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

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

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

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

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 $\widehat{\omega}_i^*$), we have

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

□

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 $\widehat{w}_i^*$),

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

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

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

□

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

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

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, unweighted standard deviation of X_p. 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 pooled, unweighted sample standard deviation for the p^{th} covariate in the r^{th} subgroup (i.e. $s_{r,p} = \sqrt{\frac{s^2_{r,p,1}+s^2_{r,p,0}}{2}}$). We define the subgroup mean within arm $z = 0, 1$ as $\bar{X}_{p,0}$ and $\bar{X}_{p,1}$, subgroup sample size within arm $z = 0, 1$ as N_{r0} and N_{r1}. Finally, the subgroup sample standard deviation within arm $z = 0, 1$ is

$$s^2_{r,p,1} = N^{-1}_{r1} \sum_i Z_i S_{ir}(X_{ip} - \bar{X}_{p,1})^2,$$

and

$$s^2_{r,p,0} = N^{-1}_{r0} \sum_i (1 - Z_i) S_{ir}(X_{ip} - \bar{X}_{p,0})^2.$$

Here, we use the unweighted sample variance to set a common denominator and therefore facilitate comparisons across different weighting methods.

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 $f^{(K)}_{rp} \leq C_1, g^{(K)}_{rp} \leq C_2$, for all r, p, and $E(\epsilon_i | X_{,i}) = 0$. 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 $k = 1, \ldots, K$, $j = 1, 2, \ldots, M_p/l_p$ and $p = 1, \ldots, P$. 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, $R_{rijpK} = \frac{f^{(K)}_{rp}(\xi'_{jp})}{K!} \tilde{X}^K_{ijp} \leq |\frac{f^{(K)}_{rp}(\xi'_{jp})}{K!} (l_p/2)^K| \leq \frac{C_1}{K!} (l_p/2)^K$, and $T_{rijpK} = \frac{g^{(K)}_{rp}(\xi'_{jp})}{K!} \tilde{X}^K_{ijp} \leq |\frac{g^{(K)}_{rp}(\xi'_{jp})}{K!} (l_p/2)^K| \leq \frac{C_2}{K!} (l_p/2)^K$ for all $r = 1, \ldots, R; j = 1, \ldots, Mp/lp; p = 1, \ldots, P$.

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

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

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(\widehat{\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}
$$

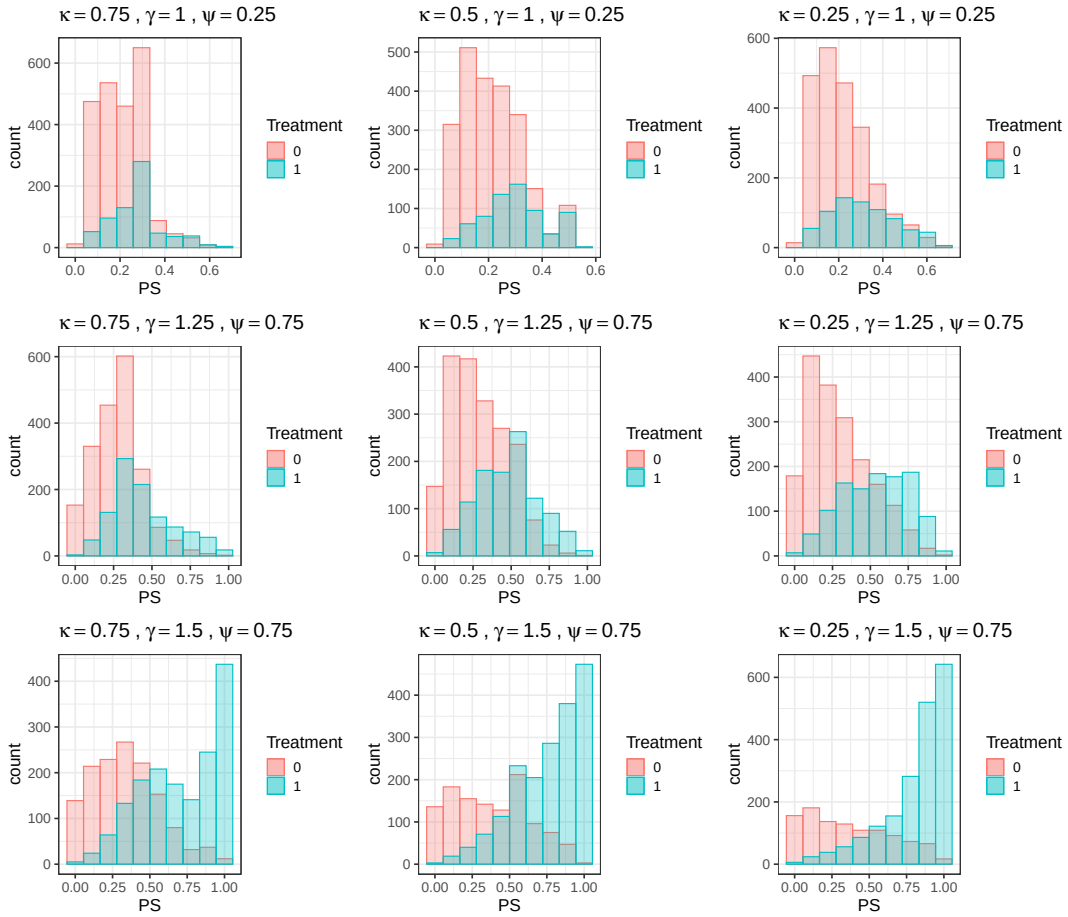
□

2 | ADDITIONAL INFORMATION ON THE SIMULATIONS

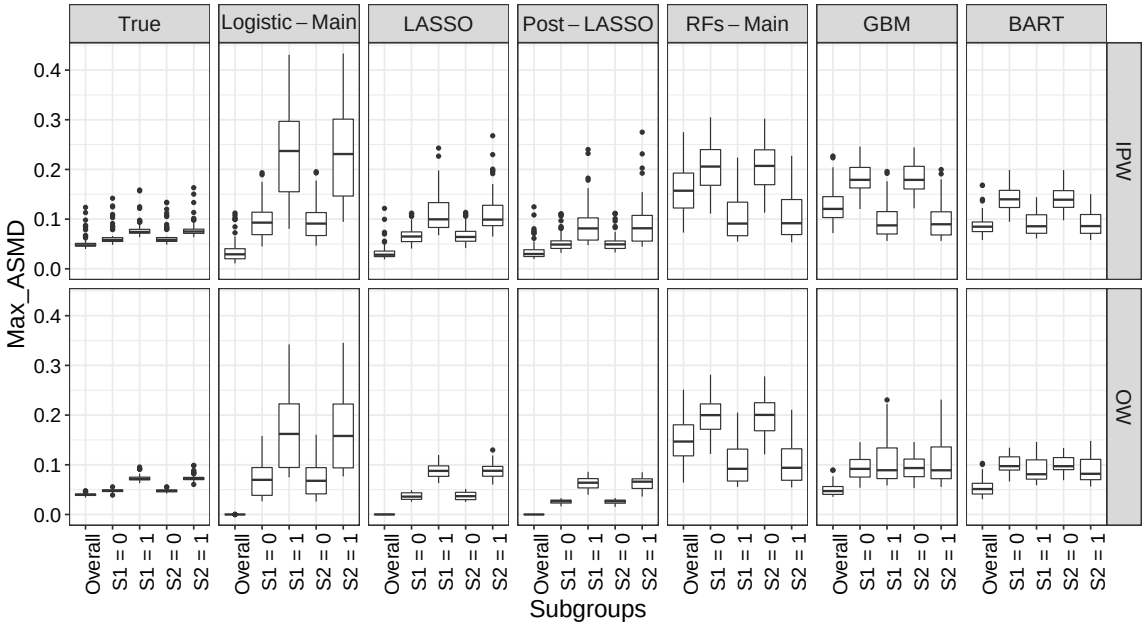
2.1 | Propensity Score Distribution Diagnostics

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 panels), to strong confounding with more probability in the tails, when $P = 50, \gamma$=1.5, $\kappa = 0.25$ and $\psi = 0.75$ (lower right panels).

