \right\}_{|\alpha=\tilde{\alpha}} A_i(\tilde{\alpha}) + o_p(1)
\end{aligned}$$

where $o_p(1)$ denotes a random variable converging to 0 in probability. When

$$\begin{aligned}
& E\left\{ \frac{\partial}{\partial \beta} U_i\left(\beta, \tilde{\gamma}(\beta^*), \tilde{\alpha}\right) \right\}_{|\beta=\beta^*} - E\left\{ \frac{\partial}{\partial \gamma} U_i(\beta^*, \gamma, \tilde{\alpha}) \right\}_{|\gamma=\tilde{\gamma}(\beta^*)} \\
& \times E^{-1}\left\{ \frac{\partial}{\partial \gamma} G_i(\beta^*, \gamma, \tilde{\alpha}) \right\}_{|\gamma=\tilde{\gamma}(\beta^*)} E\left\{ \frac{\partial}{\partial \beta} G_i\left(\beta, \tilde{\gamma}(\beta^*), \tilde{\alpha}\right) \right\}_{|\beta=\beta^*}
\end{aligned}$$

is nonsingular, it now follows that

$$\sqrt{n}(\widehat{\beta} - \beta^*) = \frac{1}{n} \sum_{i=1}^{n} E^{-1}\left\{ \frac{\partial}{\partial \beta} U_i^*(\beta, \tilde{\gamma}(\beta^*), \tilde{\alpha}) \right\}_{|\beta=\beta^*} U_i^*(\beta^*, \tilde{\gamma}(\beta^*), \tilde{\alpha}) + o_p(1) \quad (4)$$

The asymptotic distribution of $\sqrt{n}(\widehat{\beta} - \beta^*)$ under models $\mathcal{B}_{\text{id}}$ and $\mathcal{B}_{\text{exp}}$ follows from the previous equation by Slutsky's Theorem and the Central Limit Theorem. This proves part (i).

At the intersection model $\mathcal{A}_{\text{cip}} \cap \mathcal{M}_y \cap \mathcal{M}_a$, $E\left\{\frac{\partial}{\partial\gamma}U_i(\beta^*,\gamma,\tilde{\alpha})\right\}_{|\gamma=\tilde{\gamma}(\beta^*)} = 0$ and $E\left\{\frac{\partial}{\partial\alpha}U_i(\beta^*,\tilde{\gamma}(\beta^*),\alpha)\right\}_{|\alpha=\tilde{\alpha}} = 0$ and hence $U_i^*(\beta^*,\tilde{\gamma}(\beta^*),\tilde{\alpha}) = U_i(\beta^*,\tilde{\gamma}(\beta^*),\tilde{\alpha})$. It follows that the estimators $\widehat{\beta}(d, G_{(1)}, A_{(1)})$ and $\widehat{\beta}(d, G_{(2)}, A_{(2)})$ have the same influence functions at the intersection model $\mathcal{A}_{\text{cip}} \cap \mathcal{M}_y \cap \mathcal{M}_a$. This proves part (ii).

The proof of Theorem 3 is analogous and omitted here for brevity.

Remark: Relationship to the 'case-only' estimators of Tchetgen Tchetgen and Robins (2009). In the special case of binary Y with support $\{0,1\}$ and g exponential , Tchetgen Tchetgen and Robins (2009) constructed a class of CAN estimators $\left\{\widehat{\beta}_{TR}\right\}$ of the interaction parameter β^* under the union model $\mathcal{B}_{cip}^{TR} = \mathcal{A}_{cip} \cap (\mathcal{M}_1 \cup \mathcal{M}_2)$ that assumes A_1 and A_2 are conditionally independent given X and either a parametric model $\mathcal{M}_1 = \{f_1(A_1|X, Y=1, A_2=0;\omega_1);\ \omega_1 \in R^{\dim(\omega_1)}\}$ for the conditional density of A_1 given $(Y=1, A_2=0)$ or a model $\mathcal{M}_2 = \{f_2(A_2|X,\ Y=1, A_1=0;\omega_2);\ \omega_2 \in R^{\dim(\omega_2)}\}$ for the conditional density of A_2 given $(Y=1, A_2=0)$ is true, with the f_j known functions and the parameters ω_2 and ω_1 variation independent. The estimators $\widehat{\beta}_{TR}$ do not depend on the data from the subjects with $Y=0$ (the non-cases); i.e., they are functions of the case only data $\{(A_i, X_i); Y_i = 1, i \in \{1,...,n\}\}$. The model $\mathcal{A}_{cip} \cap (\mathcal{M}_{ya_1} \cup \mathcal{M}_{ya_2})$ is strictly contained in both model $\mathcal{B}_{cip}^{\exp} = \mathcal{A}_{cip} \cap (\mathcal{M}_{ya_1} \cup \mathcal{M}_{ya_2} \cup \mathcal{M}_y)$ and model $\mathcal{B}_{cip}^{TR}$ [since, owing to the identity

$$f(A_j|Y=1, A_{j'}=0, X) = \frac{\exp\{q_j(A_j,X)\} f(A_j|X)}{\int \exp\{q_j(A_j,X)\} f(A_j|X)\, d\mu(A_j)}$$

