A Appendix

A.1 Proof of Theorem 2.1

Proof. Suppose there is a sufficient reduction $R(\boldsymbol{X}) = \boldsymbol{B}^{*\prime}\boldsymbol{X}$ and the associated unspecified functions $\{g_t\}_{t\in\mathcal{T}}$, i.e., assume representation (2.5) with $\boldsymbol{B} = \boldsymbol{B}^*$. Let $g_t^*(\boldsymbol{B}^{*\prime}\boldsymbol{X}) := g_t(\boldsymbol{B}^{*\prime}\boldsymbol{X}) - \mathbb{E}[g_T(\boldsymbol{B}^{*\prime}\boldsymbol{X})|\boldsymbol{X}]$ $(t \in \mathcal{T})$, which, by rearrangement, gives $g_t(\boldsymbol{B}^{*\prime}\boldsymbol{X}) = \mathbb{E}[g_T(\boldsymbol{B}^{*\prime}\boldsymbol{X})|\boldsymbol{X}] + g_t^*(\boldsymbol{B}^{*\prime}\boldsymbol{X})$ $(t \in \mathcal{T})$, where, by definition, the term $\mathbb{E}[g_T(\boldsymbol{B}^{*\prime}\boldsymbol{X})|\boldsymbol{X}]$ does not depend on T (as T is integrated out) and the term $g_t^*(\boldsymbol{B}^{*\prime}\boldsymbol{X})$ $(t \in \mathcal{T})$ is designed to satisfy (2.7). Thus, for any contrast vector $\boldsymbol{c}$, we can rewrite (2.5) (with $\boldsymbol{B} = \boldsymbol{B}^{*\prime}$) as

$$\sum_{t=1}^{K} c_t g_t(\boldsymbol{B}^{*\prime}\boldsymbol{X}) = \sum_{t=1}^{K} c_t\big\{\mathbb{E}[g_T(\boldsymbol{B}^{*\prime}\boldsymbol{X})|\boldsymbol{X}] + g_t^*(\boldsymbol{B}^{*\prime}\boldsymbol{X})\big\} = \sum_{t=1}^{K} c_t g_t^*(\boldsymbol{B}^{*\prime}\boldsymbol{X}),$$

where the second equality follows from $\sum_{t=1}^{K} c_t\mathbb{E}[g_T(\boldsymbol{B}^{*\prime}\boldsymbol{X})|\boldsymbol{X}] = \mathbb{E}[g_T(\boldsymbol{B}^{*\prime}\boldsymbol{X})|\boldsymbol{X}]\sum_{t=1}^{K} c_t = 0$. Therefore, for representation (2.5), we can always reparametrize the set of functions $\{g_t\}_{t\in\mathcal{T}}$ by $\{g_t^*\}_{t\in\mathcal{T}}$ that satisfies (2.7), implying that we can assume $g_t = g_t^*$, without loss of generality. By definition (2.4), we can re-express (2.5) (with $\boldsymbol{B} = \boldsymbol{B}^*$) as

$$\mathcal{C}(\boldsymbol{X};\boldsymbol{c}) = \sum_{t=1}^{K} c_t\mathbb{E}\left[Y|\boldsymbol{X}, T = t\right] = \sum_{t=1}^{K} c_t g_t^*(\boldsymbol{B}^{*\prime}\boldsymbol{X}), \tag{A.1}$$

for any contrast vector $\boldsymbol{c}$. Under the general model (2.1), (A.1) indicates that the $\boldsymbol{X}$-by-T interaction term $g(\boldsymbol{X}, T = t)$ $(t \in \mathcal{T})$ in (2.1) corresponds to the term $g_t^*(\boldsymbol{B}^{*\prime}\boldsymbol{X})$ $(t \in \mathcal{T})$ in (A.1), since the second equation in (A.1) holds for any arbitrary contrast $\boldsymbol{c} = (c_1, \ldots, c_K)$. Furthermore, the term $\mu(\boldsymbol{X})$ in (2.1) corresponds to $\mu(\boldsymbol{X})$ of model (2.6), since $\mu(\boldsymbol{X})$ of model (2.6) represents the unspecified $\boldsymbol{X}$ marginal effect. Thus, under the general model (2.1), (2.5) with $\boldsymbol{B} = \boldsymbol{B}^*$ implies model (2.6).