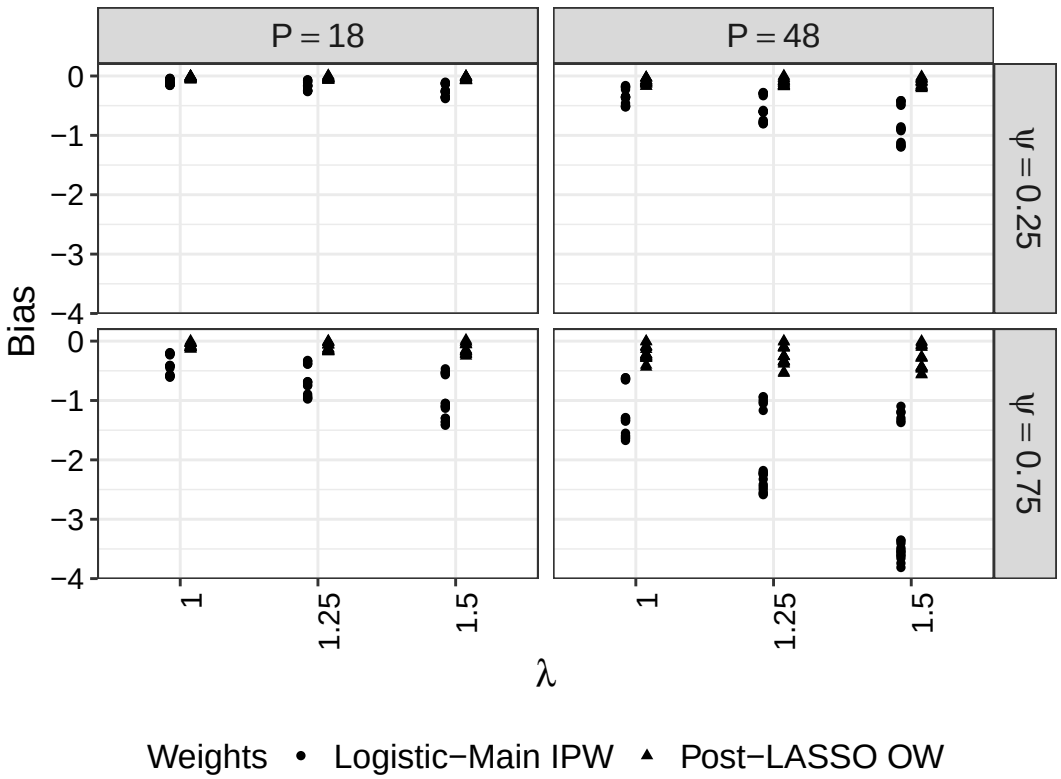
2.2 | Additional Simulation Results: ASMD, bias and RMSE



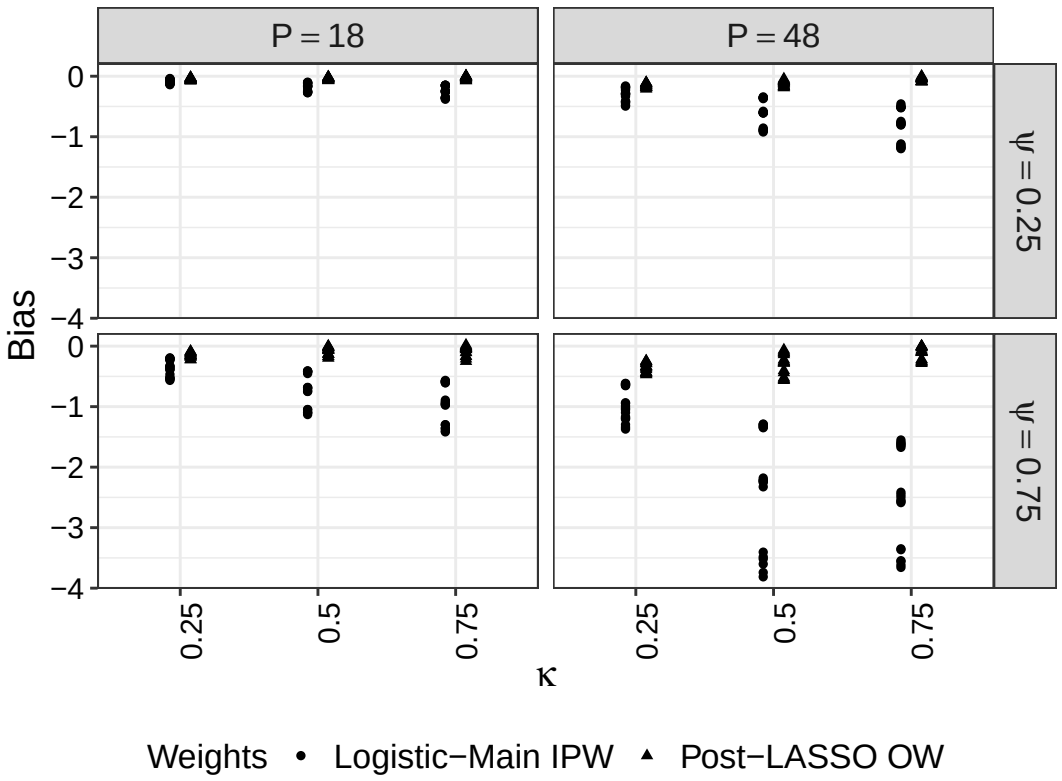
Web Figure 1 Propensity score distributions for variations of the data generation, $P \in \{18$ (the first two rows), 48 (the last row) $\}$, $\gamma \in \{1, 1.25, 1.5\}$, $\kappa \in \{0.25, 0.5, 0.75\}$, and $\psi \in \{0.25, 0.75\}$.



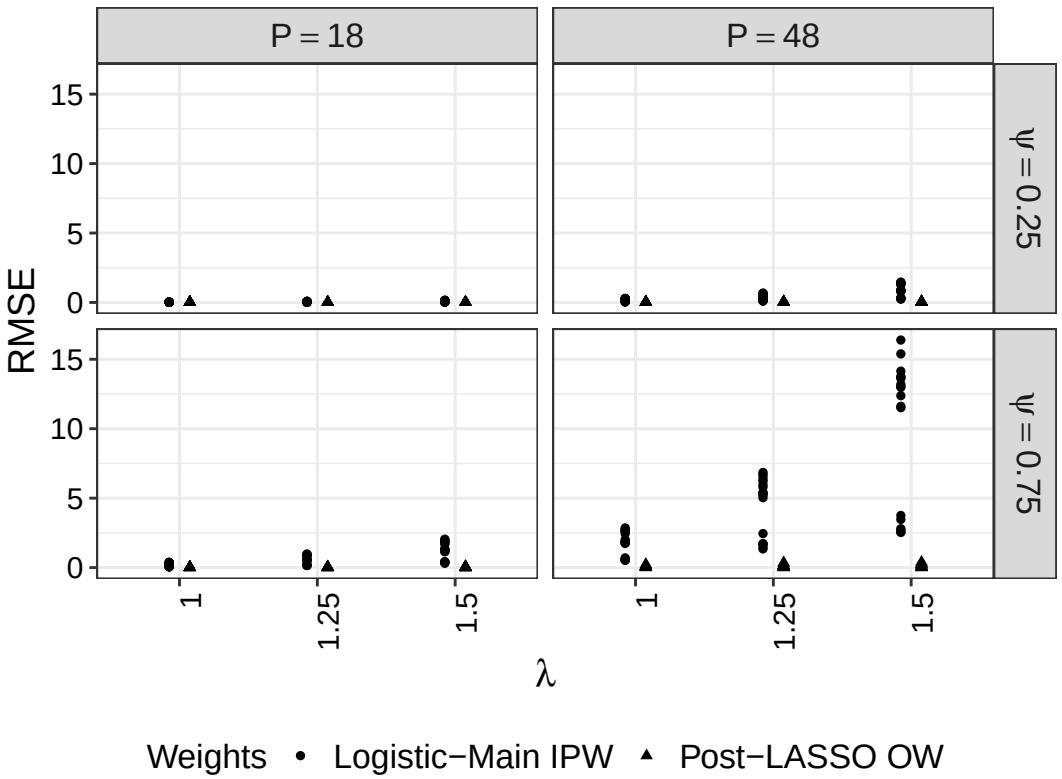
Web Figure 2 Max ASMD over all covariates, calculated in overall sample and four subgroups, across different propensity models and weighting schemes. Each dot represents one of the 72 simulation scenarios.