under model $\mathcal{A}_{cip}$, parametric models for $f(A_j|X)$ and $q_j(A_j,X)$ determine a parametric model for $f(A_j|Y=1, A_{j'}=0, X), j \neq j'$.]. Thus both $\widehat{\beta}_{cip} = \widehat{\beta}_{cip}(d)$ solving equation (8) and any $\widehat{\beta}_{TR}$ are CAN under model $\mathcal{A}_{cip} \cap (\mathcal{M}_{ya_1} \cup \mathcal{M}_{ya_2})$; however there will exist distributions in $\mathcal{B}_{cip}^{\exp} \backslash \mathcal{B}_{cip}^{TR}$ under which $\widehat{\beta}_{cip}$ is CAN but $\widehat{\beta}_{TR}$ is inconsistent and distributions in $\mathcal{B}_{cip}^{TR} \backslash \mathcal{B}_{cip}^{\exp}$ under which $\widehat{\beta}_{TR}$ is CAN but $\widehat{\beta}_{cip}$ is

inconsistent. Thus neither $\widehat{\beta}_{cip}$ nor $\widehat{\beta}_{TR}$ strictly dominates the other in terms of robustness even though $\widehat{\beta}_{cip}$ is triply robust while $\widehat{\beta}_{TR}$ is only doubly robust.

We next compare $\widehat{\beta}_{cip}$ and $\widehat{\beta}_{TR}$ in terms of cost and statistical efficiency. Use of $\widehat{\beta}_{TR}$ rather than $\widehat{\beta}_{cip}$ can lead to considerable savings in the cost of data collection since data on non-cases need not be obtained. However, a potential disadvantage of $\widehat{\beta}_{TR}$ compared to $\widehat{\beta}_{cip}$ when data on both cases and noncases have been obtained is that, as shown next, if all specified parametric models happen to be correct so that the data are actually generated under the joint intersection submodel $\mathcal{A}_{cip} \cap \mathcal{M}_y \cap \mathcal{M}_a \cap \mathcal{M}_1 \cap \mathcal{M}_2$, then (i) the asymptotic variance of the optimal estimator $\widehat{\beta}_{TR}$ in the class exceeds that of $\widehat{\beta}_{cip}\left(d_{opt}\left(\widehat{\alpha},\widehat{\eta}\right)\right)$ of Section 4.2 and (ii) the semiparametric variance bound for models $\mathcal{A}_{cip}, \mathcal{B}_{cip}^{\exp}$, and $\mathcal{B}_{cip}^{TR}$ are equal to one another and to the asymptotic variance of $\widehat{\beta}_{cip}\left(d_{opt}\left(\widehat{\alpha},\widehat{\eta}\right)\right)$. However, we show below that under the joint intersection submodel, the relative efficiency of the optimal estimator in the class $\left\{\widehat{\beta}_{TR}\right\}$ compared to $\widehat{\beta}_{cip}\left(d_{opt}\left(\widehat{\alpha},\widehat{\eta}\right)\right)$ is close to 100% whenever $E\left[Y|A,X\right]$ is small with high probability (i.e. under the rare disease assumption).

Results (i) and (ii) follows from the fact that, like the class $\left\{\widehat{\beta}_{cip}\left(d\right)\right\}$, the estimating equation solved by the class $\left\{\widehat{\beta}_{TR}\right\}$ is an (estimated) subset of the orthocomplement to the nuisance tangent space for model $\mathcal{A}_{cip}$. But, in contrast to the class $\left\{\widehat{\beta}_{cip}\left(d\right)\right\}$, the subset of the orthocomplement solved by $\left\{\widehat{\beta}_{TR}\right\}$ does not contain the efficient score S_{eff}. Specifically, with D the set defined by equation (5), the orthocomplement $\left\{d\left(A,X\right)\epsilon + \widetilde{d}\left(A,X\right); d \in \mathbf{D}, \widetilde{d} \in \mathbf{D}\right\}$ to the nuisance tangent space for β^* under model $\mathcal{A}_{\text{cip}}$ is the direct sum of the orthocomplement $\left\{d\left(A,X\right)\epsilon; d \in \mathbf{D}\right\}$ under model $\mathcal{A}$ and the set $\left\{\widetilde{d}\left(A,X\right); \widetilde{d} \in \mathbf{D}\right\}$. As a consequence, when g is exponential, the elements $d\left(A,X\right)\left\{\epsilon+1\right\} \equiv d\left(A,X\right)Y/E\left[Y|A,X\right]$ of the set $CO = \left\{d\left(A,X\right)\left\{\epsilon+1\right\}; d \in \mathbf{D}\right\}$ are contained in the orthocomplement

under $\mathcal{A}_{\text{cip}}$, depend on the case-only (CO) data, and do not contain the efficient score $d_{opt}(A,X)\,\epsilon$. The approach of Tchetgen Tchetgen and Robins is based on the identity $d(A,X)\,Y/E[Y|A,X] = Y\,d(A,X)\,f(A_2|X)\,f(A_1|X)\,/\{f(A|X,Y=1)\,E(Y=1|X)\}$ so with $d(A,X;d^*)$

$\equiv d^*(A,X) - \sum_{j=1}^{2}\int d^*(A,X)\,f(A_j|X)\,d\mu(A_j) + \int d^*(A,X)\,f(A_2|X)\,f(A_1|X)\,d\mu(A_2,A_1)$

$CO = \left\{ Z(d^*) = \frac{Y d(A,X;d^*) f(A_1|X) f(A_2|X)}{f(A|X,\mathbf{Y}=\mathbf{1})} \right\}$ as $d^*(A,X)$ varies freely.

