Web Appendix for 'Propensity Score Weighting for Subgroup Analysis" by Yang et al.

1 | PROOFS

1.1 | Proofs of Properties of Overlap Weights in Subgroups

We assume regularity conditions on $v_z = V[Y(z)|X, S]$ and $E[Y(z)|X, S]$ necessary to make the integrals defined and convergent.

(1) $\hat{\tau}_{r,h}$ is consistent for $\tau_{r,h}$.

Proof. The S-WATE for the population with density proportional to $f(x, s)h(x, s)$ with respect to base measure μ is defined as

$$\begin{aligned}
\tau_{r,h} &= \frac{\mathbb{E}[h(x,s)(Y(1)-Y(0))|S_r=1]}{\mathbb{E}[h(x,s)|S_r=1]} \\
&= \frac{\int \mathbb{E}_{Y,Z|X,S}\left\{Y(1)ZS_r[h(x,s)/e(x,s)] - Y(0)(1-Z)S_r[h(x,s)/(1-e(x,s))]\right\} f(x,s)\mu(dx,s)}{\int h(x,s)f(x,s)\mu(dx,s)} \\
&= \frac{\int \mathbb{E}_{Y,Z|X,S}\, Y(1)ZS_r[h(x,s)/e(x,s)]f(x,s)\mu(dx,s)}{\int \mathbb{E}_{Z|X,S}\, ZS_r[h(x,s)/e(x,s)]f(x,s)\mu(dx,s)} \\
&\quad - \frac{\int \mathbb{E}_{Y,Z|X,S}\, Y_0(1-Z)S_r[h(x,s)/e(x,s)]f(x,s)\mu(dx,s)}{\int \mathbb{E}_{Z|X,S}(1-Z)S_r[h(x,s)/e(x,s)]f(x,s)\mu(dx,s)}
\end{aligned} \tag{1}$$

Under the unconfoundedness assumption that $Y(1), Y(0) \perp Z \mid X, S$. The terms of (1) can be read as expectations of weighted means of $Y(z)$ in samples drawn from the population with density $f(x, s)$ respectively for the strata with $z = 1$ or $z = 0$, given $S_r = 1$. Replacing expectations by sample means, and substituting weight expressions from (**??**), we obtain the following estimator for the sample S-WATE:

$$\hat{\tau}_{r,h} = \frac{\sum_i Y_i(1)Z_iS_{ir}w_1(x_i,s_i)}{\sum_i Z_iS_{ir}w_1(x_i,s_i)} - \frac{\sum_i Y_i(0)(1-Z_i)S_{ir}w_0(x_i,s_i)}{\sum_i(1-Z_i)S_{ir}w_0(x_i,s_i)} \tag{2}$$

where each summation (divided by n) is an unbiased estimator of the corresponding integral in (1); therefore by Slutsky's theorem $\hat{\tau}_{r,h}$ is a consistent estimator of $\tau_{r,h}$. □

(2) $h(X_i, S_i) = e(X_i, S_i)\{1 - e(X_i, S_i)\}$ gives the smallest variance of the weighted estimator $\hat{\tau}_{r,h}$ over all possible h under homoscedasticity.

Proof. The variance of the estimator $\hat{\tau}_{r,h}$ is

$$\begin{aligned}
\mathbb{V}[\hat{\tau}_{r,h} \mid \mathbf{X}, S, z] &= \frac{\sum_i S_1(x_i,s_i)z_iS_{ir}w_1(x_i,s_i)^2}{\left[\sum_i z_iS_{ir}w_1(x_i,s_i)\right]^2} + \frac{\sum_i v_0(x_i,s_i)(1-z_i)S_{ir}w_0(x_i,s_i)^2}{\left[\sum_i(1-z_i)S_{ir}w_0(x_i,s_i)\right]^2} \\
&= \frac{\sum_i S_1(x_i,s_i)[z_iS_{ir}/e(x_i,s_i)][h(x_i,s_i)^2/e(x_i,s_i)]}{\left\{\sum_i[z_iS_{ir}/e(x_i,s_i)]h(x_i,s_i)\right\}^2} + \\
&\qquad \frac{\sum_i v_0(x_i,s_i)[(1-z_i)S_{ir}/(1-e(x_i,s_i))][h(x_i,s_i)^2/(1-e(x_i,s_i))]}{\left\{\sum_i[(1-z_i)S_{ir}/(1-e(x_i,s_i))]h(x_i,s_i)\right\}^2},
\end{aligned} \tag{3}$$

where $v_z = \mathbb{V}[Y(z)|X, S]$. Averaging the above first over the distribution of $\mathbf{Z}$ (using $\mathbb{E}[Z_iS_{ir}/e(x_i, s_i)] = \mathbb{E}[(1-Z_i)S_{ir}(1-e(x_i, s_i))] = 1$), and then over the joint distribution of $(\mathbf{X}, S)$, and again applying Slutsky's theorem, we have

$$n \cdot \mathbb{E}_x\, \mathbb{V}[\hat{\tau}_{r,h} \mid \mathbf{X}, S] \to \frac{\int \left(\frac{S_1(x,s)}{e(x,s)} + \frac{v_0(x,s)}{1-e(x,s)}\right) h(x,s)^2 f(x,s)\mu(dx,s)}{\left(\int f(x,s)h(x,s)\mu(dx,s)\right)^2}$$

If the residual variance is assumed to be homoscedastic across both groups, $v_1(x,s) = v_0(x,s) = v$, the above equation simplifies to

$$n \cdot \mathbb{E}_x \mathbb{V}[\hat{\tau}_{r,h} \mid \mathbf{X}, S] \to v/C_h^2 \int \frac{f(x,s)h(x,s)^2\mu(dx,s)}{e(x,s)(1-e(x,s))} = v/C_h^2\, \mathbb{E}\left\{\frac{h^2(x,s)}{e(x,s)(1-e(x,s))}\right\}. \tag{4}$$

According to the Cauchy-Schwarz inequality, we have

$$\begin{aligned}
[\mathbb{E}\{h(x,s)\}]^2 &= \left[\mathbb{E}\left\{\frac{h(x,s)}{\sqrt{e(x,s)(1-e(x,s))}}\sqrt{e(x,s)(1-e(x,s))}\right\}\right]^2 \\
&\le \mathbb{E}\left\{\frac{h^2(x,s)}{e(x,s)(1-e(x,s))}\right\}\mathbb{E}\left[e(x,s)(1-e(x,s))\right],
\end{aligned}$$

and the equality is attained when $h(x,s) \propto e(x,s)(1-e(x,s))$. Property (2) follows directly from applying the above to the right hand side of (4). □