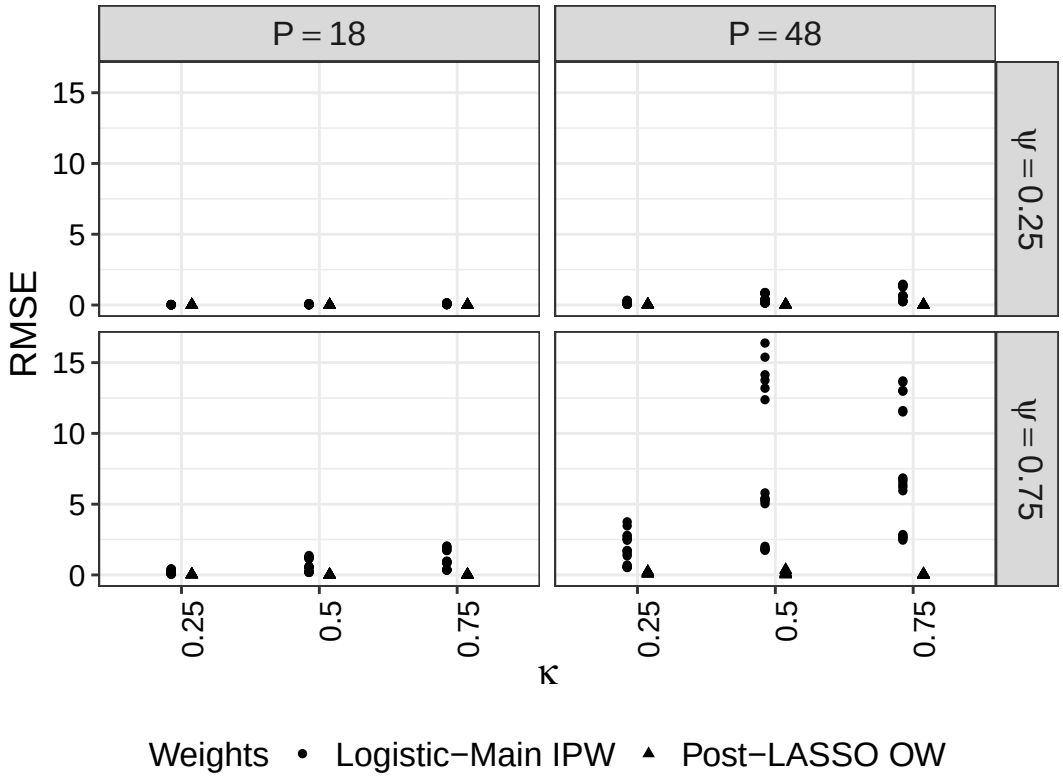
Web Figure 3 Bias of IPW main effect logistic model versus OW Post-LASSO across different λ values in estimating S-WATE $\tau_{\{S_1=1\}}$, displayed by $\psi \in \{0.25, 0.75\}$ and $P \in \{18, 48\}$. Each dot represents one of the 72 simulation scenarios.



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 Simulation Study on Variance and Coverage Rate

We have conducted an additional small simulation to examine how variable selection in propensity scores affects variance estimation and the coverage rate of different methods discussed in the paper. First, we compared three methods for estimating the variance: (1) Re-estimate the propensity score (PS) using post-LASSO on each bootstrap sample, and estimate the causal effects with this PS; (2) Estimate PS only on the original sample using post-LASSO and treat this estimated PS as fixed when estimating the causal effects in each bootstrap sample; (3) Use the robust sandwich variance. Method (2) is justified based on a large amount of literature which suggests that propensity score uncertainty does not have have to be accounted for when using propensity score approaches.[4,5] Method (3) is justified based on if the propensity score is estimated via the standard maximum-likelihood estimation of a generalized linear model, then both bootstrapping (i.e. refit the propensity score model in each bootstrap sample) or M-estimation (i.e. sandwich estimator obtained via the stacked estimating equations[6]) would accurately quantify the uncertainty due to estimating the propensity scores. Method (1) is considered here in order to investigate the performance of bootstraping the LASSO propensity score, though we caution against the use of Method (1) in practice due to lack of theoretical justification.

We compared these three variance methods with the Monte Carlo standard deviation (MC_SD) in two simulation scenarios to estimate the overall treatment effect. Scenario 1 corresponds to a simple case in our simulation study described in the Section 5 ($N = 1000, P = 18, \phi = 0.75, \gamma = 1, \kappa = 0.5, \beta_{sz}^T = (0.5, 0.5)$). Scenario 2 corresponds to a situation where the sample size is small, and the number of covariates is large with small coefficients ($N = 200, P = 100, \phi = 1, \gamma = 0.1, \kappa = 0.5, \beta_{sz}^T = (0, 0)$). Since the number of truly non-zero coefficients in the propensity score model is substantially larger than the sample size ($N = 200$ but 300 terms with small coefficients are included in the true propensity score when incorporating covariate-subgroup interactions), we expect re-fitting the LASSO to perform poorly in Scenario 2 due to instability in the covariates that will be chosen across bootstrap samples.

Web Table 1 shows the results of this simulation study. The robust sandwich estimator is very close to the bootstrap with fixed PS in both scenarios. In Scenario 1, under large sample size, the Lasso estimate is similar to the ordinary least squares estimate, and therefore the variance from bootstrap with re-estimated PS is close to other methods. In Scenario 2, the re-estimated PS has larger variance probably due to selected covariates changing often from bootstrap to bootstrap sample. Except for IPW-pLASSO in Scenario 2, the variance estimate from bootstrap with fixed PS and sandwich estimator is close to the Monte Carlo SD.

Web Table 1 Variance comparison by different estimation methods

MC_SD	Sandwich	Re-est PS	Fixed PS	PS model	Weight
Scenario 1					
0.09	0.11	0.08	0.11	LASSO	OW
0.11	0.11	0.08	0.11	pLASSO	OW
0.11	0.13	0.11	0.13	LASSO	IPW
0.15	0.15	0.15	0.15	pLASSO	IPW
Scenario 2					
0.15	0.15	0.16	0.15	LASSO	OW
0.16	0.16	0.24	0.16	pLASSO	OW
0.15	0.15	0.28	0.15	LASSO	IPW
0.31	0.19	0.35	0.21	pLASSO	IPW

We also used all three methods in our COMPARE-UF application. Results for all three methods were practically identically. The confidence intervals in Figure 4 of the manuscript are based on the robust sandwich variance estimator.

In addition, we examined the coverage rate of Logistic-Main, LASSO, pLASSO and random forests (RFs) under three scenarios, corresponding to low confounding with substantial overlap ($P = 18, \gamma = 1, \kappa = 0.5$ and $\psi = 0.25$; [1,2] in Web Figure 1), moderate confounding and overlap ($P = 18, \gamma = 1.25, \kappa = 0.5$ and $\psi = 0.75$; [2,2] in Web Figure 1), and strong confounding with more probability in the tails ($P = 48, \gamma$=1.5, $\kappa = 0.5$ and $\psi = 0.75$; [3,2] in Web Figure 1). The confidence interval is estimated by 1000 bootstrap samples with fixed PS. Web Table 2 shows that the coverage rates of pLASSO are close to the nominal 95% levels under low and moderate confounding, and slightly lower than 95% under strong confounding. LASSO fails to maintain the nominal coverage rate under moderate or strong confounding, probably due to the shrinkage bias. The coverage

rates of Logistic-Main are slightly lower than 95% under low confounding, but are substantially lower than 95% under moderate and strong confounding. RFs fail to maintain the nominal coverage rate in all scenarios. This further validates pLASSO as our preferred method for building the propensity score model.

Web Table 2 Coverage rate of different propensity score methods in subgroups. The results are based on 1000 simulations.

