Supplementary Material: Does exposure prediction bias health effect estimation?

The relationship between confounding adjustment and exposure prediction

Exposure prediction that inflates confounding bias

Without loss of generality, here and throughout the eAppendix, all models are intercept free.

Consider an oversimplified scenario where all potential confounders are used to predict the exposure and no confounding adjustment is made in the health effects regression model. We recognize that this rarely occurs in practice, but it serves as a useful example to illustrate that exposure prediction can increase the magnitude of the confounding bias.

Let $\mathbf{C}_i$ be a set of normally distributed covariates with mean $\boldsymbol{\mu}_c$ and covariance Σ_c, and assume that the outcome Y_i and the exposure X_i are generated under the following linear models:

$$Y_i = X_i\beta_0 + \mathbf{C}_i\gamma_0 + \epsilon_i^y \tag{1}$$

$$X_i = \mathbf{C}_i\alpha_0 + \epsilon_i^x \tag{2}$$

where ϵ_i^y and ϵ_i^x are independent, normally distributed, mean zero error terms with variances $\sigma^2_{y|xc}$ and $\sigma^2_{x|c}$. Suppose interest lies in the estimation of the linear exposure-outcome relationship β_0, conditional on the covariates $\mathbf{C}_i$. Here, and throughout, no restriction is placed on γ_0 or α_0, and individual components of the vectors are free to be 0.

We define bias due to the lack of adjustment for confounding as the bias in our estimation of β_0 that is due to failure to control for the covariates $\mathbf{C}_i$. That is, if one were to ignore

$\mathbf{C}_i$ when fitting the health effects regression model and instead fit $Y_i = X_i\beta + \epsilon_i$, then the least squares estimate for β, call it $\widehat{\beta}_x$, is biased. We call this the bias due to the lack of adjustment for confounding and denote it as $bias(\widehat{\beta}_x) = \mathrm{E}[\widehat{\beta}_x - \beta_0]$.

Now suppose that the exposure is unobserved. Further, let $W_i = \mathbf{C}_i\alpha_0$ be the predicted exposure with α_0 known. Consider fitting the health effects regression model that uses the predicted exposure W_i in place of the true exposure X_i and fails to control for any confounding ($Y_i = W_i\beta + \epsilon_i$). The bias of the least squares estimator for β, call it $\widehat{\beta}_w$, is given by:

$$bias(\widehat{\beta}_w) = \mathrm{E}[\widehat{\beta}_w - \beta_0] = bias(\widehat{\beta}_x)\frac{\sigma_x^2}{\sigma_w^2} \tag{3}$$

where $\sigma_x^2 = \sigma_w^2 + \sigma_{x|c}^2$ and $\sigma_w^2 = \alpha_0^T\Sigma_c\alpha_0$ denote the variances of X and W, respectively. We call the second term of Equation 3 ($\frac{\sigma_x^2}{\sigma_w^2}$) the bias inflation factor, and note that it is equal to the inverse of the population R^2 when using W to predict X. From an intuitive standpoint, we expect that the variation in the true exposure σ_x^2 will always be more than the variation in the predicted exposure σ_w^2, and hence, the bias inflation factor is always greater than 1 (i.e. the R^2 is always less than 1).

Notice that the bias of $\widehat{\beta}_w$ is the product of two pieces: (1) the bias due to lack of adjustment for confounding assuming that the true exposure is known ($bias(\widehat{\beta}_x)$); and (2) the bias inflation factor due exposure prediction ($\frac{\sigma_x^2}{\sigma_w^2}$). It is easy to see that $bias(\widehat{\beta}_x) = 0$ implies

that $bias(\widehat{\beta}_w) = 0$; therefore, bias inflation due to exposure prediction should only an issue if there is some uncontrolled confounding. However, even in the presence of uncontrolled confounding, $bias(\widehat{\beta}_x) \neq 0$ implies $bias(\widehat{\beta}_w) \neq 0$.

The bias inflation factor decreases as R^2 increases and goes to 1 as the exposure model is able to predict the true exposure X more accurately. Note that the bias inflation factor can be large even if the bias due to lack of adjustment for confounding is small. It is tempting to suggest that in an attempt to obtain an unbiased estimate of the health effect, a researcher should build an exposure model that more accurately predicts the true exposure (a model with the largest R^2). However, the relationship is not that simple. As we will show next, the bias of the health effect estimate can either increase or decrease in magnitude if a subset of the confounders are used in the exposure prediction model.