1.2 | Proof of Proposition 1

Proof. The data generating law is based on $f(x)$ and the bias property should always be stated with respect to $\mathbb{E}(\cdot)$. Suppose Condition (3) in Proposition 1 holds. For any weight w_i (including $\widehat{w}_i^*$), we have

$$\begin{aligned}
\left|\mathbb{E}\left(\widehat{\tau}_{r,h} - \tau_r\right)\right| &= \left|\mathbb{E}\left[\sum_{i=1}^N Z_i S_{ir} w_i Y_i - \sum_{i=1}^N (1-Z_i) S_{ir} w_i Y_i\right] - \tau_r\right| \\
&= \Bigg|\beta_r\left[\sum_{i=1}^N Z_i S_{ir} w_i - \sum_{i=1}^N (1-Z_i) S_{ir} w_i\right] + \\
&\qquad \sum_{p=1}^P \beta_{rp}\left[\sum_{i=1}^N Z_i S_{ir} w_i X_{ip} - \sum_{i=1}^N (1-Z_i) S_{ir} w_i X_{ip}\right] + \tau_r\left[\sum_{i=1}^N Z_i S_{ir} w_i - 1\right]\Bigg| \\
&= \left|\sum_{p=1}^P \beta_{rp}\left[\sum_{i=1}^N Z_i S_{ir} w_i X_{ip} - \sum_{i=1}^N (1-Z_i) S_{ir} w_i X_{ip}\right]\right| \\
&\qquad \text{(Normalized weights that sum to 1 among the treated in the subgroup.)} \\
&\le \sum_{p=1}^P |\beta_{rp}| \left|\sum_{i=1}^N Z_i S_{ir} w_i X_{ip} - \sum_{i=1}^N (1-Z_i) S_{ir} w_i X_{ip}\right| \qquad \text{(Triangular inequality.)} \\
&< \delta \sum_{p=1}^P |\beta_{rp}|.
\end{aligned}$$

□

1.3 | Proof of Proposition 2

Proof. The treatment effect is heterogeneous within subgroups and the subgroup average treatment effect is

$$\tau_{r,h} = \tau_r + \sum_{p=1}^P \gamma_{rp} \frac{\mathbb{E}[h(X,S)X_p | S_r = 1]}{\mathbb{E}[h(X,S) | S_r = 1]}.$$

Suppose Condition (3) in Proposition 1 and Condition (4) in Proposition 2 hold, similar to the proof of Proposition 1, for any weight w_i (including $\widehat{w}_i^*$),

$$
\begin{aligned}
\left|\mathbb{E}\left(\widehat{\tau}_{r,h}-\tau_{r,h}\right)\right| &= \left|\mathbb{E}\left[\sum_{i=1}^{N} Z_i S_{ir} w_i Y_i - \sum_{i=1}^{N}(1-Z_i)S_{ir}w_i Y_i\right] - \tau_{r,h}\right| \\
&= \Bigg|\beta_r\left[\sum_{i=1}^{N} Z_i S_{ir} w_i - \sum_{i=1}^{N}(1-Z_i)S_{ir}w_i\right] + && (5)\\
&\quad \sum_{p=1}^{P}\beta_{rp}\left[\sum_{i=1}^{N} Z_i S_{ir} w_i X_{ip} - \sum_{i=1}^{N}(1-Z_i)S_{ir}w_i X_{ip}\right] + && (6)\\
&\quad \tau_r\left[\sum_{i=1}^{N} Z_i S_{ir} w_i - 1\right] + && (7)\\
&\quad \sum_{p=1}^{P}\gamma_{rp}\left[\sum_{i=1}^{N} Z_i S_{ir} w_i X_{ip} - \frac{\sum_{i=1}^{N} h(X_i,S_i)S_{ir}X_{ip}}{\sum_{i=1}^{N} h(X_i,S_i)S_{ir}}\right] + && (8)\\
&\quad \sum_{p=1}^{P}\gamma_{rp}\left[\frac{\sum_{i=1}^{N} h(X_i,S_i)S_{ir}X_{ip}}{\sum_{i=1}^{N} h(X_i,S_i)S_{ir}} - \frac{\mathbb{E}[h(X,S)X_p|S_r=1]}{\mathbb{E}[h(X,S)|S_r=1]}\right]\Bigg|, && (9)
\end{aligned}
$$

where equation (5) = 0 by the definition of weights, equation (7) = 0 by design of weights, and the expectation of equation (9) = 0.

It follows

$$
\left|\mathbb{E}\left(\widehat{\tau}_{r,h}-\tau_{r,h}\right)\right| = (6)+(8) \le \delta\sum_{p=1}^{P}|\beta_{rp}| + \delta_2\sum_{p=1}^{P}|\gamma_{rp}|.
$$

□

1.4 | Proof of Proposition 2b

Proof. If the causal estimand in Proposition 2 is the subgroup sample weighted average treatment effect (S-SWATE),

$$
\tau_{r,\widehat{h}} = \frac{\sum_i \widehat{h}(X_i,S_i)S_{ir}[Y_i(1)-Y_i(0)]}{\sum_i \widehat{h}(X_i,S_i)S_{ir}},
$$

then (8) becomes

$$
\sum_{p=1}^{P}\gamma_{rp}\left[\sum_{i=1}^{N} Z_i S_{ir} w_i X_{ip} - \frac{\sum_{i=1}^{N}\widehat{h}(X_i,S_i)S_{ir}X_{ip}}{\sum_{i=1}^{N}\widehat{h}(X_i,S_i)S_{ir}}\right] \tag{10}
$$

It follows

$$
\left|\mathbb{E}\left(\widehat{\tau}_{r,\widehat{h}}-\tau_{r,\widehat{h}}\right)\right| = (6)+(10) \le \delta\sum_{p=1}^{P}|\beta_{rp}| + \delta_3\sum_{p=1}^{P}|\gamma_{rp}|.
$$

The quantity in (10) can be calculated. Note for the class of balancing weights, frequently the weights are estimated from the propensity score model, therefore

$$
\sum_{i=1}^{N} Z_i S_{ir}\widehat{w}_i X_{ip} - \frac{\sum_{i=1}^{N}\widehat{h}(X_i,S_i)S_{ir}X_{ip}}{\sum_{i=1}^{N}\widehat{h}(X_i,S_i)S_{ir}}
$$

is generally small because both terms are different estimates of the same thing.

□

1.5 Proof of Proposition 3

Proposition 3 is a corollary of Theorem 1 in Li et al.,[1] which proved the exact balance property of overlap weights in the overall sample. Below we will first reproduce the proof of Theorem 1 and then extend it to subgroups as in Proposition 3.