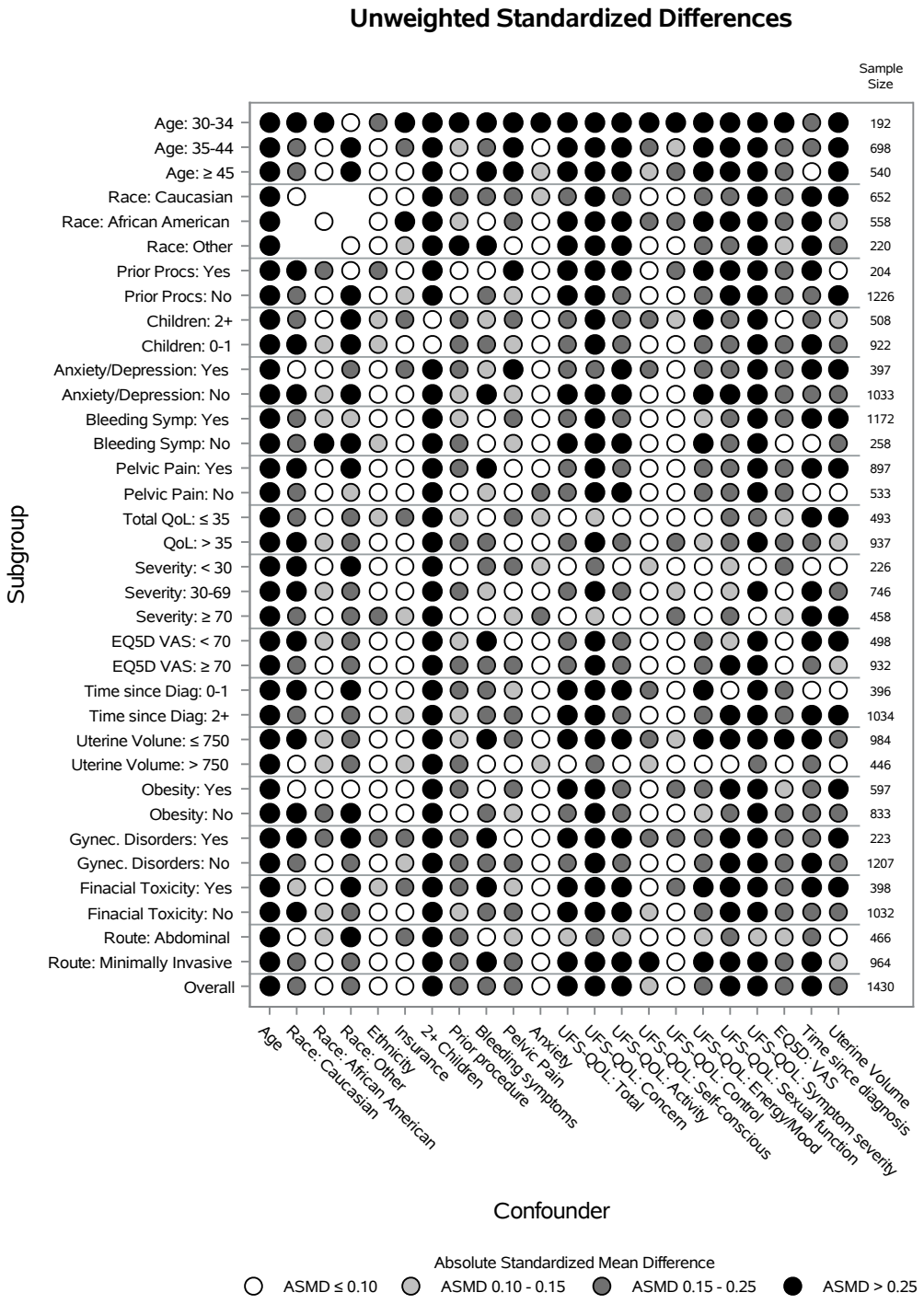
	OW				IPW			
	S1=0	S1=1	S2=0	S2=1	S1=0	S1=1	S2=0	S2=1
Low confounding								
Logistic-Main	0.91	0.83	0.93	0.84	0.79	0.85	0.81	0.86
LASSO	0.95	0.91	0.96	0.91	0.88	0.93	0.90	0.93
pLASSO	0.95	0.95	0.98	0.95	0.93	0.96	0.94	0.95
RFs	0.44	0.86	0.42	0.87	0.58	0.87	0.58	0.89
Moderate confounding								
Logistic-Main	0.29	0.02	0.29	0.03	0.15	0.02	0.15	0.02
LASSO	0.88	0.60	0.87	0.61	0.56	0.72	0.58	0.73
pLASSO	0.98	0.98	0.98	0.97	0.93	0.98	0.92	0.98
RFs	0.00	0.83	0.00	0.82	0.00	0.84	0.00	0.83
Strong confounding								
Logistic-Main	0.00	0.00	0.00	0.00	0.22	0.00	0.17	0.00
LASSO	0.06	0.00	0.06	0.00	0.18	0.20	0.16	0.19
pLASSO	0.89	0.83	0.88	0.82	0.70	0.90	0.70	0.89
RFs	0.00	0.01	0.00	0.01	0.00	0.01	0.00	0.01

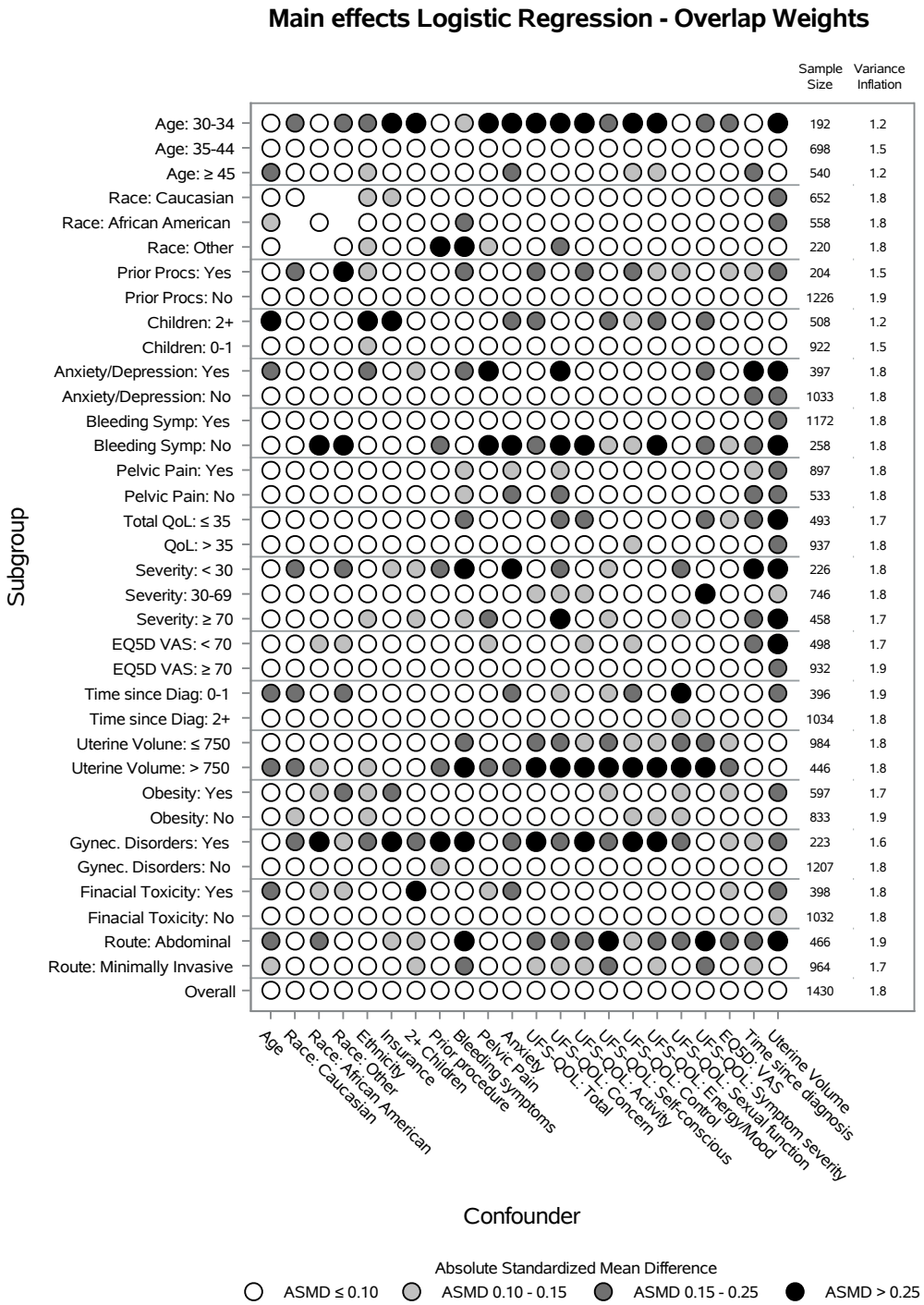
2.4 | Additional Results on COMPARE-UP Analysis