Conversely, if we assume model (2.6), then, by definition (2.4) we have

$$\mathcal{C}(\boldsymbol{X};\boldsymbol{c}) \;=\; \sum_{t=1}^{K} c_t\mathbb{E}\left[Y|\boldsymbol{X}, T = t\right] =\; 0 \;+\; \sum_{t=1}^{K} c_t g_t^*(\boldsymbol{B}^{*\prime}\boldsymbol{X}), \tag{A.2}$$

for all contrast vectors $\boldsymbol{c}$, where the $\boldsymbol{X}$ marginal effect $\mu(\boldsymbol{X})$ in (2.6) drops out due to $\sum_{t=1}^{K} c_t = 0$. Expression (A.2) implies that $\boldsymbol{B}^{*\prime}\boldsymbol{X}$ is a sufficient reduction for $\mathcal{C}(\boldsymbol{X};\boldsymbol{c})$, implying (2.5) with $\boldsymbol{B} = \boldsymbol{B}^*$. ☐

A.2 Proof of Corollary 2.1

Proof. By Theorem 2.1, $R(\boldsymbol{X}) = \boldsymbol{B}^{*\prime}\boldsymbol{X}$ of model (2.6) is a sufficient reduction (2.5). We need to show that span$(\boldsymbol{B}^*)$ is a minimal reduction, and therefore span$(\boldsymbol{B}^*) = S_{\mathcal{C}|\boldsymbol{X}}$. Due to the constraint (2.7), $\boldsymbol{B}^*$ of model (2.6) is not related to the $\boldsymbol{X}$ marginal effect, therefore there is no "nuisance" dimension contained in span$(\boldsymbol{B}^*)$. Moreover, since $\boldsymbol{B}^* \in \Theta_q$, the columns of $\boldsymbol{B}^*$ are linearly independent. This implies $\boldsymbol{B}^*$ is a basis for $S_{\mathcal{C}|\boldsymbol{X}}$.

☐

A.3 Justification for excluding the main effect term in the optimization-based representation (2.8)

Under model (2.6) of the main manuscript, we can view the treatment t-specific functions $\{g_t^*\}_{t\in\mathcal{T}}$ and the dimension reduction matrix $\boldsymbol{B}^*$ as the solution to the following optimization:

$$\begin{aligned}(g_1^*, \ldots, g_K^*, \boldsymbol{B}^*) \quad = \quad & \underset{g_t\in\mathcal{H}^{(\boldsymbol{B})}, \boldsymbol{B}\in\Theta_q}{\text{argmin}} \quad \mathbb{E}\big[\big(Y - \mu(\boldsymbol{X}) - g_T(\boldsymbol{B}'\boldsymbol{X})\big)^2\big] \\ & \text{subject to} \quad \mathbb{E}\left[g_T(\boldsymbol{B}'\boldsymbol{X})|\boldsymbol{X}\right] = 0,\end{aligned} \tag{A.3}$$

where $\mu(\boldsymbol{X})$ is the fixed term given from the assumed model (2.6). However, in (A.3), we have

$$\begin{aligned}&\underset{g_t\in\mathcal{H}^{(\boldsymbol{B})}, \boldsymbol{B}\in\Theta_q}{\arg\min} \;\mathbb{E}\Big[Y^2 + (\mu(\boldsymbol{X}))^2 + \big(g_T(\boldsymbol{B}'\boldsymbol{X})\big)^2 - 2\mu(\boldsymbol{X})Y - 2g_T(\boldsymbol{B}'\boldsymbol{X})Y + 2g_T(\boldsymbol{B}'\boldsymbol{X})\mu(\boldsymbol{X})\Big] \\ = &\underset{g_t\in\mathcal{H}^{(\boldsymbol{B})}, \boldsymbol{B}\in\Theta_q}{\arg\min} \;\mathbb{E}\Big[Y^2 + \big(g_T(\boldsymbol{B}'\boldsymbol{X})\big)^2 - 2g_T(\boldsymbol{B}'\boldsymbol{X})Y + 2g_T(\boldsymbol{B}'\boldsymbol{X})\mu(\boldsymbol{X})\Big] \\ = &\underset{g_t\in\mathcal{H}^{(\boldsymbol{B})}, \boldsymbol{B}\in\Theta_q}{\arg\min} \;\mathbb{E}\Big[Y^2 + \big(g_T(\boldsymbol{B}'\boldsymbol{X})\big)^2 - 2g_T(\boldsymbol{B}'\boldsymbol{X})Y + 2\mathbb{E}\big[g_T(\boldsymbol{B}'\boldsymbol{X})\mu(\boldsymbol{X})|\boldsymbol{X}\big]\Big] \\ = &\underset{g_t\in\mathcal{H}^{(\boldsymbol{B})}, \boldsymbol{B}\in\Theta_q}{\arg\min} \;\mathbb{E}\Big[Y^2 + \big(g_T(\boldsymbol{B}'\boldsymbol{X})\big)^2 - 2g_T(\boldsymbol{B}'\boldsymbol{X})Y\Big],\end{aligned}$$

where the first equality follows from the fact that $\mu(\boldsymbol{X})$ is not a component that we optimize over, the second equality follows from an application of the iterated expectation rule to condition on $\boldsymbol{X}$ and the last equality follows from the constraint $\mathbb{E}\left[g_T(\boldsymbol{B}'\boldsymbol{X})|\boldsymbol{X}\right] = 0$ in (A.3). Therefore, representation (A.3) can be simplified to (2.8) of the main manuscript, which does not involve the unspecified term $\mu(\boldsymbol{X})$ of the underlying model.