Under model $\mathcal{A}_{\text{cip}} \cap M_1 \cap M_2$, the parameters $(\beta,\omega_1,\omega_2)$ completely specify the density of A given $(X,Y=1)$, which we will denote by $f(A|X,Y=1;\beta,\omega_1,\omega_2)$. Similarly under model $\mathcal{A}_{\text{cip}} \cap \mathcal{M}_j$, (β,ω_j) completely specify the density of A_j given $(X,Y=1,A_{j'})$, $j \neq j'$, which we denote by $f(A_j|Y=1,A_{j'},X;\beta,\omega_j)$. The estimator $\widehat{\beta}_{TR} = \widehat{\beta}_{TR}\left(d^*, f^\dagger\right)$ is defined to be the solution to $\sum_i Z_i\left(d^*;\beta,\widehat{\omega}_1(\beta),\widehat{\omega}_2(\beta),f^\dagger\right) = 0$ where $\widehat{\omega}_j(\beta)$ is the maximizer of $\prod_i f(A_{ji}|Y_i=1,A_{j'i},X_i;\beta,\omega_j)$ and $Z\left(d^*;\beta,\omega_1,\omega_2,f^\dagger\right)$ is defined like $Z(d^*)$ except with $f(A|X,Y=1;\beta,\omega_1,\omega_2)$ replacing the true density $f(A|X,Y=1)$ and $d\left(A,X;d^*,f^\dagger\right)$ replacing $d(A,X;d^*)$ with $d\left(A,X;d^*,f^\dagger\right)$ defined as $d(A,X;d^*)$ except with arbitrary user-supplied densities $f_j^\dagger(A_j|X)$, $j=1,2$ replacing the true densities $f(A_j|X)$. The estimator $\widehat{\beta}_{TR}\left(d^*,f^\dagger\right)$ is CAN under model $\mathcal{B}_{cip}^{TR}$ because $Z\left(d^*;\beta^*,\omega_1^*,\omega_2^*,f^\dagger\right)$ has mean zero under model $\mathcal{A}_{\text{cip}} \cap \mathcal{M}_j$ when the true value of the parameters are $\left(\beta^*,\omega_j^*\right)$ even when model $\mathcal{M}_{j'}$ is misspecified, $\omega_{j'}^*$ is arbitrary, and $f^\dagger$ did not generate the data.

Under the the joint intersection submodel, the relative efficiency of $\widehat{\beta}_{TR}\left(d^*,f^\dagger\right)$ and $\widehat{\beta}_{cip}(d_{opt})$ when β^* is one dimensional, $f^\dagger$ is the true density, and $d\left(A,X;d^*,f^\dagger\right) = d(A,X;d^*)$ equals $E(Y=1|X)d_{opt}(A,X)$, is $E[\{d_{opt}(A,X)\}^2\left\{\{E[Y|A,X]\}^{-1}-1\right\}]$ $/E[\{E[Y|A,X]\}^{-1}\{d_{opt}(A,X)\}^2]$ which is nearly one whenever $E[Y|A,X]$ is small with high probability. The expression follows from the fact that the ratio of asymptotic variances is $\text{Var}\{d_{opt}(A,X)\,\epsilon\}\ /\text{Var}\{d_{opt}(A,X)\,Y\ /E[Y|A,X]\}$, that $\text{Var}\{d_{opt}(A,X)\,\epsilon/E[Y|A,X]\} = E\left[\{d_{opt}(A,X)\}^2\left\{\{E[Y|A,X|]\}^{-1}-1\right\}\right]$, and

that $\mathrm{Var}\left\{d_{opt}\left(A,X\right)Y/E\left[Y|A,X\right]\right\} = \mathrm{Var}\left\{d_{opt}\left(A,X\right)\epsilon/E\left[Y|A,X\right]\right\}+E\left[\left\{d_{opt}\left(A,X\right)\right\}^2\right]$
$= E\left[\left\{E\left[Y\left|A,X\right|\right]\right\}^{-1}\left\{d_{opt}\left(A,X\right)\right\}^2\right].$

A.4 Proof of Theorem 4

Part (i). The efficient score S_{eff} for β^* under model $\mathcal{A}$ is the projection $S_{eff} = \Pi\left[S_\beta|\Lambda^\perp_{nuis}\right]$ of the score S_β for β^* onto $\Lambda^\perp_{nuis}$ (as defined in the proof of Theorem 1) in the Hilbert space with covariance inner product. Now $S_{eff} = \Pi\left[S_\beta|\Lambda^\perp_{nuis}\right] = \Pi\left\{\Pi\left[S_\beta|\Lambda^{SR,\perp}_{nuis}\right]|\Lambda^\perp_{nuis}\right\}$ since $\Lambda^\perp_{nuis} \subset \Lambda^{SR,\perp}_{nuis}$, where

$$\Pi\left[S_\beta|\Lambda^{SR,\perp}_{nuis}\right] = \left[\frac{\partial q_3(A,X;\beta)}{\partial\beta} - \frac{E\left\{\sigma^{-2}\left(A,X\right)\frac{\partial q_3(A,X;\beta)}{\partial\beta}|X\right\}}{E\left\{\sigma^{-2}\left(A,X\right)|X\right\}}\right]\sigma^{-2}(A,X)\epsilon$$