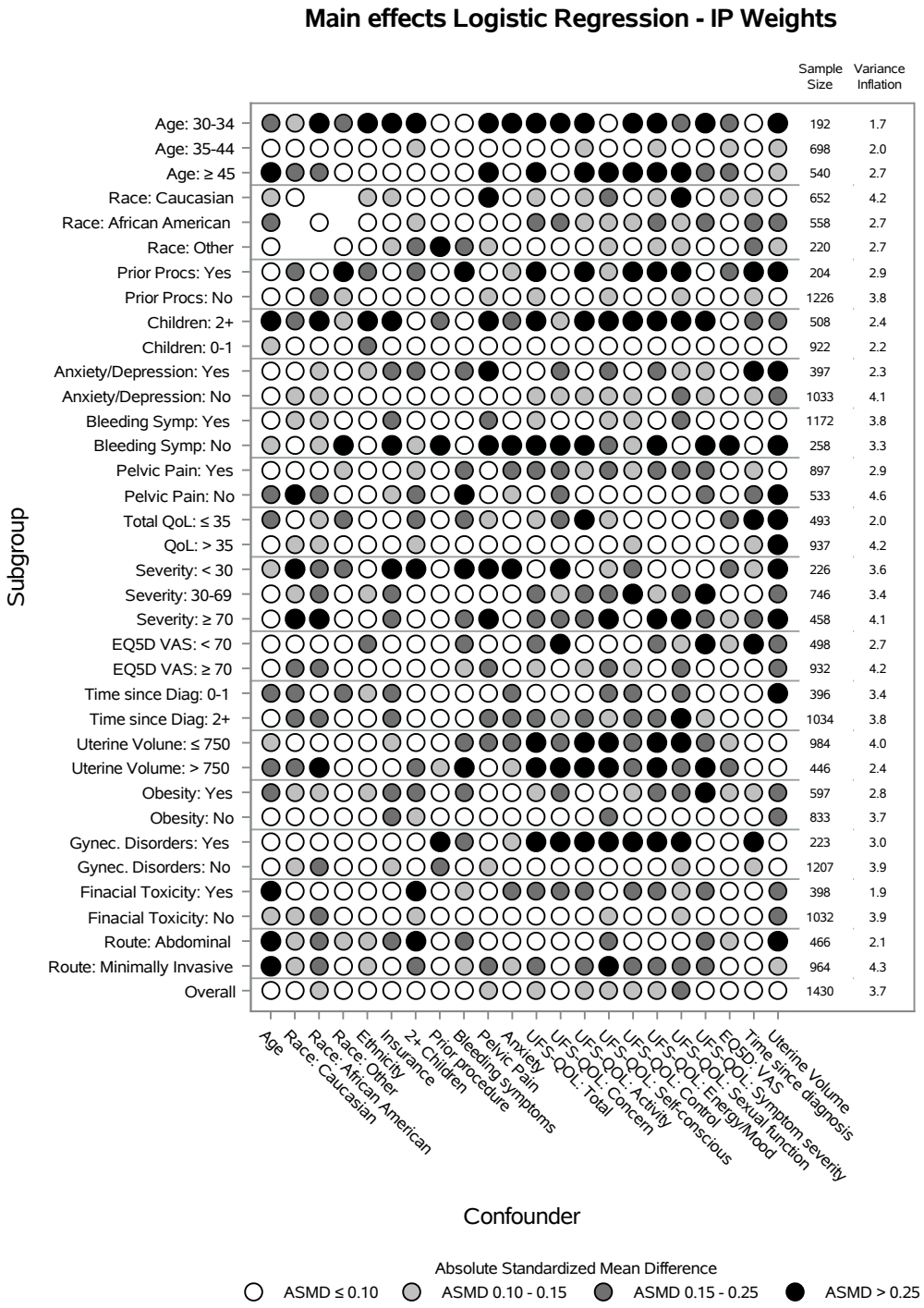
Characteristic	Unweighted			Weighted (Post-LASSO overlap weights)		
	Myomectomy (N = 567)	Hysterectomy (N = 863)	Overall (N = 1430)	Myomectomy (N = 567)	Hysterectomy (N = 863)	Overall (N = 1430)
Age	38.0	44.7	42.0	41.6	41.6	41.6
Race						
White	38.6	50.2	45.6	45.5	45.5	45.5
Black	40.7	37.9	39.0	40.1	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

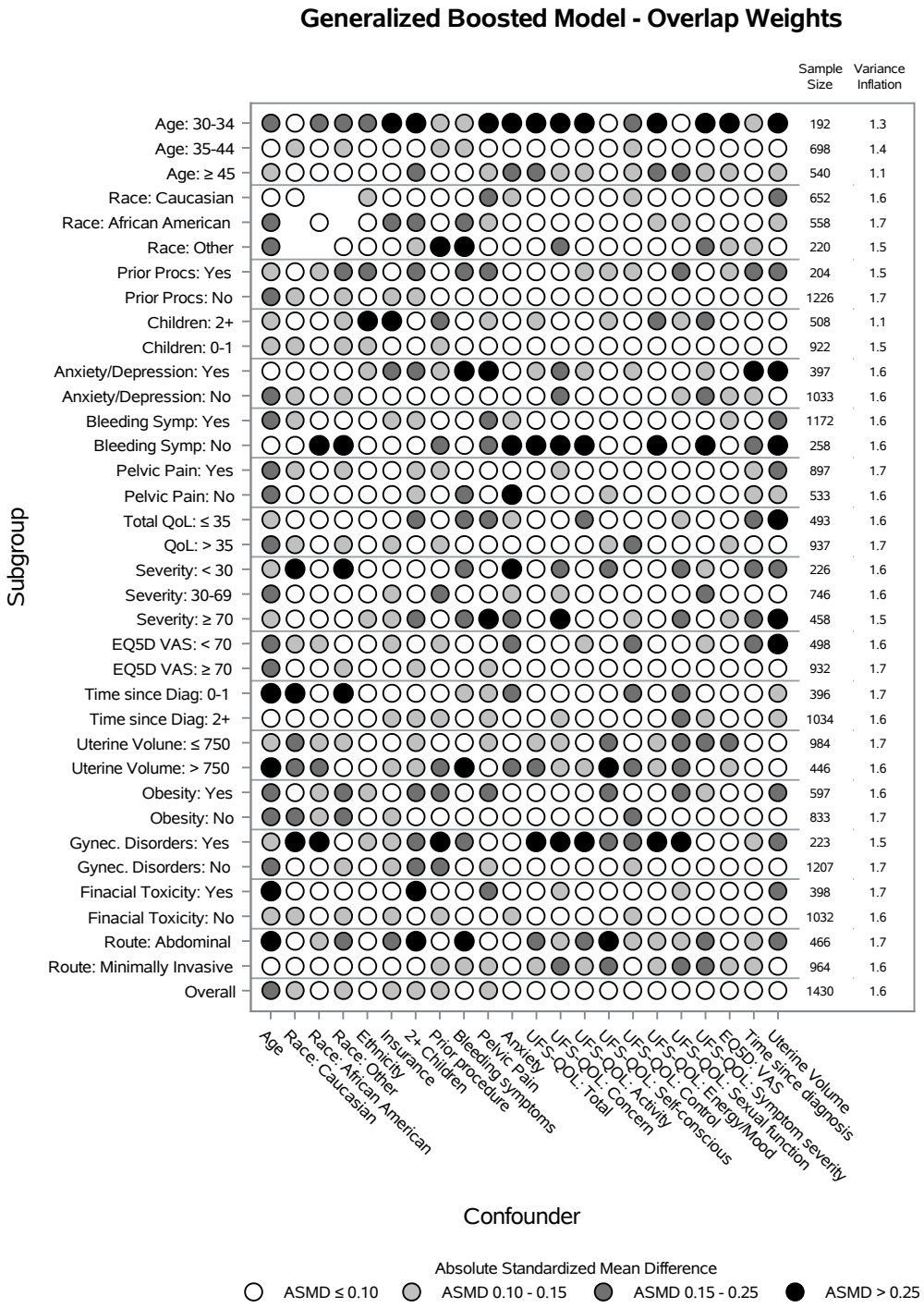
Values in cells are means and percentages.