Theorem 1. *When the propensity scores are estimated by maximum likelihood under a logistic regression model,* $\text{logit}\{e(X_i)\} = \beta_0 + X_i\beta'$*, the overlap weights lead to exact balance in the means of any included covariate between treatment and control groups. That is, for any covariate* j*, we have*

$$\frac{\sum_i X_{ij} Z_i (1-\hat{e}_i)}{\sum_i Z_i(1-\hat{e}_i)} = \frac{\sum_i X_{ij}(1-Z_i)\hat{e}_i}{\sum_i (1-Z_i)\hat{e}_i}, \quad \text{for } j = 1, \dots, P, \tag{11}$$

where $\hat{e}_i = \{1+\exp[-(\hat{\beta}_0 + X_i\hat{\beta}')]\}^{-1}$ *and* $\hat{\beta} = (\hat{\beta}_1, ..., \hat{\beta}_P)$ *is the MLE for the regression coefficients.*

Proof. The score functions of the logistic propensity score model, $\text{logit}\{e(x_i;\beta)\} = \beta_0 + x_i\beta'$ with $\beta = (\beta_1, ..., \beta_P)$, are:

$$\frac{\partial \log L}{\partial \beta_j} = \sum_i x_{ij}(Z_i - e_i), \quad \text{for } j = 0, 1, \dots, P, \tag{12}$$

where $x_{0j} \equiv 1$ and $e_i \equiv e(x_i;\beta) = \left[1+\exp(-(\beta_0 + x_i\beta')\right]^{-1}$.

Equate the score functions to 0 and solve for the MLE $\hat{\beta}$, we have

$$\sum Z_i = \sum \hat{e}_i, \quad \text{and} \quad \sum x_{ij} Z_i = \sum x_{ij}\hat{e}_i.$$

where $= e(x_i;\hat{\beta})$. It follows that

$$\sum_i Z_i(1-\hat{e}_i) = \sum \hat{e}_i - \sum_i Z_i\hat{e}_i = \sum \hat{e}_i(1-Z_i), \tag{13}$$

$$\sum_i x_{ij} Z_i(1-\hat{e}_i) = \sum x_{ij}\hat{e}_i - \sum_i x_{ij} Z_i\hat{e}_i = \sum x_{ij}\hat{e}_i(1-Z_i), \tag{14}$$

for $j = 1, \dots, P$. Therefore, for any j, we have

$$\frac{\sum_i x_{ij} Z_i(1-\hat{e}_i)}{\sum_i Z_i(1-\hat{e}_i)} = \frac{\sum_i x_{ij}(1-Z_i)\hat{e}_i}{\sum_i (1-Z_i)\hat{e}_i}. \tag{15}$$

□

Corollary (i.e. Proposition 3): If the postulated logistic model for propensity score, $\text{logit}\{e(X_i)\} = \beta_0 + X_i\beta'$, includes any interaction term of a binary covariate as a predictor, then the overlap weights lead to exact balance in the means in the subgroups defined by that binary covariate.

Proof. For simplicity, consider the case where there are only two covariates X_1 and X_2, where X_1 is binary. The interaction term is $X_1X_2 = X_2$ for units with $X_1 = 1$, and $X_1X_2 = 0$ for units with $X_1 = 0$. If the postulated propensity score model include the interaction term X_1X_2 as a predictor, e.g. $\text{logit}\{e(X_i)\} = \beta_0 + X_{i1}\beta_1 + X_{i2}\beta_2 + X_{i1}X_{i2}\beta_{12}$, then

i For units i with $X_{i1} = 0$, the exact balance of the interaction term trivially stands because the interaction term equals zero for these units.

ii For units i with $X_{i1} = 1$, the exact balance of the interaction term also holds by directly applying the interaction term to Equation (14) (and thus Equation (15)), which becomes:

$$\frac{\sum_{i:X_{i1}=1} X_{i2} Z_i(1-\hat{e}_i)}{\sum_i Z_i(1-\hat{e}_i)} = \frac{\sum_{i:X_{i1}=1} X_{i2}(1-Z_i)\hat{e}_i}{\sum_i (1-Z_i)\hat{e}_i}.$$

□

The *absolute standardized mean difference*(ASMD)[2] is widely used for measuring covariate balance. The ASMD is the difference in weighted means further scaled by the pooled, weighted standard deviation. That is

$$\text{ASMD}_{r,p} = \frac{\sum_{i=1}^N Z_i S_{ir} w_i X_{ip} - \sum_{i=1}^N (1-Z_i) S_{ir} w_i X_{ip}}{s_{r,p}} \tag{16}$$

where $s_{r,p}$ is the weighted, pooled standard deviation for the r^{th} subgroup and the p^{th} covariate. Following the literature,[2] we define the subgroup weighted mean within arm $z = 0, 1$ as $\bar{X}^w_{p,1} = \sum_i Z_i w_i S_{ir} X_{ip} / \sum_i Z_i w_i S_{ir}$ and $\bar{X}^w_{p,0} = \sum_i (1 - Z_i) w_i S_{ir} X_{ip} / \sum_i (1 - Z_i) w_i S_{ir}$, respectively, and the subgroup weighted variance of X_p within arm $z = 0, 1$ as

$$s^2_{r,p,1} = \frac{\sum_i Z_i w_i S_{ir}}{(\sum_i Z_i w_i S_{ir})^2 - \sum_i (Z_i w_i S_{ir})^2} \sum_i Z_i w_i S_{ir} (X_{ip} - \bar{X}^w_{p,1})^2,$$

and

$$s^2_{r,p,0} = \frac{\sum_i (1 - Z_i) w_i S_{ir}}{(\sum_i (1 - Z_i) w_i S_{ir})^2 - \sum_i ((1 - Z_i) w_i S_{ir})^2} \sum_i (1 - Z_i) w_i S_{ir} (X_{ip} - \bar{X}^w_{p,0})^2.$$

1.6 | Proposition 4 (Non-linear covariates)

Suppose the regression of Y on X is non-linear and the outcome surface satisfies a generalized additive model

$$Y_i(z) = \sum_{r=1}^{R} \beta_r S_{ir} + \sum_{r=1}^{R} \sum_{p=1}^{P} f_{rp}(X_{ip}) S_{ir} + \sum_{r=1}^{R} \lambda_r S_{ir} z + \sum_{p=1}^{P} g_{rp}(X_{ip}) S_{ir} z + \epsilon_i(z),$$

where each $f_{rp}(X), g_{rp}(X)$ is a K-times differentiable, non-linear function at all X_{ip}, with $f^{(K)}_{rp} \le C_1, g^{(K)}_{rp} \le C_2$, for all r, p, and $E(\epsilon_i | X_{i,i}) = 0$. If balance holds in higher order terms for the transformed covariates $\tilde{X}^k_{ijp}$, i.e.,

$$\left| \sum_{i=1}^{N} Z_i S_{ir} w_i \tilde{X}^k_{ijp} - \sum_{i=1}^{N} (1 - Z_i) S_{ir} w_i \tilde{X}^k_{ijp} \right| < \delta_{1,pk}, \tag{17}$$