Bias when confounding has been partially controlled or different subsets of confounders are used to predict exposure

First, consider the same set up as before, with the exposure-outcome-confounder relationship given by eEquation 1 and 2. Let $\mathbf{C} = (\mathbf{C}^{(1)}, \mathbf{C}^{(2)})$ and $\Sigma_C = \text{var}(\mathbf{C}_i) = \begin{pmatrix} \Sigma_1 & \Sigma_{12} \\ \Sigma_{21} & \Sigma_2 \end{pmatrix}$. Further, let $W = \mathbf{C}\alpha$ be the predicted exposure if the exposure model from eEquation 2 were known exactly, $W_1 = \mathbf{C}^{(1)}\alpha_1^*$ be the predicted exposure if the misspecified exposure model $X_i = \mathbf{C}_i^{(1)}\alpha_1^* + \epsilon$ were known exactly, and $W_2 = \mathbf{C}_i^{(2)}\alpha_2^*$ be the predicted exposure if the misspecified exposure model $X_i = \mathbf{C}_i^{(2)}\alpha_2^* + \epsilon$ were known exactly.

eTable 1 provides the bias of the health effect estimate for each choice of the predicted

exposure and a health effects regression model that either fails to control for any confounding ($Y = W\beta + \epsilon$) or a health effects regression model that controls for only $\mathbf{C}^{(1)}$ ($Y = W\beta + \mathbf{C}^{(1)}\gamma + \epsilon$). Further, let $\widetilde{R}^2_z$ denote the population value of the R^2 from the exposure model that uses arbitrary Z as a prediction of X. eTable 2 provides the R^2 and its corresponding population value for each of the predicted exposures W, W_1, or W_2.

The bias of $\widehat{\beta}_w$ given in eTable 1 is the bias of the health effect estimate provided in eEquation 3 that was previously described under the situation that the predicted exposure W is used in an health effects regression model that fails to control for any confounding. Recall that is was shown that this bias is the product of the bias due to lack of adjustment for any confounding and a bias inflation factor due to exposure prediction that is the inverse of the $\widetilde{R}^2_w$.

This relationship holds true for any collection of covariates, regardless of their association with the exposure and the outcome. For example, suppose all $\mathbf{C}$ are only related to the exposure. Then, there is no confounding and as a result, the bias of $\widehat{\beta}_w$ is 0. Similarly, suppose that all $\mathbf{C}$ are only related to the outcome. Then, $\widetilde{R}^2_w = 0$ because $\mathbf{C}$ has no power to predict exposure, and the bias of $\widehat{\beta}_w$ increases in magnitude to infinity.

Next, consider a situation where the true set of confounders $\mathbf{C}$ is unknown to the researcher but the true exposure X is observed, and instead of controlling for the full set of $\mathbf{C}$s, the decision is made to only control for the subset $\mathbf{C}^{(1)}$ (first row, second column). The bias of the health effect estimate from the misspecified health effects regression model $Y = X\beta + \gamma\mathbf{C}^{(1)} + \epsilon$ is given by $bias(\widehat{\beta}^{(1)}_x)$ in eTable 1. This corresponds to the bias due to the

failure to control for the confounding due to $\mathbf{C}^{(2)}$. In other words, it is the bias due to confounding that remains after controlling for $\mathbf{C}^{(1)}$, but failing to control for the full set of necessary confounders $\mathbf{C}$. Suppose that $\mathbf{C}^{(1)}$ contains all covariates that are confounders and $\mathbf{C}^{(2)}$ contains any remaining covariates. Then, $bias(\widehat{\beta}_x^{(1)}) = 0$ because confounding has sufficiently been controlled by $\mathbf{C}^{(1)}$ alone. However, suppose that $\mathbf{C}^{(2)}$ contains all covariates that are confounders, $\mathbf{C}^{(1)}$ contains any remaining covariates, and $\mathbf{C}^{(1)}$ and $\mathbf{C}^{(2)}$ are uncorrelated. Then, $bias(\widehat{\beta}_x^{(1)}) = bias(\widehat{\beta}_x)\widetilde{R}_{w_2}^{-2}$ so that the bias of the health effect estimate is inflated by controlling for covariates that are not confounders. This is a specific example of bias inflation that arises from conditioning on instrumental variables.[1,2]