A.4 Proof of Proposition 3.1

Proof. Note that $\boldsymbol{\eta}_t - \bar{\boldsymbol{\eta}} \in \text{span}(\boldsymbol{\Xi})$ and hence $(\boldsymbol{\eta}_t - \bar{\boldsymbol{\eta}})'\boldsymbol{X}$ is measurable with respect to $\boldsymbol{X}'\boldsymbol{\Xi}$. If model (3.1) holds, then

$$\begin{aligned}
\mathcal{C}\big(\boldsymbol{X}'\boldsymbol{\Xi};\boldsymbol{c}\big) &= \sum_{t=1}^{K} c_t \mathbb{E}[Y \mid \boldsymbol{X}'\boldsymbol{\Xi}, T=t] \\
&= \sum_{t=1}^{K} c_t \mathbb{E}[\mathbb{E}[Y|\boldsymbol{X}, T=t] \mid \boldsymbol{X}'\boldsymbol{\Xi}, T=t] \\
&= \sum_{t=1}^{K} c_t \mathbb{E}[\mu_0(\boldsymbol{X}) + \boldsymbol{\eta}_t'\boldsymbol{X} \mid \boldsymbol{X}'\boldsymbol{\Xi}, T=t] \ \text{ by (3.1)} \\
&= \sum_{t=1}^{K} c_t \mathbb{E}[(\boldsymbol{\eta}_t - \bar{\boldsymbol{\eta}})'\boldsymbol{X} \mid \boldsymbol{X}'\boldsymbol{\Xi}, T=t] \ \text{ (by zero-sum constraint on contrast } \boldsymbol{c}\text{)} \\
&= \sum_{t=1}^{K} c_t (\boldsymbol{\eta}_t - \bar{\boldsymbol{\eta}})'\boldsymbol{X} \ \text{ (by the measurability condition)} \\
&= \sum_{t=1}^{K} c_t (\mu_0(\boldsymbol{X}) + \boldsymbol{\eta}_t'\boldsymbol{X}) \ \text{ (by the zero-sum constraint on contrast } \boldsymbol{c}\text{)} \\
&= \sum_{t=1}^{K} c_t \mathbb{E}[Y \mid \boldsymbol{X}, T=t] \ \text{ (by (3.1))} \\
&= \mathcal{C}\left(\boldsymbol{X};\boldsymbol{c}\right).
\end{aligned}$$

That $(\boldsymbol{\eta}_1, \ldots, \boldsymbol{\eta}_K)$ are distinct and $\pi_t > 0$ is sufficient to guarantee that there are $K-1$ nonzero eigenvalues in the matrix $\boldsymbol{H}$ in (3.2). Since the "between" group dispersion matrix $\boldsymbol{H}$ in (3.2) has $K-1$ nonzero eigenvalues and the rank of $\boldsymbol{\Xi}$ is $K-1$, it is clear span$(\boldsymbol{\Xi}) = S_{\mathcal{C}|\boldsymbol{X}}$.

□

A.5 Proof of Proposition 3.2

Proof. Let Y_t denote Y given $T = t$ $(t = 1, \ldots, K)$, i.e., the T-specific outcome. For a given $\boldsymbol{\beta}$, consider the expression:

$$\mathbb{E}\left[(Y - \gamma_T\boldsymbol{\beta}'\boldsymbol{X})^2\right] = \sum_{t=1}^{K} \pi_t \mathbb{E}\left[(Y_t - \gamma_t\boldsymbol{\beta}'\boldsymbol{X})^2 1_{(T=t)}\right] = \sum_{t=1}^{K} \pi_t \mathbb{E}\left[(Y_t - \gamma_t\boldsymbol{\beta}'\boldsymbol{X})^2\right],$$

which can be minimized by minimizing each of the K terms with respect to γ_t $(t = 1, \ldots, K)$ separately. For the uncentered $\tilde{\gamma}_t$, standard least-squares theory gives the solution as

$$\tilde{\gamma}_t = \frac{\text{cov}(\boldsymbol{\beta}'\boldsymbol{X}, Y_t)}{\text{var}(\boldsymbol{\beta}'\boldsymbol{X})} = \frac{\boldsymbol{\beta}'\text{cov}(\boldsymbol{X}, Y_t)}{\boldsymbol{\beta}'\boldsymbol{\Sigma}_X\boldsymbol{\beta}} \quad (t = 1, \ldots, K).$$