And if

$$\left| \sum_{i=1}^{N} Z_i S_{ir} w_i \tilde{X}^k_{ijp} - \frac{\sum_{i=1}^{N} h(X_i, S_i) S_{ir} \tilde{X}^k_{ijp}}{\sum_{i=1}^{N} h(X_i, S_i) S_{ir}} \right| < \delta_{2,pk}, \tag{18}$$

for each $k = 1, \dots, K$, $j = 1, 2, \dots, M_p/l_p$ and $p = 1, \dots, P$. Then the bounding bias is

$$\mathbb{E}\left(\hat{\tau}_{r,h} - \tau_{r,h}\right) < \delta_{1,pk} \sum_{p=1}^{P} \sum_{k=0}^{K} \sum_{j=1}^{M_p/l_p} |\eta_{rjpk}| + \delta_{2,pk} \sum_{p=1}^{P} \sum_{k=0}^{K-1} \sum_{j=1}^{M_p/l_p} |\phi_{rjpk}| + 2P \frac{M_p}{l_p} \frac{C_1}{K!} (l_p/2)^K + 2P \frac{M_p}{l_p} \frac{C_2}{K!} (l_p/2)^K,$$

where η_{rjpk}, ϕ_{rjpk}'s are the coefficients of the Taylor expansion of $f_{rp}(X_{ip}), g_{rp}(X_{ip})$ at order k around ξ_{jp}.

***Proof*.** Similar to Zubizarreta,[3] for each X_{ip} we break its support $[-M_p/2, M_p/2] \in R$ into M_p/l_p disjoint intervals of length l_p and midpoint ξ_{jp}, and define the transformed piecewise covariates centered around ξ_{jp} as $\tilde{X}_{ijp} = (X_{ip} - \xi_{jp}) 1_{X_{ip} \in [\xi_{jp} - l_p/2, \xi_{jp} + l_p/2]}$.

The k^{th} order Taylor expansion of $f_{rp}(X_{ip}), g_{rp}(X_{ip})$ at ξ_{jp} is

$$f_{rp}(X_{ip}) = \sum_{k=0}^{K-1} \frac{f^{(k)}_{rp}(\xi_{jp})}{k!} (X_{ip} - \xi_{jp})^k + \frac{f^{(K)}_{rp}(\xi'_{jp})}{K!} (X_{ip} - \xi_{jp})^K = \sum_{j=1}^{M_p/l_p} \left(\sum_{k=0}^{K-1} \eta_{rjpk} \tilde{X}^k_{ijp} + R_{rijpK} \right)$$
$$g_{rp}(X_{ip}) = \sum_{k=0}^{K-1} \frac{g^{(k)}_{rp}(\xi_{jp})}{k!} (X_{ip} - \xi_{jp})^k + \frac{g^{(K)}_{rp}(\xi'_{jp})}{K!} (X_{ip} - \xi_{jp})^K = \sum_{j=1}^{M_p/l_p} \left(\sum_{k=0}^{K-1} \phi_{rjpk} \tilde{X}^k_{ijp} + T_{rijpK} \right),$$

for some ξ'_{jp} between ξ_{jp} and X_{ip}, $\eta_{rjpk} = \frac{f^{(k)}_{rp}(\xi_{jp})}{k!}$, and $\phi_{rjpk} = \frac{g^{(k)}_{rp}(\xi_{jp})}{k!}$. By the Lagrange Error Bound, $R_{rijpK} = \frac{f^{(K)}_{rp}(\xi'_{jp})}{K!} \tilde{X}^K_{ijp} \le |\frac{f^{(K)}_{rp}(\xi'_{jp})}{K!} (l_p/2)^K| \le \frac{C_1}{K!} (l_p/2)^K$, and $T_{rijpK} = \frac{g^{(K)}_{rp}(\xi'_{jp})}{K!} \tilde{X}^K_{ijp} \le |\frac{g^{(K)}_{rp}(\xi'_{jp})}{K!} (l_p/2)^K| \le \frac{C_2}{K!} (l_p/2)^K$ for all $r = 1, \dots, R; j = 1, \dots, Mp/lp; p = 1, \dots, P$.

Our target estimand is

$$\tau_{r,h} = \frac{\mathbb{E}\left\{ h(\mathbf{X},) \left[\lambda_r + \sum_{p=1}^{P} g_{rp}(X_p) | S_r = 1 \right] \right\}}{\mathbb{E}[h(\mathbf{X},) | S_r = 1]}$$

Then

$$
\begin{aligned}
\left|\mathbb{E}\left(\widehat{\tau}_{r,h}-\tau_{r,h}\right)\right| &= \left|\mathbb{E}\left[\sum_{i=1}^{N} Z_i S_{ir} w_i Y_i - \sum_{i=1}^{N}(1-Z_i) S_{ir} w_i Y_i\right] - \tau_{r,h}\right| \\
&= \left|\sum_{i=1}^{N} Z_i S_{ir} w_i \left(\beta_r + \sum_{p=1}^{P} f_{rp}(X_{ip}) + \lambda_r + \sum_{p=1}^{P} g_{rp}(X_{ip})\right) - \right. \\
&\qquad \left. \sum_{i=1}^{N}(1-Z_i) S_{ir} w_i \left(\beta_r + \sum_{p=1}^{P} f_{rp}(X_{ip})\right) - \tau_{r,h}\right| \\
&= \left|\beta_r \left[\sum_{i=1}^{N} Z_i S_{ir} w_i - \sum_{i=1}^{N}(1-Z_i) S_{ir} w_i\right] + \right. && (19) \\
&\qquad \sum_{p=1}^{P}\left[\sum_{i=1}^{N} Z_i S_{ir} w_i f_{rp}(X_{ip}) - \sum_{i=1}^{N}(1-Z_i) S_{ir} w_i f_{rp}(X_{ip})\right] + && (20) \\
&\qquad \lambda_r \left[\sum_{i=1}^{N} Z_i S_{ir} w_i - 1\right] + && (21) \\
&\qquad \sum_{p=1}^{P}\left[\sum_{i=1}^{N} Z_i S_{ir} w_i g_{rp}(X_{ip}) - \frac{\sum_{i=1}^{N} h(X_i, S_i) S_{ir} g_{rp}(X_{ip})}{\sum_{i=1}^{N} h(X_i, S_i) S_{ir}}\right] && (22) \\
&\qquad \left. \sum_{p=1}^{P}\left[\frac{\sum_{i=1}^{N} h(X_i, S_i) S_{ir} g_{rp}(X_{ip})}{\sum_{i=1}^{N} h(X_i, S_i) S_{ir}} - \frac{\mathbb{E}[h(X,S) g_{rp}(X_p) | S_r = 1]}{\mathbb{E}[h(X,S)|S_r = 1]}\right]\right| && (23)
\end{aligned}
$$