Now consider a situation where the true exposure X is unobserved, and instead is predicted with a subset of the $\mathbf{C}$s (second row, first column). The $bias(\widehat{\beta}_{w_1})$ is the bias of the health effect estimate in the situation that the predicted exposure $W_1 = \mathbf{C}^{(1)}\alpha_1^*$ is used in the health effects regression model that fails to control for any confounding. From eTable 1, we note that this bias decomposes into two parts, with the first one being the bias due to the failure to control for confounding due to $\mathbf{C}^{(2)}$. Therefore, ignoring the second term, using $\mathbf{C}^{(1)}$ to predict the exposure appears to help control the confounding due to $\mathbf{C}^{(1)}$. However, this is not exactly the case, as the second term of $bias(\widehat{\beta}_{w_1})$ in eTable 1 can either decrease or increase the magnitude of the bias. Further we note that $bias(\widehat{\beta}_{w_1})$ depends on the inverse of $\widetilde{R}_{w_1}^2$; therefore, the bias of $\widehat{\beta}_{w_1}$ is a function of how well W_1 predicts X. As $\widetilde{R}_{w_1}^2$ goes to 1, $bias(\widehat{\beta}_{w_1}) = bias(\widehat{\beta}_x)$, so that if W_1 predicts X perfectly, we are left with the bias due to lack of adjustment for confounding in the situation where the true exposure X is known. Similarly, as $\widetilde{R}_{w_1}^2$ goes to 0, the $bias(\widehat{\beta}_{w_1})$ increases in magnitude to infinity,

suggesting that if we cannot accurately predict the exposure, we cannot return a valid effect estimate. However, as $\widetilde{R}^2_{w_1}$ varies between 0 and 1, no general statement can be made about the magnitude of the bias. Similar results hold for $bias(\widehat{\beta}_{w_2})$.

Suppose that $\mathbf{C}^{(1)}$ contains all covariates that are confounders and $\mathbf{C}^{(2)}$ contains any remaining covariates. Then, $bias(\widehat{\beta}_{w_1}) = bias(\widehat{\beta}_x)\widetilde{R}^{-2}_{w_1}$, or in other words, we have an expression similar to $bias(\widehat{\beta}_w)$ in that we are inflating the bias due to lack of adjustment for confounding. By moving covariates that are not confounders from $\mathbf{C}^{(2)}$ into $\mathbf{C}^{(1)}$, we would increase $\widetilde{R}^2_{w_1}$ and as a result $bias(\widehat{\beta}_{w_1})$ would decrease. Therefore, if all confounders are used to predict the exposure, we decrease the bias of the health effect estimate by improving the prediction accuracy.

Another situation provided in eTable 1 is a situation where the true exposure X is unobserved, instead is predicted with a subset of the $\mathbf{C}$s, and a different set of $\mathbf{C}$s are used to control confounding in the health effects regression model (third row, second column). Specifically, the $bias(\widehat{\beta}^{(1)}_{w_2})$ is the bias of the health effect estimate in the situation that the predicted exposure $W_2 = \mathbf{C}^{(2)}\alpha^*_2$ is used in the health effects regression model that controls for only $\mathbf{C}^{(1)}$ ($Y = W_2\beta + \mathbf{C}^{(1)}\gamma + \epsilon$). We wish to only point out a few features of the expression for this bias. First, the bias depends on the true underlying effect β_0. As the true effect size increases, so does the magnitude of bias. Second, the expression for the bias of $\widehat{\beta}^{(1)}_{w_2}$ is much more complex than any of the other biases given in eTable 1 and will not be described in detail. However, suppose again that $\mathbf{C}^{(1)}$ contains all covariates that are confounders and $\mathbf{C}^{(2)}$ contains any remaining covariates. Further, assume that $\mathbf{C}^{(1)}$ and $\mathbf{C}^{(2)}$ are uncorrelated. Then, $bias(\widehat{\beta}^{(1)}_{w_2}) = 0$. This occurs because: (1) confounding has been sufficiently controlled

through $\mathbf{C}^{(1)}$; and (2) the exposure is predicted with covariates that are uncorrelated with confounders. However, if $\mathbf{C}^{(1)}$ and $\mathbf{C}^{(2)}$ are correlated, then $bias(\widehat{\beta}_{w_2}^{(1)}) \neq 0$.