evaluated at β^*, since $E(S_\beta|A,X) = 0, E(S_\beta\epsilon|A,X) = \partial q_3(A,X;\beta)/\partial\beta$ for both $g\left(x\right) = x$ and $g\left(x\right) = \exp\left(x\right)$. Let us further write $\Lambda^\perp_{nuis} = \Lambda^\perp_1 \cap \Lambda^\perp_2$, where $\Lambda^\perp_j = \{d(A,X)\epsilon : E\{d(A,X)|A_j,X\} = 0\}$ for $j = 1,2$. Then Von Neumann's theorem (Bickel et al., 1993) shows that the projection onto $\Lambda^\perp_{nuis}$ can be obtained by repeatedly projecting onto $\Lambda^\perp_1$ and $\Lambda^\perp_2$ until convergence. The projection $\Pi\left[d^*(A,X)\epsilon|\Lambda^\perp_j\right]$ of any $d^*(A,X)\epsilon$ on $\Lambda^\perp_j$ is

$$\Pi\left[d^*(A,X)\epsilon|\Lambda^\perp_j\right] = \left\{d^*(A,X) - \frac{\sigma^{-2}(A,X)E\left\{d^*(A,X)|A_j,X\right\}}{E\left\{\sigma^{-2}(A,X)|A_j,X\right\}}\right\}\epsilon$$

since for all $d(A,X)\epsilon \in \Lambda^\perp_j$

$$\begin{aligned}
0 &= E\left[\left\{d^*(A,X)\epsilon - \Pi\left[d^*(A,X)\epsilon|\Lambda^\perp_j\right]\right\}d(A,X)\epsilon\right] \\
&= E\left[\frac{\sigma^{-2}(A,X)E\left\{d^*(A,X)|A_j,X\right\}}{E\left\{\sigma^{-2}(A,X)|A_j,X\right\}}d(A,X)\sigma^2(A,X)\right] \\
&= E\left[\frac{E\left\{d^*(A,X)|A_j,X\right\}}{E\left\{\sigma^{-2}(A,X)|A_j,X\right\}}E\left[d(A,X)|A_j,X\right]\right] = 0
\end{aligned}$$

Initiating the repeated projections with $d^*(A,X)\epsilon = \frac{\partial q_3(A,X;\beta)}{\partial \beta}_{|\beta=\beta^*}\sigma^{-2}(A,X)\epsilon$ proves Part (i.2), upon noting that the projection of $E\left\{\sigma^{-2}(A,X)\frac{\partial q_3(A,X;\beta)}{\partial \beta}|X\right\}$ $\times\sigma^{-2}(A,X)\epsilon/E\left\{\sigma^{-2}(A,X)|X\right\}$ onto $\Lambda_j^{\perp}$ is zero for $j=1,2$.

Part (i.1) is proved by explicitly verifying that $d^{\dagger}_{opt}(X_i)\Delta(A_i,X_i)\epsilon_i$ is $\Pi\left\{\Pi\left[S_\beta|\Lambda^{SR,\perp}_{nuis}\right]|\Lambda^{\perp}_{nuis}\right\}$ with $\Pi\left[S_\beta|\Lambda^{SR,\perp}_{nuis}\right]$ as given above and with $\Lambda^{\perp}_{nuis} = \{d(X)\Delta(A,X)\epsilon : d(X) \text{ arbitrary}\}$. This is immediate upon noting that

$$0 \;=\; E\left[\left\{\Pi\left[S_\beta|\Lambda^{SR,\perp}_{nuis}\right] - d^{\dagger}_{opt}(X)\Delta(A,X)\epsilon\right\}\Delta(A,X)\epsilon|X\right]$$

Part (ii). That $\widehat{\beta}\left(d\left(\widehat{\alpha}\right)\right)$ and $\widehat{\beta}(d_{\text{opt}}\left(\widehat{\alpha},\widehat{\eta}\right))$ are RAL estimators in models $\mathcal{B}^{\text{id}}$ or $\mathcal{B}^{\text{exp}}$ follows from Theorem 3. The last part of the theorem holds by the results of Robins and Rotnitzky (2001) as discussed in the main text just below the theorem.

Extension: In fact we now show we can actually obtain a closed form expression for S_{eff} for β^* under model $\mathcal{A}$ whenever A_1 or A_2 has finite support. In the following we assume A_1 has finite support $\{\varrho_1,...,\varrho_R\}$.

Consider the Hilbert space $\mathcal{H}$ of functions of (A,X) with inner product given by the inverse variance weighted expectation $E\left(\sigma^{-2}(A,X)b_1\left(A,X\right)b_2\left(A,X\right)\right)$, for all $b_1,b_2\in\mathcal{H}$. Let $\mathcal{C}_0=\{b_0\left(X\right)\}\cap\mathcal{H}$, $\mathcal{C}_1=\{b_1\left(A_1,X\right)\}\cap\mathcal{H}$, $\mathcal{C}_2=\{b_2\left(A_2,X\right)\}\cap\mathcal{H}$. Then, let $\mathcal{K}\left(D\right)=\Pi_{\mathcal{H}}\left(D|\left(\mathcal{C}_0\cup\ \mathcal{C}_1\cup\mathcal{C}_2\right)^{\perp}\right)$ be the projection of $D\in\mathcal{H}$ onto $\left(\mathcal{C}_0\cup\ \mathcal{C}_1\cup\mathcal{C}_2\right)^{\perp}$ in $\mathcal{H}$. We shall need a closed form expression for this projection. To do so we first orthogonalize $\mathcal{C}_0\cup\mathcal{C}_1\cup\mathcal{C}_2$ as $\mathcal{C}_0+\mathcal{C}_1^*+\mathcal{C}_2^*$, where $\mathcal{C}_1^*=\Pi_{\mathcal{H}}\left(\mathcal{C}_1|\mathcal{C}_2^{\perp}\right)$, $\mathcal{C}_2^*=\Pi_{\mathcal{H}}\left(\mathcal{C}_2|\mathcal{C}_0^{\perp}\right)$; a straightforward calculation shows