where equation (19) = 0 by the definition of weights, equation (21) = 0 by design of weights, and the expectation of equation (23) = 0. It follows

$$
\begin{aligned}
&\left|\mathbb{E}\left(\hat{\tau}_{r,h} - \tau_{r,h}\right)\right| = |(20) + (22)| \\
&= \left|\sum_{p=1}^{P}\left[\sum_{i=1}^{N} Z_i S_{ir} w_i \left(\sum_{j=1}^{M_p/l_p}\sum_{k=0}^{K-1} \eta_{rjpk}\tilde{X}_{ijp}^{k} + R_{rijpK}\right) - \sum_{i=1}^{N}(1-Z_i)S_{ir}w_i\left(\sum_{j=1}^{M_p/l_p}\sum_{k=0}^{K-1} \eta_{rjpk}\tilde{X}_{ijp}^{k} + R_{rijpK}\right)\right] + \right. \\
&\qquad \left. \sum_{p=1}^{P}\left[\sum_{i=1}^{N} Z_i S_{ir} w_i \left(\sum_{j=1}^{M_p/l_p}\sum_{k=0}^{K-1} \phi_{rjpk}\tilde{X}_{ijp}^{k} + T_{rijpK}\right) - \frac{\sum_{i=1}^{N} h(X_i,S_i)S_{ir}(\sum_{j=1}^{M_p/l_p}\sum_{k=0}^{K-1}\phi_{rjpk}\tilde{X}_{ijp}^{k} + T_{rijpK})}{\sum_{i=1}^{N} h(X_i,S_i)S_{ir}}\right]\right| \\
&= \left|\sum_{p=1}^{P}\sum_{k=0}^{K-1}\sum_{j=1}^{M_p/l_p}\eta_{rjpk}\left[\sum_{i=1}^{N} Z_i S_{ir} w_i \tilde{X}_{ijp}^{k} - \sum_{i=1}^{N}(1-Z_i)S_{ir}w_i\tilde{X}_{ijp}^{k}\right] + \right. \\
&\qquad \sum_{p=1}^{P}\sum_{k=0}^{K-1}\sum_{j=1}^{M_p/l_p}\phi_{rjpk}\left[\sum_{i=1}^{N} Z_i S_{ir} w_i \tilde{X}_{ijp}^{k} - \frac{\sum_{i=1}^{N} h(X_i,S_i)S_{ir}\tilde{X}_{ijp}^{k}}{\sum_{i=1}^{N} h(X_i,S_i)S_{ir}}\right] + \\
&\qquad \sum_{p=1}^{P}\sum_{j=1}^{M_p/l_p}\left(\sum_{i=1}^{N} Z_i S_{ir} w_i R_{rijpK} - \sum_{i=1}^{N}(1-Z_i)S_{ir}w_i R_{rijpK}\right) + \\
&\qquad \left. \sum_{p=1}^{P}\sum_{j=1}^{M_p/l_p}\left(\sum_{i=1}^{N} Z_i S_{ir} w_i T_{rijpK} - \frac{\sum_{i=1}^{N} h(X_i,S_i)S_{ir}T_{rijpK}}{\sum_{i=1}^{N} h(X_i,S_i)S_{ir}}\right)\right| \\
&\le \sum_{p=1}^{P}\sum_{k=0}^{K-1}\sum_{j=1}^{M_p/l_p}\delta_{1,pk}|\eta_{rjpk}| + \sum_{p=1}^{P}\sum_{k=0}^{K-1}\sum_{j=1}^{M_p/l_p}\delta_{2,pk}|\phi_{rjpk}| + \\
&\qquad \sum_{p=1}^{P}\sum_{j=1}^{M_p/l_p}\left(\sum_{i=1}^{N}|R_{rijpK}|\left|Z_i S_{ir} w_i - (1-Z_i)S_{ir}w_i\right|\right) + \sum_{p=1}^{P}\sum_{j=1}^{M_p/l_p}\left(\sum_{i=1}^{N}\left|T_{rijpK}\right|\left|Z_i S_{ir} w_i - \frac{h(X_i,S_i)S_{ir}}{\sum_{i=1}^{N} h(X_i,S_i)S_{ir}}\right|\right) \\
&\le \sum_{p=1}^{P}\sum_{k=0}^{K-1}\sum_{j=1}^{M_p/l_p}\delta_{1,pk}|\eta_{rjpk}| + \sum_{p=1}^{P}\sum_{k=0}^{K-1}\sum_{j=1}^{M_p/l_p}\delta_{2,pk}|\phi_{rjpk}| + 2P\frac{M_p}{l_p}\frac{C_1}{K!}(l_p/2)^K + 2P\frac{M_p}{l_p}\frac{C_2}{K!}(l_p/2)^K.
\end{aligned}
$$