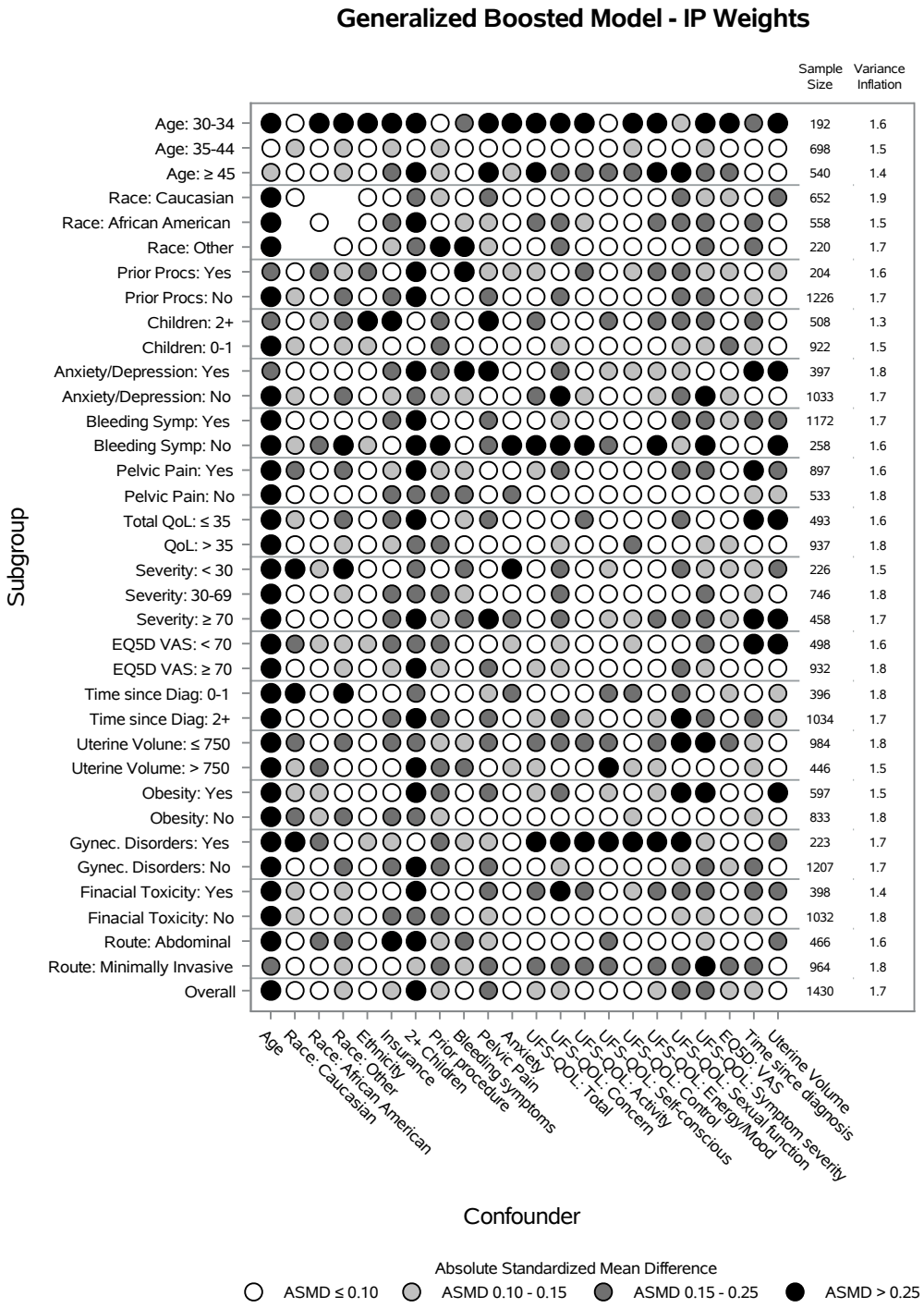
Web Figure 7 The Weighted Baseline Characteristics Table by Procedure Type.

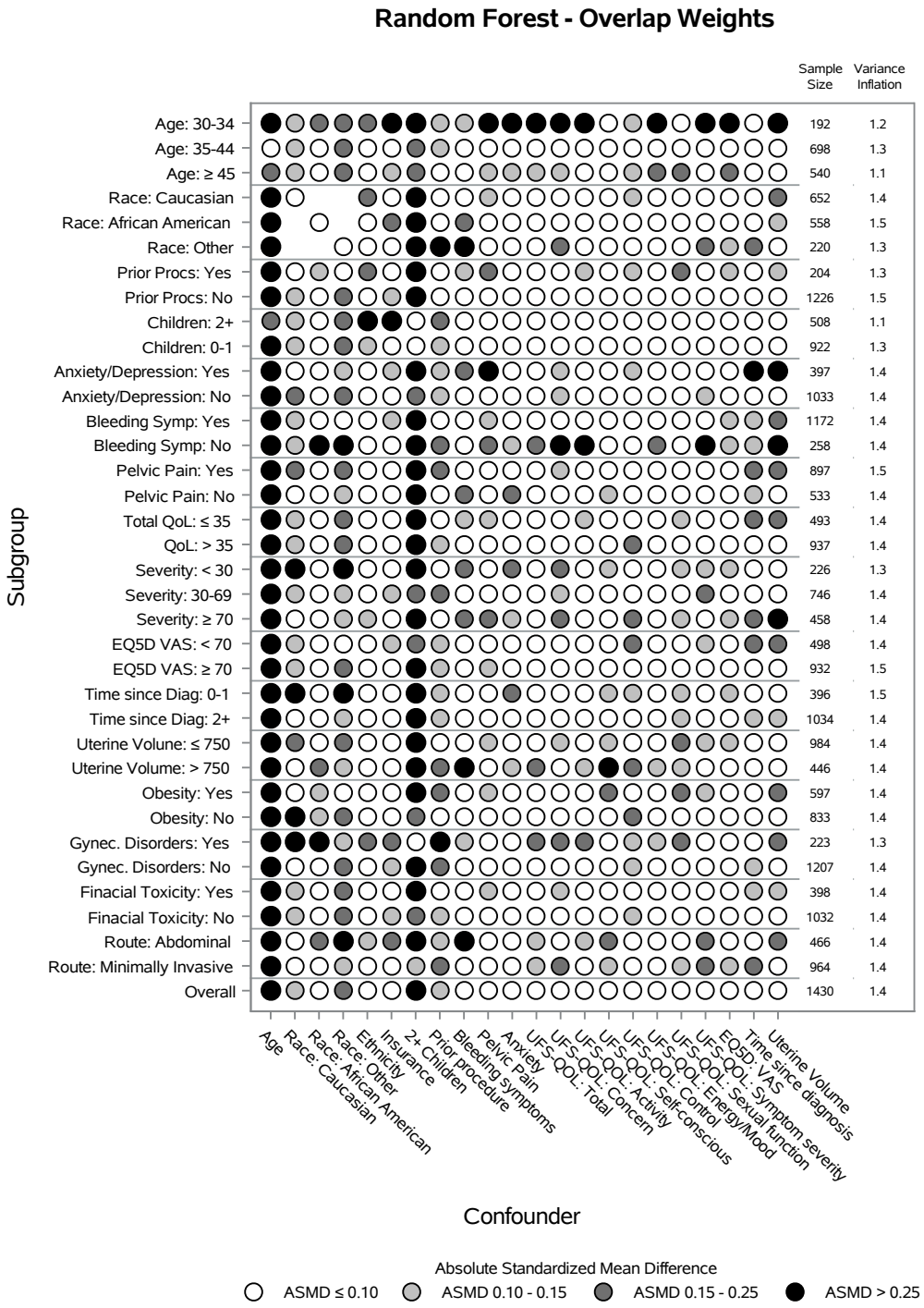

Unweighted Standardized Differences
Sample Size
Age: 30-34 192
Age: 35-44 698
Age: ≥ 45 540
Race: Caucasian 652
Race: African American 558
Race: Other 220
Prior Procs: Yes 204
Prior Procs: No 1226
Children: 2+ 508
Children: 0-1 922
Anxiety/Depression: Yes 397
Anxiety/Depression: No 1033
Bleeding Symp: Yes 1172
Bleeding Symp: No 258
Pelvic Pain: Yes 897
Pelvic Pain: No 533
Total QoL: ≤ 35 493
QoL: > 35 937
Severity: < 30 226
Severity: 30-69 746
Severity: ≥ 70 458
EQ5D VAS: < 70 498
EQ5D VAS: ≥ 70 932
Time since Diag: 0-1 396
Time since Diag: 2+ 1034
Uterine Volume: ≤ 750 984
Uterine Volume: > 750 446
Obesity: Yes 597
Obesity: No 833
Gynec. Disorders: Yes 223
Gynec. Disorders: No 1207
Finacial Toxicity: Yes 398
Finacial Toxicity: No 1032
Route: Abdominal 466
Route: Minimally Invasive 964
Overall 1430
Subgroup
Age
Race: Caucasian
Race: African American
Race: Other
Ethnicity
Insurance
2+ Children
Prior procedure
Bleeding symptoms
Pelvic Pain
Anxiety
UFS-QOL: Total
UFS-QOL: Concern
UFS-QOL: Activity
UFS-QOL: Self-conscious
UFS-QOL: Control
UFS-QOL: Energy/Mood
UFS-QOL: Sexual function
UFS-QOL: Symptom severity
EQ5D: VAS
Time since diagnosis
Uterine Volume
Confounder
Absolute Standardized Mean Difference
ASMD ≤ 0.10
ASMD 0.10 - 0.15
ASMD 0.15 - 0.25
ASMD > 0.25