$$\begin{aligned}
\Pi_{\mathcal{H}}\left(\mathcal{C}_1|\mathcal{C}_2^{\perp}\right) &= \left\{\begin{array}{c} h\left(X\right)\widetilde{A}_1 : \widetilde{A}_1=\left(A_1-\frac{E(A_1\sigma^{-2}(A,X)|A_2,X)}{E(\sigma^{-2}(A,X)|A_2,X)}\right), \\ h(X) \text{ arbirtary p}\times(R-1) \text{ functions}\end{array}\right\}\cap\mathcal{H}, \text{ and}\\
\Pi_{\mathcal{H}}\left(\mathcal{C}_2|\mathcal{C}_0^{\perp}\right) &= \{J\left(B_2\right) : \text{arbirtrary } B_2=b_2\left(A_2,X\right)\}\cap\mathcal{H},
\end{aligned}$$

where $J(\cdot)$ is defined in section A.1, and that $\Pi_{\mathcal{H}}\left(\mathcal{C}_1|\mathcal{C}_2^{\perp}\right)$ and $\mathcal{C}_0$ are orthogonal in $\mathcal{H}$. Therefore, $\mathcal{K}(D) = \Pi_{\mathcal{H}}\left(D|\mathcal{C}_0^{\perp}\right) - \Pi_{\mathcal{H}}(D|\mathcal{C}_1^*) - \Pi_{\mathcal{H}}(D|\mathcal{C}_2^*)$, where $\Pi_{\mathcal{H}}\left(D|\mathcal{C}_0^{\perp}\right) = J(D)$, $\Pi_{\mathcal{H}}(D|\mathcal{C}_1^*) = E\left(\sigma^{-2}(A,X)D\widetilde{A}_1^T|X\right)E\left(\sigma^{-2}(A,X)\widetilde{A}_1\widetilde{A}_1^T|X\right)^{-1}\widetilde{A}_1$ and finally $\Pi_{\mathcal{H}}(D|\mathcal{C}_2^*) = J\left(\frac{E(D\sigma^{-2}(A,X)|A_2,X)}{E(\sigma^{-2}(A,X)|A_2,X)}\right)$. Now it follows essentially from its definition that efficient score for β is given by

$$S_{eff} = \Pi\left(\Pi\left(S_\beta|\left(\Lambda_{nuis}^{rm}\right)^{\perp}\right)|\left\{\begin{array}{c}(b_0(X)+b_1^*(A_1,X)+b_2^*(A_2,X))\sigma^{-2}(A,X)\epsilon \\ : b_0\in\mathcal{C}_0, b_1\in\mathcal{C}_1, b_2\in\mathcal{C}_2\end{array}\right\}^{\perp}\right)$$

where $\Pi(\cdot|\cdot)$ is the orthogonal projection operator in $L_2(P)$, and $\left(\Lambda_{nuis}^{rm}\right)^{\perp}$ is the orthogonal complement to the nuisance tangent space in a restricted mean model where q_1, q_2 and $h(X)$ are assumed known; it is well known that $\Pi\left(S_\beta|\left(\Lambda_{nuis}^{rm}\right)^{\perp}\right) = \frac{\partial}{\partial\beta}q_3(A,X;\beta)|_{\beta=\beta^*}\sigma^{-2}(A,X)\epsilon$, so that we have

$$\begin{aligned}
S_{\beta,eff} &= \Pi\left(\frac{\partial}{\partial\beta}q_3(A,X;\beta)|_{\beta=\beta^*}\sigma^{-2}(A,X)\epsilon|\left\{\begin{array}{c}(b_0(X)+b_1^*(A_1,X)+b_2^*(A_2,X))\sigma^{-2}(A,X)\epsilon \\ : b_0\in\mathcal{C}_0, b_1\in\mathcal{C}_1, b_2\in\mathcal{C}_2\end{array}\right\}^{\perp}\right) \\
&= \left[\Pi_{\mathcal{H}}\left(\frac{\partial}{\partial\beta}q_3(A,X;\beta)|_{\beta=\beta^*}|\left\{\begin{array}{c}(b_0(X)+b_1^*(A_1,X)+b_2^*(A_2,X)) \\ : b_0\in\mathcal{C}_0, b_1\in\mathcal{C}_1, b_2\in\mathcal{C}_2\end{array}\right\}^{\perp}\right)\right]\sigma^{-2}(A,X)\epsilon \\
&= \left[\begin{array}{c}J\left(\frac{\partial}{\partial\beta}q_3(A,X;\beta)|_{\beta=\beta^*}\right) \\ -E\left(\sigma^{-2}(A,X)\frac{\partial}{\partial\beta}q_3(A,X;\beta)|_{\beta=\beta^*}\widetilde{A}_1^T|X\right)E\left(\sigma^{-2}(A,X)\widetilde{A}_1\widetilde{A}_1^T|X\right)^{-1}\widetilde{A}_1 \\ -J\left(\frac{E(\frac{\partial}{\partial\beta}q_3(A,X;\beta)|_{\beta=\beta^*}\sigma^{-2}(A,X)|A_2,X)}{E(\sigma^{-2}(A,X)|A_2,X)}\right)\end{array}\right]\sigma^{-2}(A,X)\epsilon,
\end{aligned}$$

the desired closed form expression.