Considering these results, if we can separate our covariates into two orthogonal sets, one of which contains all necessary confounders, then we can hope to construct an exposure prediction model along with a health effects regression model that yields an unbiased health effect estimate.

The final situation provided in eTable 1 (row 4, column 2) is such that the subset $\mathbf{C}^{(1)}$ is used to control confounding while the full set $\mathbf{C}$ is used to predict the exposure. Specifically, the $bias(\widehat{\beta}_w^{(1)})$ is the bias of the health effect estimate in the situation that the predicted exposure $W = \mathbf{C}\alpha$ is used in the health effects regression model that controls for only $\mathbf{C}^{(1)}$ ($Y = W\beta + \mathbf{C}^{(1)}\gamma + \epsilon$). The main feature of $bias(\widehat{\beta}_w^{(1)})$ is that if $\mathbf{C}^{(1)}$ contains all confounders, then $b_x^{(1)} = 0$ implying $bias(\widehat{\beta}_w^{(1)}) = 0$. Therefore, if all confounders are included in the health effects regression model, an unbiased health effect estimate can be estimated if the exposure prediction model is correctly specified.

The biases given in eTable 1 are difficult to compare, except for in the simplest situations as in $bias(\widehat{\beta}_x)$ and $bias(\widehat{\beta}_w)$. Therefore, it is difficult to make any general conclusions about whether including or excluding a potential confounder from either the exposure model or the health effects regression model is beneficial or detrimental to the final goal of effect estimation.

The previous point warrant further discussion; when the goal of a study is effect estimation, the decision to include or exclude a potential confounder from either the outcome or the

exposure model needs to be based on more than just the predictive power of the potential confounder on the exposure or the strength of the relationship with the outcome, but instead the decision needs to be based on some tradeoff between the two. Current statistical methods for model selection fail in this regard, as they have been designed to control confounding and ignore exposure prediction all together.

All previous results can be extended to situations where the outcome, exposure, and confounders are not assumed to be normally distributed by replacing expectations with convergence in probability. The probability limits have the exact form of the biases given in eTable 1.

Bias due to exposure prediction under exposure model misspecification

Assume that the data is generated under the following linear models:

$$Y_i = X_i\beta_x + \mathbf{C}_i^{(1)}\gamma_1 + \epsilon_i^y \tag{4}$$

$$X_i = \mathbf{C}_i^{(1)}\alpha_1 + \mathbf{C}_i^{(2)}\alpha_2 + \mathbf{C}_i^{(3)}\alpha_3 + \epsilon_i^x \tag{5}$$

where $\mathbf{C}^{(1)}$, $\mathbf{C}^{(2)}$, and $\mathbf{C}^{(3)}$ denote subsets of $\mathbf{C}$. Notice that in this data generating scheme, there is only partial overlap in the sets of covariates in the two models, and that the necessary set of confounders is $\mathbf{C}^{(1)}$.

Consider the situation where both $\mathbf{C}^{(1)}$ and $\mathbf{C}^{(2)}$ are used to predict the exposure, and

the predicted exposure $W = \mathbf{C}^{(1)}\alpha_1^* + \mathbf{C}^{(2)}\alpha_2^*$ from the misspecified exposure model $X_i = \mathbf{C}_i^{(1)}\alpha_1^* + \mathbf{C}_i^{(2)}\alpha_2^* + \epsilon$ were known exactly. If the predicted exposure W is used in the health effects regression model that properly adjusts for confounding ($Y_i = W_i\beta + \mathbf{C}_i^{(1)}\gamma_1 + \epsilon_i^y$), then the resulting estimate of β is unbiased.