Because $\boldsymbol{X}$ is centered and Y_t is centered within each treatment t, the covariance in the numerator can be written as

$$\text{cov}(\boldsymbol{X}, Y_t) = \mathbb{E}[\boldsymbol{X}Y_t] = \mathbb{E}[\boldsymbol{X}\mathbb{E}[Y_t|\boldsymbol{X}]] = \mathbb{E}[\boldsymbol{X}\boldsymbol{X}'\boldsymbol{\eta}_t] = \mathbb{E}[\boldsymbol{X}\boldsymbol{X}']\boldsymbol{\eta}_t = \boldsymbol{\Sigma}_X\boldsymbol{\eta}_t,$$

and hence

$$\tilde{\gamma}_t = \frac{\boldsymbol{\beta}'\boldsymbol{\Sigma}_X\boldsymbol{\eta}_t}{\boldsymbol{\beta}'\boldsymbol{\Sigma}_X\boldsymbol{\beta}} \quad (t = 1, \ldots, K).$$

Centering the $\tilde{\gamma}_t$ finishes the proof.

□

A.6 Proof of Proposition 3.4

Proof. This equivalency, presented in Proposition 3.4, follows from Proposition 3.3 that gives an explicit expression of the minimizer $(\gamma_1, \gamma_2, \boldsymbol{\beta})$ of (3.6) in terms of the population parameters in (3.1), and the expression of $\boldsymbol{\xi}_1$ available in a closed form.

Consider the criterion of (3.6) at the minimum:

$$\begin{aligned}(**) &= \min_{(\gamma_1,\gamma_2,\boldsymbol{\beta})} \mathbb{E}[(Y - \boldsymbol{X}'\boldsymbol{\beta}\gamma_t)^2] \\ &= \min_{(\gamma_1,\gamma_2,\boldsymbol{\beta})} \pi_1\mathbb{E}[(Y - \boldsymbol{X}'\boldsymbol{\beta}\gamma_1)^2 \mid T = 1] + (1-\pi_1)\mathbb{E}[(Y - \boldsymbol{X}'\boldsymbol{\beta}\gamma_2)^2 \mid T = 2]\end{aligned} \tag{A.4}$$

By Proposition 3.3, the minimum $(**)$ occurs at $\boldsymbol{\beta} = \boldsymbol{\xi}_1$ and $\gamma_t = (\boldsymbol{\xi}_1{}'\boldsymbol{\Sigma}_X\boldsymbol{\xi}_1)^{-1}\boldsymbol{\xi}_1{}'\boldsymbol{\Sigma}_X(\boldsymbol{\eta}_t - \bar{\boldsymbol{\eta}}) = (\boldsymbol{\xi}_1{}'\boldsymbol{\Sigma}_X\boldsymbol{\xi}_1)^{-1}\boldsymbol{\xi}_1{}'\boldsymbol{\Sigma}_X(\boldsymbol{\eta}_t - \{\pi_1\boldsymbol{\eta}_1 + (1-\pi_1)\boldsymbol{\eta}_2\})$ $(a = 1, 2)$, that is:

$$\begin{aligned}\gamma_1 &= (\boldsymbol{\xi}_1{}'\boldsymbol{\Sigma}_X\boldsymbol{\xi}_1)^{-1}\boldsymbol{\xi}_1{}'\boldsymbol{\Sigma}_X(\boldsymbol{\eta}_2 - \boldsymbol{\eta}_1)(\pi_1 - 1) = \|\boldsymbol{\eta}_2 - \boldsymbol{\eta}_1\|(\pi_1 - 1) \quad \text{and} \\ \gamma_2 &= (\boldsymbol{\xi}_1{}'\boldsymbol{\Sigma}_X\boldsymbol{\xi}_1)^{-1}\boldsymbol{\xi}_1{}'\boldsymbol{\Sigma}_X(\boldsymbol{\eta}_2 - \boldsymbol{\eta}_1)\pi_1 = \|\boldsymbol{\eta}_2 - \boldsymbol{\eta}_1\|\pi_1,\end{aligned} \tag{A.5}$$