A.5 Derivation of equation (23)

The proposition is valid when model $\mathcal{M}_y$ holds because then the lefthand side of equation (23) in the article is zero and so is the righthand side by the fact that $\tilde{\gamma}(\beta^*) = \gamma^*$ and $E\{\epsilon(\beta^*,\gamma^*)|A,X\} = 0$ under $\mathcal{M}_y$. Suppose now that model

$\mathcal{M}_a$ holds and that $g(x) = x$. Define $S_2(\alpha) = \partial \log f(A_2|A_1, X; \alpha)/\partial\alpha$. Under weak regularity conditions (for interchanging the integral and derivative), it then follows from equation (5) in the article that

$$\begin{aligned} E\left\{\frac{\partial}{\partial\alpha}d(A, X; \alpha)|A_1, X; \alpha\right\} &= -E\left\{d(A, X; \alpha)S_2(\alpha)|A_1, X; \alpha\right\} \\ &= -E\left[d(A, X; \alpha)\left\{S(\alpha) - E[S(\alpha)|A_1, X]\right\}|A_1, X; \alpha\right\} \\ &= -E\left[d(A, X; \alpha)S(\alpha)|A_1, X; \alpha\right\} \end{aligned}$$

where $S(\alpha) = \partial \log f(A|X; \alpha)/\partial\alpha$, and likewise that

$$E\left\{\frac{\partial}{\partial\alpha}d(A, X; \alpha)|A_2, X; \alpha\right\} = -E\left[d(A, X; \alpha)S(\alpha)|A_2, X; \alpha\right\}.$$

Suppose that $E(Y|A, X) = q_3\left(A, X; \beta^*\right)A_1A_2 + q_2^*\left(X, A_2\right) + q_1^*\left(X, A_1\right) + h^*(X)$. Then, because $\tilde{\alpha} = \alpha^*$ under $\mathcal{M}_a$,

$$\begin{aligned} E\left\{\frac{\partial}{\partial\alpha}U_i^*(\beta^*, \tilde{\gamma}(\beta^*), \alpha)\right\}_{|\alpha=\tilde{\alpha}} &= E\left\{\frac{\partial}{\partial\alpha}d(A, X; \alpha)\epsilon(\beta^*, \tilde{\gamma}(\beta^*))\right\}_{|\alpha=\alpha^*} \\ &= E\left\{\frac{\partial}{\partial\alpha}d(A, X; \alpha)\left[q_2^*\left(X, A_2\right) - q_2\left(X, A_2; \tilde{\gamma}(\beta^*)\right)\right.\right. \\ &\qquad \left.\left.+q_1^*\left(X, A_1\right) - q_1\left(X, A_1; \tilde{\gamma}(\beta^*)\right) + h^*(X) - h(X; \tilde{\gamma}(\beta^*))\right]\right\}_{|\alpha=\alpha^*} \\ &= -E\left\{d(A, X; \alpha^*)S(\alpha^*)\left[q_2^*\left(X, A_2\right) - q_2\left(X, A_2; \tilde{\gamma}(\beta^*)\right)\right.\right. \\ &\qquad \left.\left.+q_1^*\left(X, A_1\right) - q_1\left(X, A_1; \tilde{\gamma}(\beta^*)\right) + h^*(X) - h(X; \tilde{\gamma}(\beta^*))\right]\right\} \\ &= -E\left\{d(A, X; \alpha^*)S(\alpha^*)\epsilon(\beta^*, \tilde{\gamma}(\beta^*))\right\} \end{aligned}$$

Suppose now that model $\mathcal{M}_{ya2}$ holds and that $g(x) = x$. Define $S_1(\alpha) = \partial \log f(A_1|X; \alpha)/\partial\alpha$ where the law $f(A_1|X; \alpha)$ may be misspecified. Then, with $\alpha = (\alpha_1', \alpha_2')'$, we have that

$$\begin{aligned} E\left\{\frac{\partial}{\partial\alpha}d(A, X; \alpha)|A_1, X; \alpha_2\right\} &= -E\left\{d(A, X; \alpha)S_2(\alpha)|A_1, X; \alpha_2\right\} \\ &= -E\left[d(A, X; \alpha)\left\{S(\alpha) - S_1(\alpha)\right\}|A_1, X; \alpha_2\right\} \\ &= -E\left[d(A, X; \alpha)S(\alpha)|A_1, X; \alpha_2\right\} \end{aligned}$$

by the fact that $S_1(\alpha)$ is a function of only A_1 and X. It follows, because $\tilde{\alpha}_2 = \alpha_2^*$ under $\mathcal{M}_{ya2}$, that