However, if we were to purposefully exclude confounders $\mathbf{C}^{(1)}$ from the exposure prediction model, so that only $\mathbf{C}^(2)$ is used to predict the exposure, then the resulting health effect will biased. Specifically, assume the predicted exposure $W = \mathbf{C}^{(2)}\alpha_2^*$ from the misspecified exposure model $X_i = \mathbf{C}_i^{(2)}\alpha_2^* + \epsilon$ were known exactly. If the predicted exposure W is used in the health effects regression model that properly adjusts for confounding ($Y_i = W_i\beta + \mathbf{C}_i^{(1)}\gamma_1 + \epsilon_i^y$), then the resulting estimate of β is biased, with bias given by:

$$E[\widehat{\beta} - \beta] = \beta \frac{-\alpha_2^{*T}\Sigma_{21}\Sigma_1^{-1}\left\{(\Sigma_1 - \Sigma_{12}\Sigma_2^{-1}\Sigma_{21})\alpha_1 + (\Sigma_{13} - \Sigma_{12}\Sigma_2^{-1}\Sigma_{23})\alpha_3\right\}}{\alpha_2^{*T}(\Sigma_2 - \Sigma_{21}\Sigma_1^{-1}\Sigma_{12})\alpha_2^*}$$

where $\alpha_2^* = \alpha_2 + \Sigma_2^{-1}\Sigma_{21}\alpha_1 + \Sigma_2^{-1}\Sigma_{23}\alpha_3$. Note that this bias is zero if $\mathbf{C}^{(2)}$ is uncorrelated with the confounders $\mathbf{C}^{(1)}$. In general, the bias will be non-zero.

Equivalence of including confounders into exposure prediction model versus orthogonalization

Consider the following two approaches for predicting the exposure: (1) including all confounders into the exposure prediction model and (2) orthogonalizing the covariates used to

predict the exposure to the confounders. Under the health effect regression model that includes all confounders, these two approaches are analytically equivalent. Considering the same set up as before, let $\mathbf{C}^{(1)}$ be the confounders, and let $\mathbf{C}^{(2)}$ be additional predictors of the exposure.

A known result from linear models implies that health effect estimate from strategy 1 is given by $\widehat{\beta}_1 = \frac{<Y,w^{\perp}>}{||w^{\perp}||^2}$, where $w^{\perp} = P_{C_1^{\perp}} w = P_{C_1^{\perp}} P_{C_1,C_2} X = P_{C_1^{\perp}} P_{C_2} X$ and P_L denotes the projection matrix onto the space spanned by L. This is equivalent to a univariate regression where $w^{\perp}$ is the only predictor of the outcome Y. In strategy 2, we are using a predicted exposure $w^* = P_{C_1^{\perp}} C_2 (C_2^T P_{C_1^{\perp}} C_2)^{-1} C_2^T P_{C_1^{\perp}} X = P_{C_1^{\perp}} P_{C_2} X$. Since w^* is orthogonal to C_1, the estimate from strategy 2 is given by $\widehat{\beta}_2 = \frac{<Y,w^*>}{||w^*||^2}$. Since $w^* = w^{\perp}$, $\widehat{\beta}_1 = \widehat{\beta}_2$.

Additional Simulated Results

Following the simulation setup of the main text exactly, we provide additional simulated results for two additional choices of the parameter γ. Specifically, let

$$\gamma^a = (0, -0.044, -0.075, 0.105, 0.090, -0.082, 0.096, 0.0897, -0.041, 0.011)$$

$$\gamma^b = (0.025, 0.0067, -0.0058, 0.005, 0.0208, 0.0033, 0.025, 0.025, 0.0125, 0)$$

The purpose of these two additional specifications is to illustrate that in some cases, increasing the R^2 always decreases the bias, while in others, increasing the R^2 always increases

the bias. From eFigure 1, we note that the bias increases with the R^2. Therefore, adding additional covariates to the exposure prediction model adds bias to the estimated health effect. From eFigure 2, we note that the bias decreases as R^2 increases. Therefore, adding additional covariates to the exposure prediction model improves the health effect estimate.

These results, in addition to those in the main text, provide evidence that the bias of a health effect estimate can either reduce or increase when predicting the exposure.