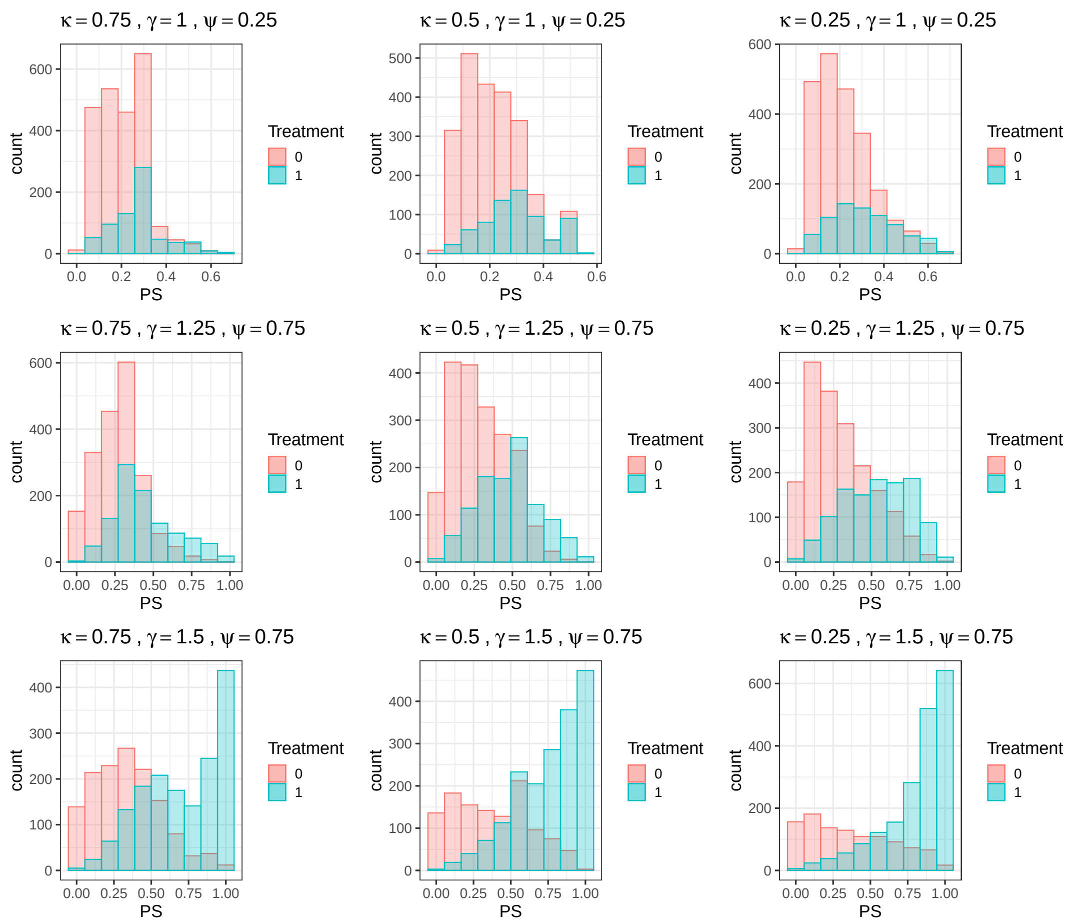
□

2 | ADDITIONAL INFORMATION ON THE SIMULATIONS

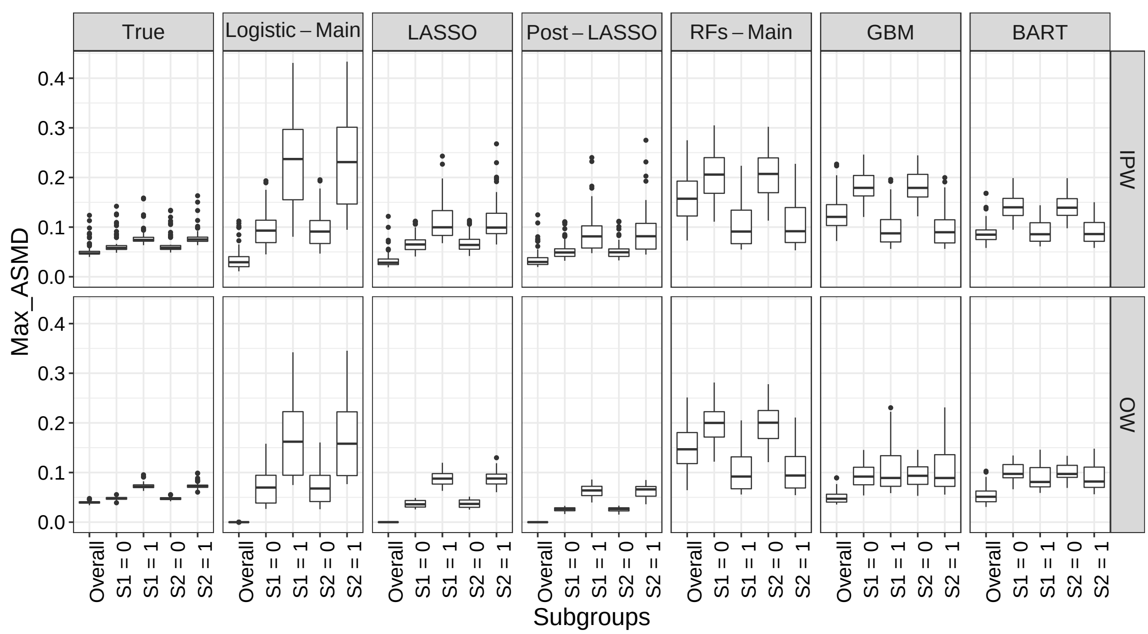
2.1 | Propensity Score Distribution Diagnostics

We first investigate the distributions of true PS by each treatment arm in the simulated data. The data are evaluated in terms of the extreme PS values and degree of overlap between two treatment arms. Figure 1 summarizes the true propensity score distribution across values of κ (across columns) and ψ, γ (down rows). The resulting propensity score distributions vary from showing low confounding with substantial overlap, when $P = 20, \gamma = 1$, $\kappa = 0.75$ and $\psi = 0.25$ (upper left corner), to strong confounding with more probability in the tails, when $P = 50, \gamma$=1.5, $\kappa = 0.25$ and $\psi = 0.75$ (lower right corner).