$$\begin{aligned}
E\left\{\frac{\partial}{\partial\alpha}U_i^*(\beta^*,\tilde{\gamma}(\beta^*),\alpha)\right\}_{|\alpha=\tilde{\alpha}} &= E\left\{\frac{\partial}{\partial\alpha}d(A,X;\alpha)\epsilon(\beta^*,\tilde{\gamma}(\beta^*))\right\}_{|\alpha=\tilde{\alpha}} \\
&= E\left\{\frac{\partial}{\partial\alpha}d(A,X;\alpha)\left[q_1^*(X,A_1) - q_1(X,A_1;\tilde{\gamma}(\beta^*)) + h^*(X) - h(X;\tilde{\gamma}(\beta^*))\right]\right\}_{|\alpha=\tilde{\alpha}} \\
&= -E\left\{d(A,X;\tilde{\alpha})S(\tilde{\alpha})\left[q_1^*(X,A_1) - q_1(X,A_1;\tilde{\gamma}(\beta^*)) + h^*(X) - h(X;\tilde{\gamma}(\beta^*))\right]\right\} \\
&= -E\left\{d(A,X;\tilde{\alpha})S(\tilde{\alpha})\epsilon(\beta^*,\tilde{\gamma}(\beta^*))\right\}
\end{aligned}$$

It is immediate that this result continues to hold when $g(x) = \exp(x)$.

Finally, suppose that model $\mathcal{M}_{ya2}^*$ holds and that $g(x) = \exp(x)$. Then, because in particular model $\mathcal{M}_a$ holds, we have for $j = 1, 2$ that

$$E\left\{\frac{\partial}{\partial\alpha}d(A,X;\alpha)|A_j,X;\alpha\right\} = -E\left[d(A,X;\alpha)S(\alpha)|A_j,X;\alpha\right\}.$$

It follows, because $\tilde{\alpha} = \alpha^*$ under $\mathcal{M}_{ya2}^*$, that

$$\begin{aligned}
E\left\{\frac{\partial}{\partial\alpha}U_i^*(\beta^*,\tilde{\gamma}(\beta^*),\alpha)\right\}_{|\alpha=\tilde{\alpha}} &= E\left\{\frac{\partial}{\partial\alpha}d(A,X;\alpha)\epsilon(\beta^*,\tilde{\gamma}(\beta^*))\right\}_{|\alpha=\tilde{\alpha}} \\
&= E\left\{\frac{\partial}{\partial\alpha}d(A,X;\alpha)\left(\exp\left[q_1^*(X,A_1) - q_1(X,A_1;\tilde{\gamma}(\beta^*)) + h^*(X) - h(X;\tilde{\gamma}(\beta^*))\right] - 1\right)\right\}_{|\alpha=\alpha^*} \\
&= -E\left\{d(A,X;\alpha^*)S(\alpha^*)\left(\exp\left[q_1^*(X,A_1) - q_1(X,A_1;\tilde{\gamma}(\beta^*)) + h^*(X) - h(X;\tilde{\gamma}(\beta^*))\right] - 1\right)\right\} \\
&= -E\left\{d(A,X;\alpha^*)S(\alpha^*)\epsilon(\beta^*,\tilde{\gamma}(\beta^*))\right\}
\end{aligned}$$

A.6 Relation beween statistical interactions and sufficient cause interactions

Under the nonparametric structural equation model of Pearl (2000) whenever there exists a binary outcome D with two binary causes A_1 and A_2, there always

exists latent binary random variables U_0, U_1, U_2, U_3, U_4, U_5, U_6, U_7, U_8 which are not affected by A_1 and A_2 such that D is the deterministic function

$$\begin{aligned} D &= U_0 \bigvee U_1 A_1 \bigvee U_2 \overline{A_1} \bigvee U_3 A_2 \bigvee U_4 \overline{A_2} \bigvee U_5 A_1 A_2 \bigvee U_6 \overline{A_1} A_2 \\ &\quad \bigvee U_7 A_1 \overline{A_2} \bigvee U_8 \overline{A}_1 \overline{A}_2 \end{aligned} \tag{5}$$

where $\bigvee$ is the Boolean OR operator (VanderWeele and Robins 2007, 2008). Equation (5) is referred to as a sufficient cause representation of D by the vector $U = (U_0, U_1, U_2, U_3, U_4, U_5, U_6, U_7, U_8)$ and the 9 conjunctions occurring in (5) are referred to as the representation's sufficient causes. The vector U will not in general be unique (VanderWeele and Robins 2007, 2008), in the sense that (5) will also hold with U replaced by another vector of binary variables U^*, with U and U^* having different distributions conditional on the measured variables $(A = (A_1, A_2), X)$, where X denotes some other measured variables not caused by A. However, certain sufficient cause representations can be empirically excluded, the following lemma being one example.

Supplementary Lemma 1. Define $Y = 1 - D$. Then, if $\beta^* \neq 0$ in model $\mathcal{A}$ with g the exponential function (i.e., there is a multiplicative interaction between A_1 and A_2 with $Y = 1 - D$) then there cannot exist a sufficient cause representation by a vector U satisfying

(i) $U_5 = U_6 = U_7 = U_8 = 0$ almost surely and

(ii) conditional on X, the variables U_0, U_1, U_2, U_3, U_4,and A are jointly independent.