Proof of Results

The proofs for all results in eTable 1 are similar; therefore, we will only provide the proof of $bias(\widehat{\beta}_w)$, as all other results follow in a similar fashion. Additionally, we provide a proof of the probability limit of the estimator without any parametric distributional assumptions.

Let $\mathbf{C}_i$ be a set of normally distributed covariates with mean $\boldsymbol{\mu}_c$ and covariance Σ_c, and assume that the outcome Y_i and the exposure X_i are generated under the following linear models:

$$Y_i = \beta_0 X_i + \mathbf{C}_i \gamma_0 + \epsilon_i^y$$

$$X_i = \mathbf{C}_i \alpha_0 + \epsilon_i^x$$

where ϵ_i^y and ϵ_i^x are independent, normally distributed, mean zero error terms with variances $\sigma^2_{y|xc}$ and $\sigma^2_{x|c}$. The hierarchical structure, along with the assumption of normality at each

level, implies that:

$$\begin{pmatrix} Y \\ X \\ \mathbf{C} \end{pmatrix} \sim \mathcal{N}\left(\mu = \begin{pmatrix} \mu_y \\ \mu_x \\ \mu_c \end{pmatrix}, \Sigma = \begin{pmatrix} \sigma_y^2 & \sigma_{yx} & \Sigma_{yc} \\ & \sigma_x^2 & \Sigma_{xc} \\ & & \Sigma_c \end{pmatrix} \right)$$

where

$$\begin{aligned} \sigma_x^2 &= \sigma_{x|c}^2 + \alpha_0^T \Sigma_c \alpha_0 \\ \Sigma_{xc} &= \alpha_0^T \Sigma_c \\ \Sigma_{yc} &= \beta_0 \alpha_0^T \Sigma_c + \gamma_0 \Sigma_c \\ \sigma_{yx} &= \beta_0 (\sigma_{x|c}^2 + \alpha_0^T \Sigma_c \alpha_0) + \gamma_0 \Sigma_c \alpha_0 \end{aligned}$$

Once this joint distribution is defined, all that is left is to find the conditional distribution of interest. In this case, we are interested in:

$$(Y|W = \alpha_0^T \mathbf{C}) \sim \mathcal{N}\left(\mu_y + (\beta_0 \alpha_0^T \Sigma_c \alpha_0 + \gamma_0 \Sigma_c \alpha_0)(\alpha_0^T \Sigma_c \alpha_0)^{-1}(W - \alpha_0^T \mu_c), \sigma_y^2 - (\Sigma_{yc} \alpha_0)(\alpha_0^T \Sigma_c \alpha_0)^{-1}(\Sigma_{yc} \alpha_0)^T\right)$$

which implies that

$$\begin{aligned}
\mathrm{E}[\widehat{\beta}_w] &= \beta_0 + \frac{\alpha_0^T \Sigma_c \gamma_0}{\alpha_0^T \Sigma_c \alpha_0} \\
&= \beta_0 + \frac{\alpha_0^T \Sigma_c \gamma_0}{\sigma^2_{x|c} + \alpha_0^T \Sigma_c \alpha_0} \frac{\sigma^2_{x|c} + \alpha_0^T \Sigma_c \alpha_0}{\alpha_0^T \Sigma_c \alpha_0} \\
&= \beta_0 + \frac{\alpha_0^T \Sigma_c \gamma_0}{\sigma^2_{x|c} + \alpha_0^T \Sigma_c \alpha_0} \frac{\sigma^2_x}{\sigma^2_w}
\end{aligned}$$

All other results of eTable 1 follow in a similar manner.

Suppose that we loosen the distributional assumptions so that $\mathbf{C}_i$ is a set of covariates with mean $\boldsymbol{\mu}_c$ and finite covariance Σ_c, and assume that the outcome Y_i and the exposure X_i are generated under the same linear model, but with ϵ_i^y and ϵ_i^x are independent distributed, mean zero error terms with finite variances $\sigma^2_{y|xc}$ and $\sigma^2_{x|c}$. Without loss of generality, assume $\boldsymbol{\mu}_c = 0$. Then,

$$\begin{aligned}
\widehat{\beta}_w &= (W^T W)^{-1} W^T Y \\
&= (\alpha_0^T \mathbf{C}^T \mathbf{C} \alpha_0)^{-1} \alpha_0^T \mathbf{C}^T Y \\
&= \frac{\alpha_0^T \mathbf{C}^T Y}{\alpha_0^T \mathbf{C}^T \mathbf{C} \alpha_0} \\
&\xrightarrow{p} \frac{\alpha_0^T \Sigma_{cy}}{\alpha_0^T \Sigma_c \alpha_0} = \beta_0 + \frac{\alpha_0^T \Sigma_c \gamma_0}{\alpha_0^T \Sigma_c \alpha_0}
\end{aligned}$$