which follows from $\boldsymbol{\xi}_1 = (\boldsymbol{\eta}_2 - \boldsymbol{\eta}_1)/\|\boldsymbol{\eta}_2 - \boldsymbol{\eta}_1\|$. Plugging (A.5) and $\boldsymbol{\beta}(= \boldsymbol{\xi}_1) = (\boldsymbol{\eta}_2 - \boldsymbol{\eta}_1)/\|\boldsymbol{\eta}_2 - \boldsymbol{\eta}_1\|$ into the second line of (A.4) gives:

$$\begin{aligned}(**) =& \pi_1\mathbb{E}[(Y - \boldsymbol{X}'(\boldsymbol{\eta}_2 - \boldsymbol{\eta}_1)(\pi_1 - 1))^2 \mid T = 1] + (1-\pi_1)\mathbb{E}[(Y - \boldsymbol{X}'(\boldsymbol{\eta}_2 - \boldsymbol{\eta}_1)\pi_1)^2 \mid T = 2] \\ =& \pi_1\mathbb{E}[(Y - \boldsymbol{X}'\boldsymbol{\beta}(\pi_1 - 1))^2 \mid T = 1] + (1-\pi_1)\mathbb{E}[(Y - \boldsymbol{X}'\boldsymbol{\beta}\pi_1)^2 \mid T = 2] \\ =& \pi_1\mathbb{E}[(Y - \boldsymbol{X}'\boldsymbol{\beta}(T + \pi_1 - 2))^2 \mid T = 1] + (1-\pi_1)\mathbb{E}[(Y - \boldsymbol{X}'\boldsymbol{\beta}(T + \pi_1 - 2))^2 \mid T = 2] \\ =& \mathbb{E}[(Y - \boldsymbol{X}'\boldsymbol{\beta}(T + \pi_1 - 2))^2],\end{aligned} \tag{A.6}$$

in which we set $\boldsymbol{\beta} = (\boldsymbol{\eta}_2 - \boldsymbol{\eta}_1) \in \mathbb{R}^p$. The last line of (A.6) is the least squares criterion on the right-hand side of (3.10) associated with $\boldsymbol{\beta}^*$ of model (3.8). Since the minimum $(**)$ (A.4) is unique, it follows that $\boldsymbol{\beta}^* = (\boldsymbol{\eta}_2 - \boldsymbol{\eta}_1)$, which is proportional to $\boldsymbol{\xi}_1 = (\boldsymbol{\eta}_2 - \boldsymbol{\eta}_1)/\|\boldsymbol{\eta}_2 - \boldsymbol{\eta}_1\|$. □

A.7 Depression treatment study: Boxplots of the estimated Values of ITRs with a larger number of pretreatment covariates

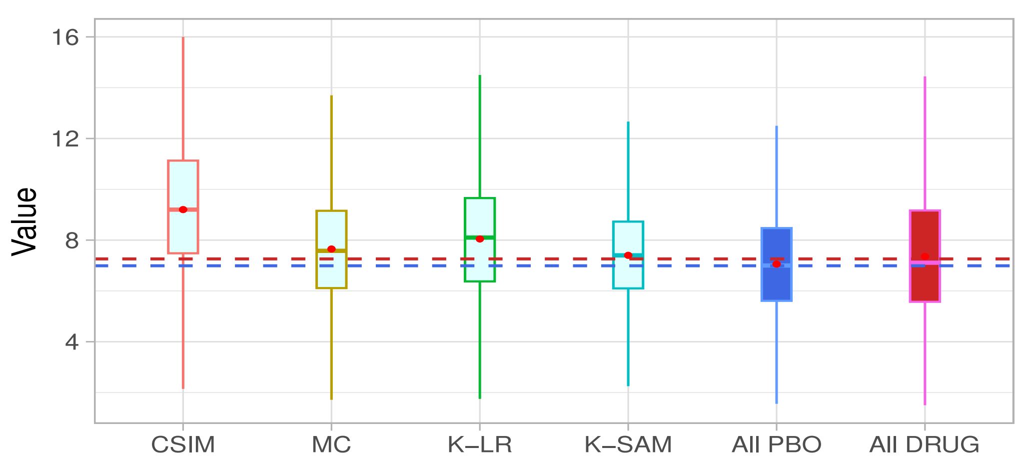

Figure A1: Comparison of the boxplots of the estimated Values (5.1) of the treatment decision rules (ITRs) (the 4 methods [CSIM, MC, K-LR, K-SAM] considered in the main manuscript and the two naive rules of assigning everyone placebo [All PBO] and everyone the active drug [All DRUG]), obtained from 500 randomly split testing sets. Higher Values are preferred.

Code (Github Repository)

http://github.com/syhyunpark/hd-csim