Before proving the lemma we discuss its interpretation. A sufficient cause representation in which (i) holds is said to exhibit *no sufficient cause interaction between A_1 or its complement and A_2 or its complement,* because D can be

writttten

$$D = U_0 \bigvee U_1 A_1 \bigvee U_2 \overline{A_1} \bigvee U_3 A_2 \bigvee U_4 \overline{A_2}. \tag{6}$$

Suppose a scientist believed a scientific theory that the actual (physical) co-causes (if any) C_1, C_2, C_3, C_4 needed for $A_1, \overline{A_1}, A_2$, and $\overline{A_2}$ respectively to cause D and the background causes C_0 were distributed independently of one another and of $\mathbf{A}$ within levels of X. One would then be principally interested in a sufficient cause representation in which U_0, U_1, U_2, U_3, U_4 stood for C_0, C_1, C_2, C_3, C_4. Suppose the scientist then wanted to test the hypothesis that (6) held because he hypothesized that there did not exist a physical causal mechanism that required the prescence of A_1 *(or its complement) and* A_2 *(or its complement)* to cause D. Then, according to the Lemma, he could test the hypothesis by testing whether $\beta^* = 0$ in model $\mathcal{A}$ with $Y = 1 - D$ and g the exponential function. Of course it may be exceedingly rare for a scientist to believe that the physical cocauses $(C_0, C_1, C_2, C_3, C_4)$ were jointly independent given X and thus that (ii) held, so the practical utility of the result may be small.

Results contained in the following lemma concerning model $\mathcal{A}$ with an additive link tend to be more useful than those in the previous lemma for testing for *sufficient cause interactions*.

Supplementary Lemma 2 (VanderWeele and Robins). Define $Y = D$. If $\beta^* > 0$ in model $\mathcal{A}$ with g the identity function (i.e., there is a additive interaction between A_1 and A_2 with $Y = D$) then there cannot exist a sufficient cause representation by any vector U satisfying

(i')$U_2 = U_4 = U_6 = U_7 = U_8 = 0$ almost surely,

(ii') $U_5 = 0$, almost surely

(iii') conditional on X, (U_0, U_1, U_3) is independent of A.

Suppose our scientist now believed A_1 and A_2 never prevent disease so that with U equal to the physical co-causes C, we can modify (5) and (6) by removing terms containing either $\overline{A}_1$ or $\overline{A}_2$, which we do mathematically by imposing (i'). Suppose the scientist also accepts hypothesis (iii'), which is the assumption that within levels of X, there is no confounding for the effect of A on D. Finally, suppose the scientist again wanted to test the hypothesis of *no sufficient cause interaction between* A_1 *and* A_2 encoded in the (now modified) equation (6) with U being the physical co-causes C. Given (i'), this hypothesis is equivalent to (ii'). He could therefore test the hypothesis by estimating model $\mathcal{A}$ with g the identity function. If he concludes $\beta^* > 0$, the hypothesis of *no sufficient cause interaction is rejected. As* it would not be unusual for a scientist to believe A_1 and A_2 never prevent disease and that A is unconfounded, the practical utility of the second Lemma may be considerable. Note in particular that assumption (iii') does not impose the often unrealistic assumption that U_0, U_1, U_3 are jointly independent given X.

Proof of Supplementary Lemma 1. By contradiction. Assume (i) and (ii) hold. Define $P(U_0|X = x) = a_0^x$, $P(U_1|X = x) = a_1^x$, $P(U_2|X = x) = a_2^x$, $P(U_3|X = x) = a_3^x$ and $P(U_4|X = x) = a_4^x$. We then have by (i) and (ii) that

$$\begin{aligned} 1 - E(D|X = x, A_1 = 0, A_2 = 0) &= \{1 - E(U_0 \bigvee U_2 \bigvee U_4|X = x)\} \\ &= E(\overline{U}_0\overline{U}_2\overline{U}_4|X = x) \\ &= E(\overline{U}_0|X = x)E(\overline{U}_2|X = x)E(\overline{U}_4|X = x) \\ &= (1 - a_0^x)(1 - a_2^x)(1 - a_4^x) \end{aligned}$$

and

$$\begin{aligned} 1 - E(D|X = x, A_1 = 0, A_2 = 1) &= \{1 - E(U_0 \bigvee U_2 \bigvee U_3 | X = x)\} \\ &= (1 - a_0^x)(1 - a_2^x)(1 - a_3^x) \end{aligned}$$

and

$$\begin{aligned} 1 - E(D|X = x, A_1 = 1, A_2 = 0) &= \{1 - E(U_0 \bigvee U_1 \bigvee U_4 | X = x)\} \\ &= (1 - a_0^x)(1 - a_1^x)(1 - a_4^x) \end{aligned}$$

and

$$\begin{aligned} 1 - E(D|X = x, A_1 = 1, A_2 = 1) &= \{1 - E(U_0 \bigvee U_1 \bigvee U_3 | X = x)\} \\ &= (1 - a_0^x)(1 - a_1^x)(1 - a_3^x) \end{aligned}$$

So

$$\begin{aligned} & E(Y|X = x, A_1 = 1, A_2 = 1)\}E(Y|X = x, A_1 = 0, A_2 = 0)\} \\ &= (1 - a_0^x)^2(1 - a_1^x)(1 - a_2^x)(1 - a_3^x)(1 - a_4^x) \\ &= \{E(Y|X = x, A_1 = 0, A_2 = 1)\}\{E(Y|X = x, A_1 = 1, A_2 = 0)\}, \end{aligned}$$

which implies $\beta^* = 0$, a contradiction.