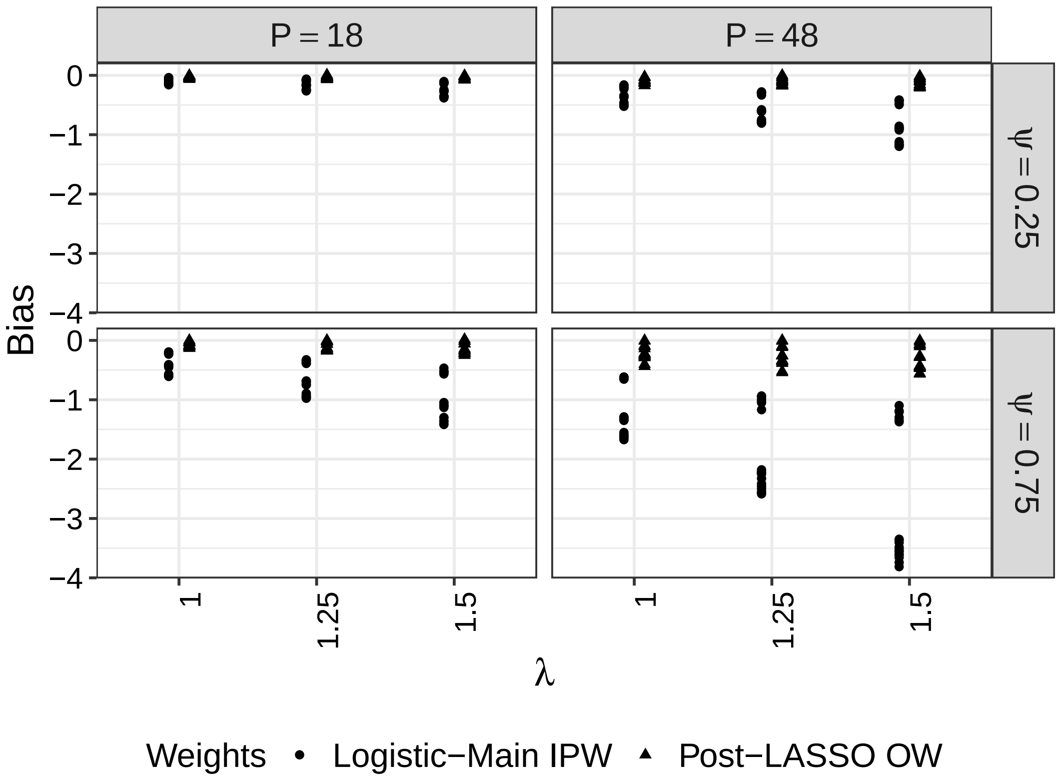
2.2 | Additional Simulation Results



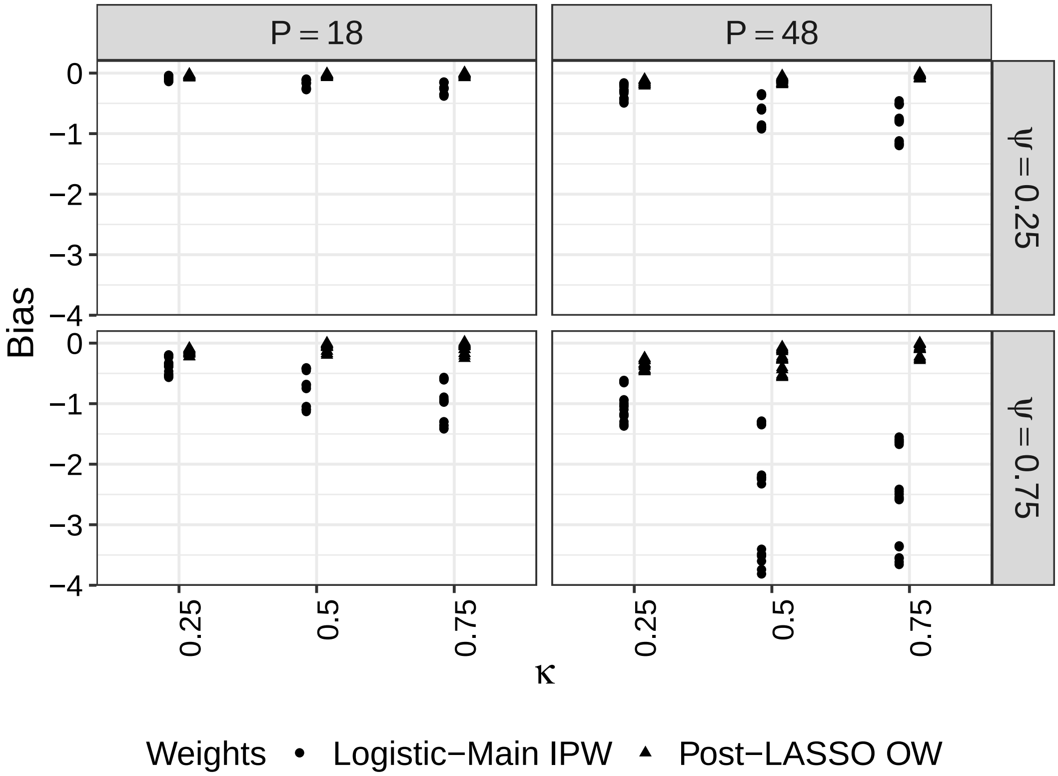
Web Figure 1 Propensity score distributions for variations of the data generation, $\gamma \in \{1, 1.25, 1.5\}$, $\kappa \in \{0.25, 0.5, 0.75\}$, and $\psi \in \{0.25, 0.75\}$.



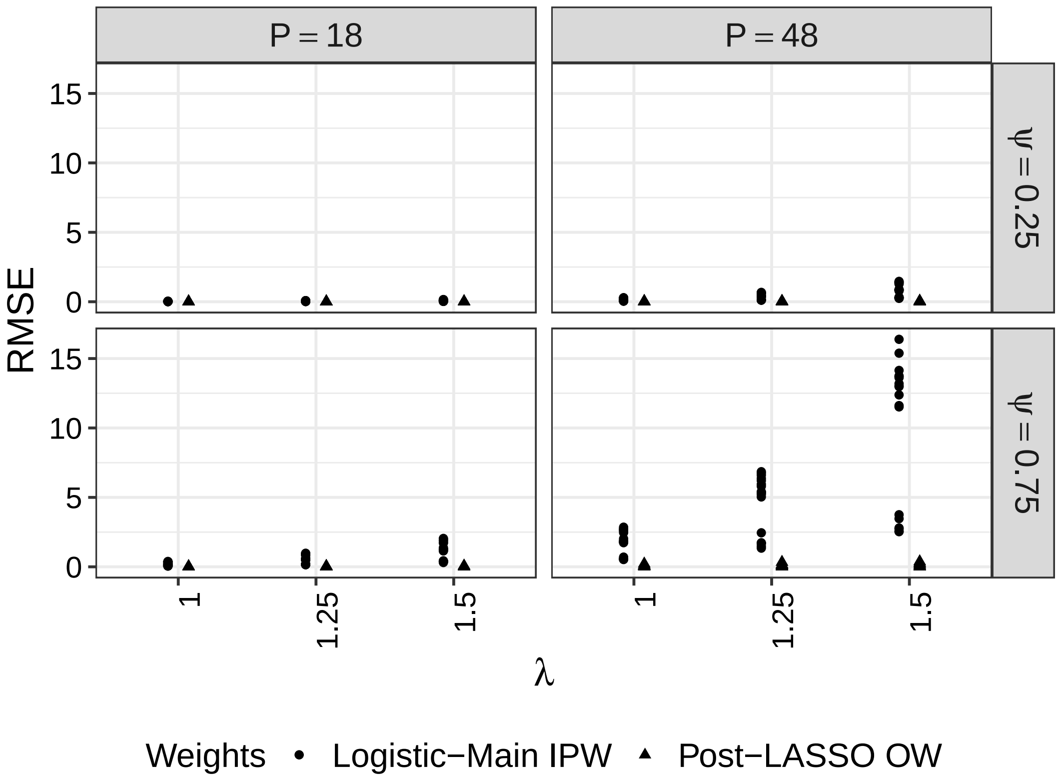
Web Figure 2 Max ASMD over all covariates, calculated in overall sample and four subgroups, across different propensity models and weighting schemes. Each dot represents one of the 72 simulation scenarios.