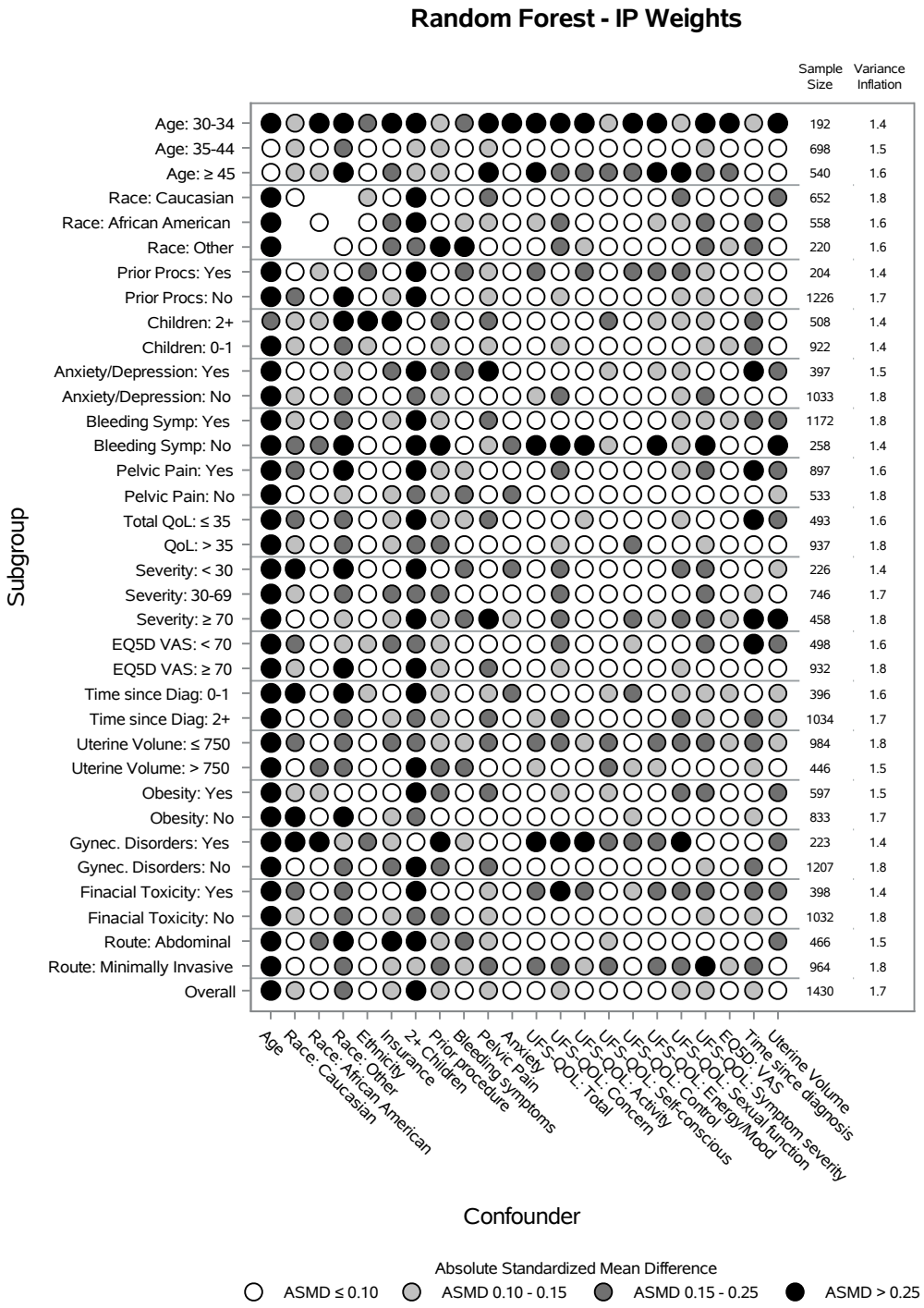

Main effects Logistic Regression - Overlap Weights
Sample Size
Variance Inflation
Age: 30-34 192 1.2
Age: 35-44 698 1.5
Age: ≥ 45 540 1.2
Race: Caucasian 652 1.8
Race: African American 558 1.8
Race: Other 220 1.8
Prior Procs: Yes 204 1.5
Prior Procs: No 1226 1.9
Children: 2+ 508 1.2
Children: 0-1 922 1.5
Anxiety/Depression: Yes 397 1.8
Anxiety/Depression: No 1033 1.8
Bleeding Symp: Yes 1172 1.8
Bleeding Symp: No 258 1.8
Pelvic Pain: Yes 897 1.8
Pelvic Pain: No 533 1.8
Total QoL: ≤ 35 493 1.7
QoL: > 35 937 1.8
Severity: < 30 226 1.8
Severity: 30-69 746 1.8
Severity: ≥ 70 458 1.7
EQ5D VAS: < 70 498 1.7
EQ5D VAS: ≥ 70 932 1.9
Time since Diag: 0-1 396 1.9
Time since Diag: 2+ 1034 1.8
Uterine Volume: ≤ 750 984 1.8
Uterine Volume: > 750 446 1.8
Obesity: Yes 597 1.7
Obesity: No 833 1.9
Gynec. Disorders: Yes 223 1.6
Gynec. Disorders: No 1207 1.8
Finacial Toxicity: Yes 398 1.8
Finacial Toxicity: No 1032 1.8
Route: Abdominal 466 1.9
Route: Minimally Invasive 964 1.7
Overall 1430 1.8
Subgroup
Age
Race: Caucasian
Race: African American
Race: Other
Ethnicity
Insurance
2+ Children
Prior procedure
Bleeding Symptoms
Pelvic Pain
Anxiety
UFS-QOL: Total
UFS-QOL: Concern
UFS-QOL: Activity
UFS-QOL: Self-conscious
UFS-QOL: Control
UFS-QOL: Energy/Mood
UFS-QOL: Sexual function
EQ5D: VAS
Time since diagnosis
Uterine Volume
Confounder
Absolute Standardized Mean Difference
ASMD ≤ 0.10
ASMD 0.10 - 0.15
ASMD 0.15 - 0.25
ASMD > 0.25