References

1. Bhattacharya, J. and Vogt, W. B. (2012), Do Instrumental Variables Belong in Propensity Scores? International Journal of Statistics & Economics, 9, 107127.

2. Pearl, J. (2012), On a class of bias-amplifying variables that endanger effect estimates, arXiv preprint arXiv:1203.3503.

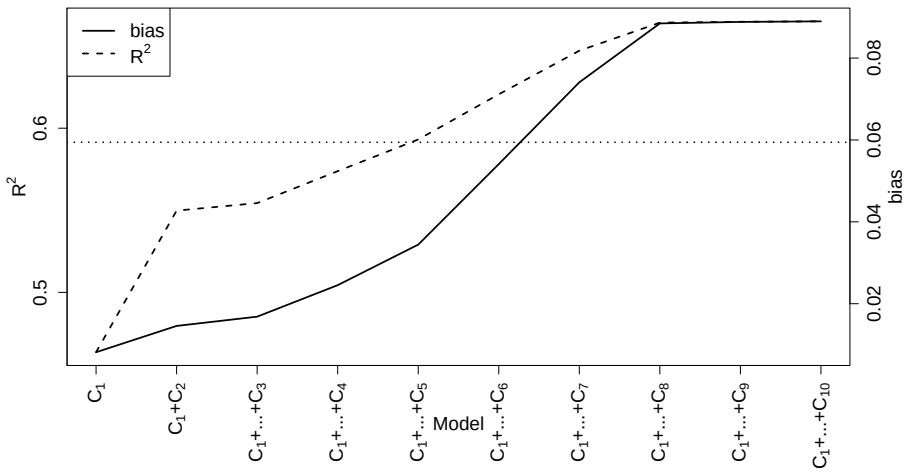
eTable 1: The bias of the health effect estimate

	Outcome model	
Exposure	$Y = W + \epsilon$	$Y = W + C^{(1)} + \epsilon$
$W = X$	$\mathrm{E}[\widehat{\beta}_x - \beta_0] = b_x = \frac{\alpha^T \Sigma_c \gamma}{\sigma^2_{x\|c} + \alpha^T \Sigma_C \alpha}$	$\mathrm{E}[\widehat{\beta}^{(1)}_x - \beta_0] = b^{(1)}_x = \frac{\alpha_2^T(\Sigma_2 - \Sigma_{21}\Sigma_1^{-1}\Sigma_{12})\gamma_2}{\sigma^2_{x\|c} + \alpha_2^T(\Sigma_2 - \Sigma_{21}\Sigma_1^{-1}\Sigma_{12})\alpha_2}$
$W = C^{(1)}\alpha_1^*$	$\mathrm{E}[\widehat{\beta}_{w_1} - \beta_0] = b_{w_1} = b^{(1)}_x + (b_x - b^{(1)}_x)\widetilde{R}^{-2}_{w_1}$	NA[a]
$W = C^{(2)}\alpha_2^*$	$\mathrm{E}[\widehat{\beta}_{w_2} - \beta_0] = b_{w_2} = b^{(2)}_x + (b_x - b^{(2)}_x)\widetilde{R}^{-2}_{w_2}$	$\mathrm{E}[\widehat{\beta}^{(1)}_{w_2} - \beta_0] = \beta_0 \frac{\alpha_2^{*T}\Sigma_{21}\Sigma_1^{-1}\Sigma_{12}\alpha_2^* - \alpha_1^{*T}\Sigma_{12}\alpha_2^*}{\alpha_2^{*T}(\Sigma_2 - \Sigma_{21}\Sigma_1^{-1}\Sigma_{12})\alpha_2^*} + \frac{\alpha_2^{*T}(\Sigma_2 - \Sigma_{21}\Sigma_1^{-1}\Sigma_{12})\gamma_2}{\alpha_2^{*T}(\Sigma_2 - \Sigma_{21}\Sigma_1^{-1}\Sigma_{12})\alpha_2^*}$
$W = C\alpha$	$\mathrm{E}[\widehat{\beta}_w - \beta] = b_w = b_x \widetilde{R}^{-2}_w$	$\mathrm{E}[\widehat{\beta}^{(1)}_w - \beta_0] = b^{(1)}_x \frac{\sigma^2_x - \sigma^2_{w_1}}{\sigma^2_w - \sigma^2_{w_1}}$