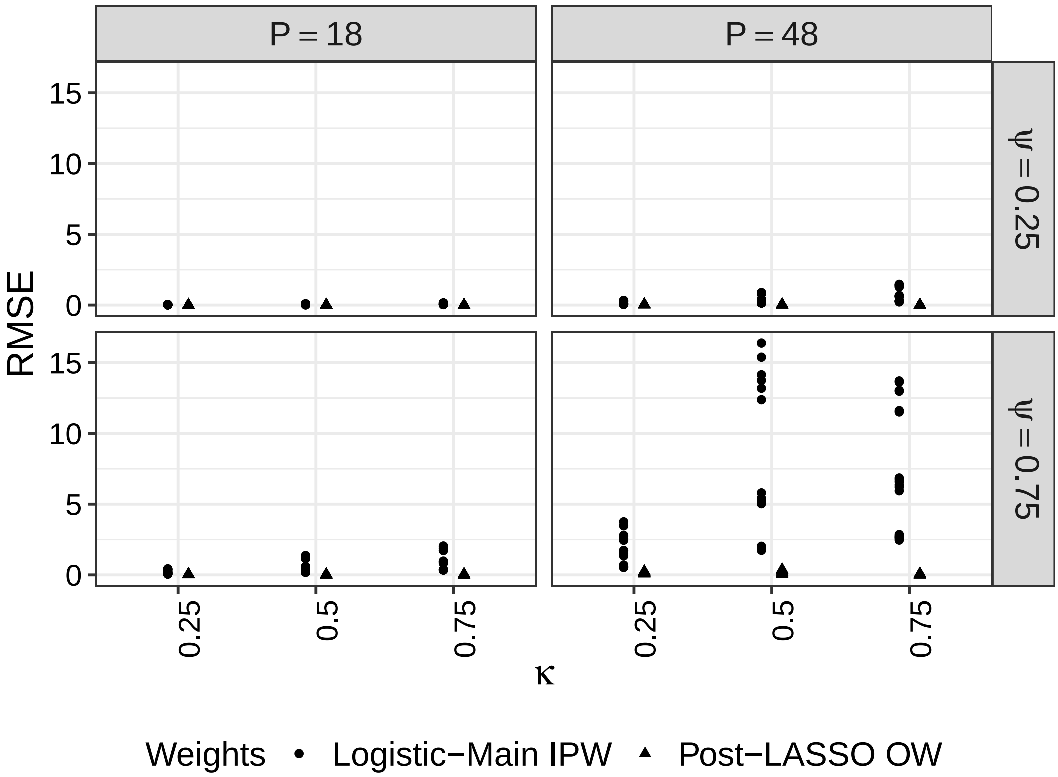
Web Figure 3 Bias of IPW main effect logistic model versus OW Post-LASSO across different λ values in estimating S-WATE $\tau_{\{S_1=1\}}$, displayed by $\psi \in \{0.25, 0.75\}$ and $P \in \{18, 48\}$. Each dot represents one of the 72 simulation scenarios.



Web Figure 4 Bias of IPW main effect logistic model versus OW Post-LASSO across different κ values in estimating S-WATE $\tau_{\{S_1=1\}}$, displayed by $\psi \in \{0.25, 0.75\}$ and $P \in \{18, 48\}$. Each dot represents one of the 72 simulation scenarios.



Web Figure 5 RMSE of IPW main effect logistic model versus OW Post-LASSO across different λ values in estimating S-WATE $\tau_{\{S_1=1\}}$, displayed by $\psi \in \{0.25, 0.75\}$ and $P \in \{18, 48\}$. Each dot represents one of the 72 simulation scenarios.



Web Figure 6 RMSE of IPW main effect logistic model versus OW Post-LASSO across different κ values in estimating S-WATE $\tau_{\{S_1=1\}}$, displayed by $\psi \in \{0.25, 0.75\}$ and $P \in \{18, 48\}$. Each dot represents one of the 72 simulation scenarios.

2.3 Additional Results on COMPARE-UP Analysis

Web Figure 7 The Baseline Characteristics Table by Procedure Type.

Characteristic	Unweighted			Weighted (Post-LASSO overlap weights)		
	Myomectomy (N = 567)	Hysterectomy (N = 863)	Overall (N = 1430)	Myomectomy (N = 567)	Hysterectomy (N = 863)	Overall (N = 1430)
Age	38.0	44.7	42.0	41.6	41.6	41.6
Race						
White	38.6	50.2	45.6	45.5	45.5	45.5
Black	40.7	37.9	39.0	40.1	40.1	40.1
Other	20.6	11.9	15.4	14.3	14.3	14.3
Hispanic Ethnicity	8.3	7.4	7.8	5.0	5.0	5.0
Private Insurance	84.8	81.2	82.7	83.3	83.3	83.3
Parity: 2+	9.9	52.4	35.5	25.0	25.0	25.0
Prior procedure	11.1	16.3	14.3	15.0	15.0	15.0
Bleeding Symptoms	76.9	85.3	82.0	79.9	79.9	79.9
Pelvic Pain	58.2	65.7	62.7	57.9	57.9	57.9
Anxiety/Depression	28.4	27.3	27.8	27.3	27.3	27.3
UFS-QoL Total Score	50.4	43.7	46.3	48.3	48.3	48.3
UFS-QoL Concern	49.9	37.5	42.4	44.1	44.1	44.1
UFS-QoL Activities	52.4	44.9	47.8	49.2	49.2	49.2
UFS-QoL Self-concious	45.4	42.2	43.4	44.9	44.9	44.9
UFS-QoL Control	49.9	48.4	49.0	51.0	51.0	51.0
UFS-QoL Energy	50.9	44.8	47.3	49.6	49.6	49.6
UFS-QoL Sexual function	53.6	44.5	48.1	51.6	51.6	51.6
UFS-QoL Symptom severity	50.4	60.5	56.5	53.6	53.6	53.6
EQ5D: VAS	73.3	69.4	71.0	71.3	71.3	71.3
Time since diagnosis, years	5.2	7.0	6.2	5.6	5.4	5.5
Uterine volume	744	603	659	665	694	680

Values in cells are means and percentages.

References

1. Li F, Morgan KL, Zaslavsky AM. Balancing Covariates via Propensity Score Weighting. *Journal of the American Statistical Association* 2018; 113(521): 390-400. doi: 10.1080/01621459.2016.1260466

2. Austin PC, Stuart EA. Moving towards best practice when using inverse probability of treatment weighting (IPTW) using the propensity score to estimate causal treatment effects in observational studies. *Statistics in medicine* 2015; 34(28): 3661–3679.

3. Zubizarreta JR. Stable weights that balance covariates for estimation with incomplete outcome data. *Journal of the American Statistical Association* 2015; 110(511): 910–922.