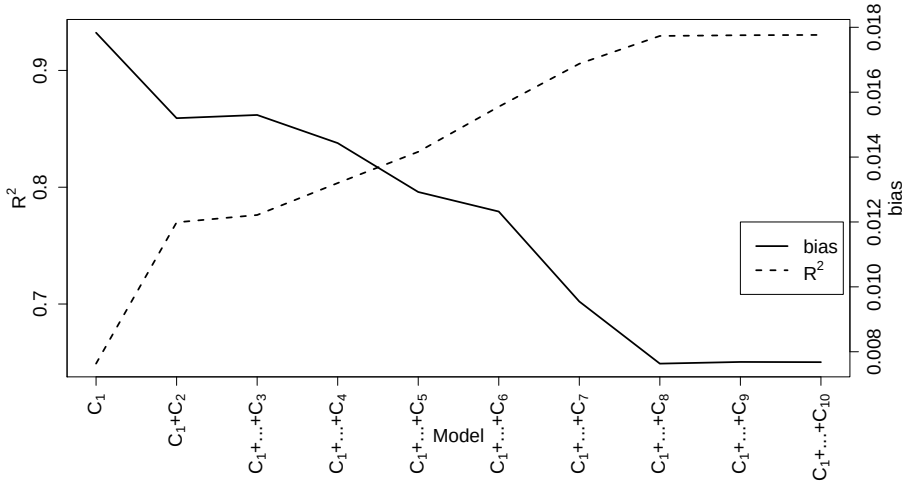
[a] Does not exist due to collinearity between predicted exposure and confounding adjustment.

eTable 2: Coefficient of determination (R^2) and its corresponding population value ($\widetilde{R}^2$)

Predicted exposure	**Coefficient of determination**	**Population based** R^2
$W = \mathbf{C}\alpha_0$	$R^2_w = \frac{\sum(W_i - X_i)^2}{\sum(X_i - \bar{X})^2}$	$\widetilde{R}^2_w = \frac{\alpha^T \Sigma_C \alpha}{\sigma^2_{x\|c} + \alpha^T \Sigma_c \alpha} = \frac{\sigma^2_w}{\sigma^2_x}$
$W_1 = \mathbf{C}^{(1)}\alpha_1^*$	$R^2_{w_1} = \frac{\sum(W_{1i} - X_i)^2}{\sum(X_i - \bar{X})^2}$	$\widetilde{R}^2_{w_1} = \frac{\alpha_1^{*T}\Sigma_1\alpha_1^*}{\sigma^2_{x\|c} + \alpha^T \Sigma_c \alpha} = \frac{\sigma^2_{w_1}}{\sigma^2_x}$
$W_2 = \mathbf{C}^{(2)}\alpha_2^*$	$R^2_{w_2} = \frac{\sum(W_{2i} - X_i)^2}{\sum(X_i - \bar{X})^2}$	$\widetilde{R}^2_{w_2} = \frac{\alpha_2^{*T}\Sigma_C\alpha_2^*}{\sigma^2_{x\|c} + \alpha^T \Sigma_c \alpha} = \frac{\sigma^2_{w_2}}{\sigma^2_x}$



eFigure 1: Tradeoff between bias and R^2,
$\gamma^a = (0, -0.044, -0.075, 0.105, 0.090, -0.082, 0.096, 0.0897, -0.041, 0.011)$



eFigure 2: Tradeoff between bias and R^2,
$\gamma^a = (0.025, 0.0067, -0.0058, 0.005, 0.0208, 0.0033, 0.025, 0.025, 0.0125, 0)$