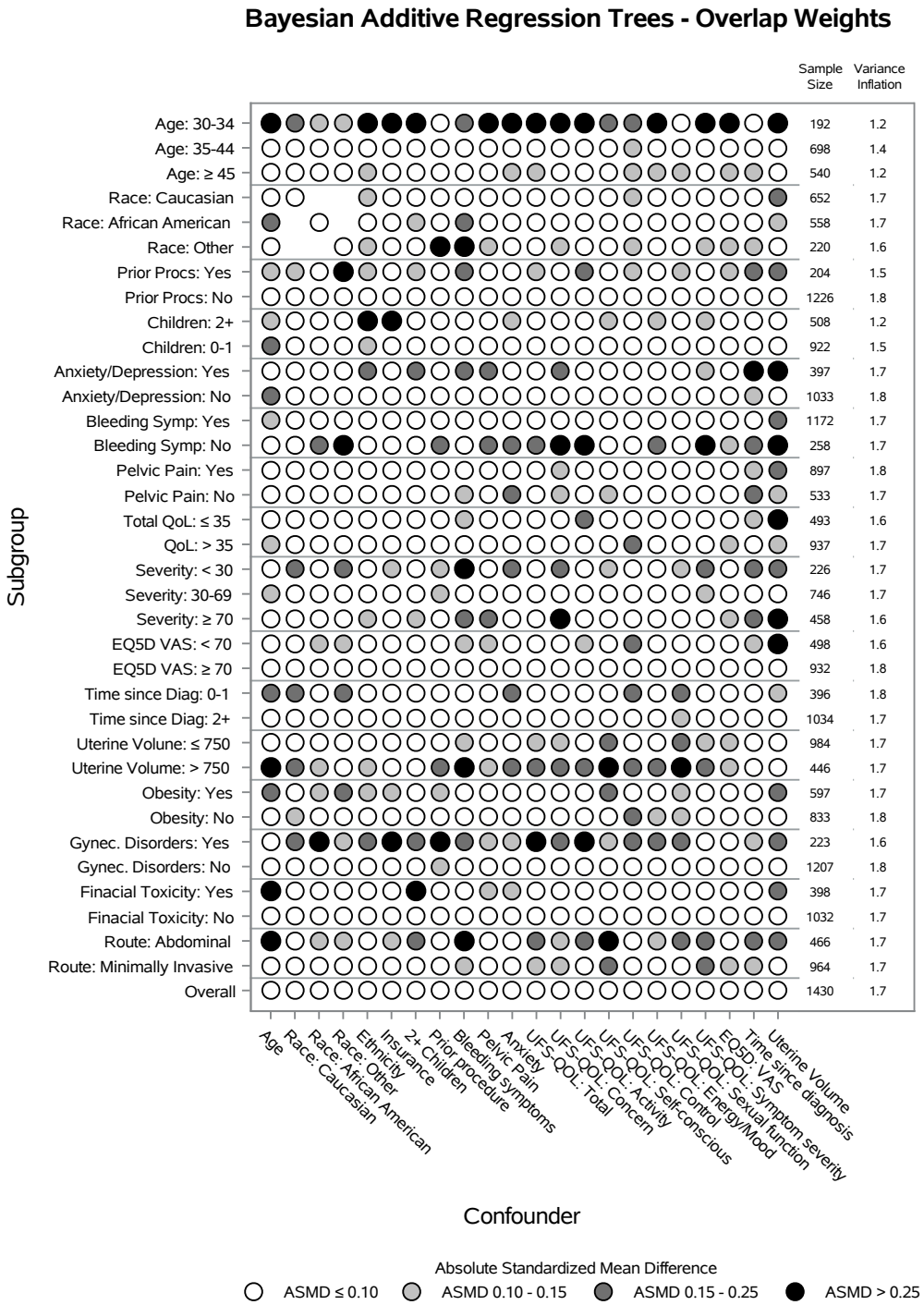

Main effects Logistic Regression - IP Weights
Sample Size
Variance Inflation
Age: 30-34 192 1.7
Age: 35-44 698 2.0
Age: ≥ 45 540 2.7
Race: Caucasian 652 4.2
Race: African American 558 2.7
Race: Other 220 2.7
Prior Procs: Yes 204 2.9
Prior Procs: No 1226 3.8
Children: 2+ 508 2.4
Children: 0-1 922 2.2
Anxiety/Depression: Yes 397 2.3
Anxiety/Depression: No 1033 4.1
Bleeding Symp: Yes 1172 3.8
Bleeding Symp: No 258 3.3
Pelvic Pain: Yes 897 2.9
Pelvic Pain: No 533 4.6
Total QoL: ≤ 35 493 2.0
QoL: > 35 937 4.2
Severity: < 30 226 3.6
Severity: 30-69 746 3.4
Severity: ≥ 70 458 4.1
EQ5D VAS: < 70 498 2.7
EQ5D VAS: ≥ 70 932 4.2
Time since Diag: 0-1 396 3.4
Time since Diag: 2+ 1034 3.8
Uterine Volume: ≤ 750 984 4.0
Uterine Volume: > 750 446 2.4
Obesity: Yes 597 2.8
Obesity: No 833 3.7
Gynec. Disorders: Yes 223 3.0
Gynec. Disorders: No 1207 3.9
Finacial Toxicity: Yes 398 1.9
Finacial Toxicity: No 1032 3.9
Route: Abdominal 466 2.1
Route: Minimally Invasive 964 4.3
Overall 1430 3.7
Subgroup
Age
Race: Caucasian
Race: African American
Race: Other
Ethnicity
Insurance
2+ Children
Prior procedure
Bleeding Symptoms
Pelvic Pain
Anxiety
UFS-QOL: Total
UFS-QOL: Concern
UFS-QOL: Activity
UFS-QOL: Self-conscious
UFS-QOL: Control
UFS-QOL: Energy/Mood
UFS-QOL: Sexual function
UFS-QOL: Symptom severity
EQ5D: VAS
Time since diagnosis
Uterine Volume
Confounder
Absolute Standardized Mean Difference
ASMD ≤ 0.10
ASMD 0.10 - 0.15
ASMD 0.15 - 0.25
ASMD > 0.25

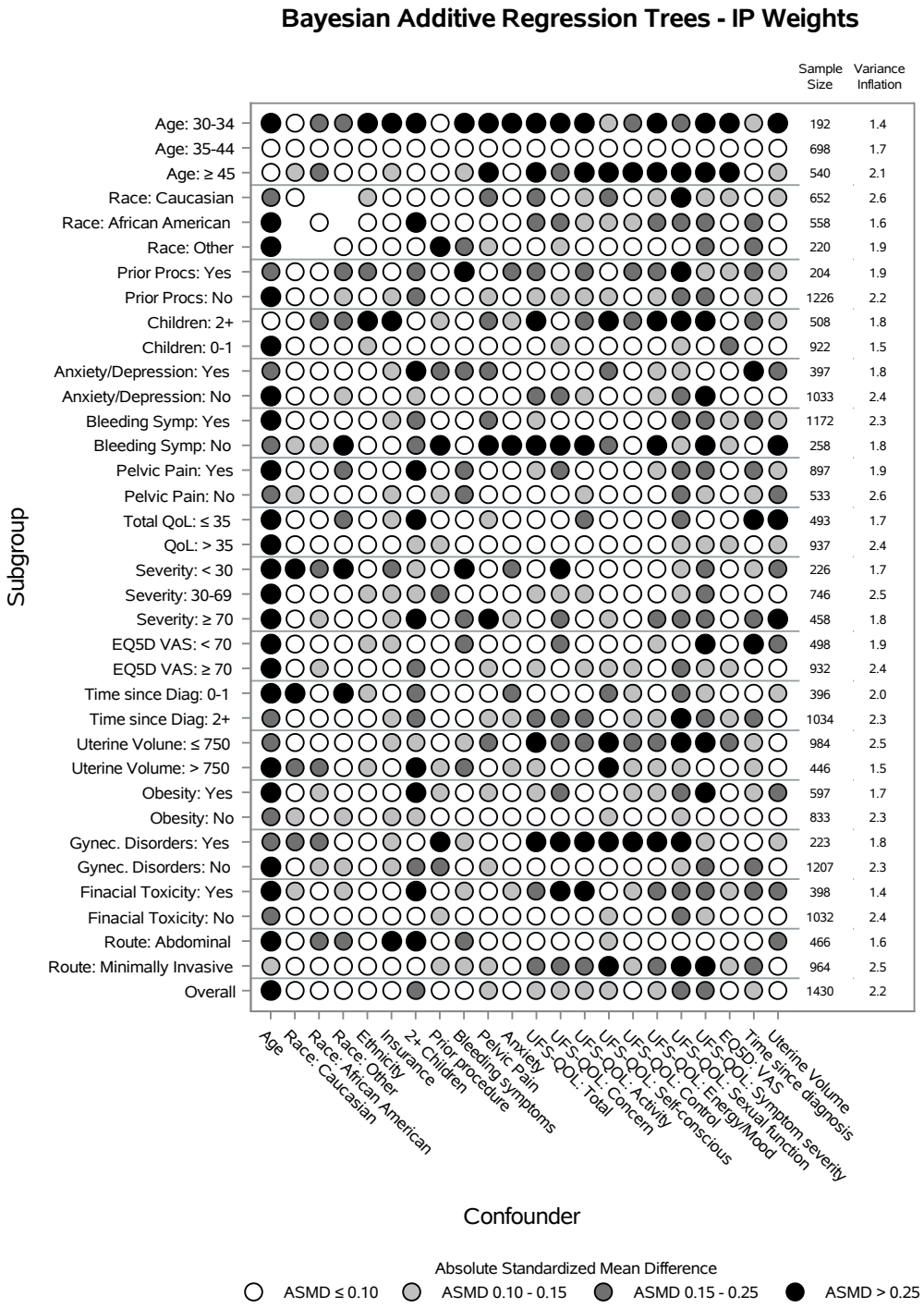

Generalized Boosted Model - Overlap Weights
Sample Size
Variance Inflation
Age: 30-34 192 1.3
Age: 35-44 698 1.4
Age: ≥ 45 540 1.1
Race: Caucasian 652 1.6
Race: African American 558 1.7
Race: Other 220 1.5
Prior Procs: Yes 204 1.5
Prior Procs: No 1226 1.7
Children: 2+ 508 1.1
Children: 0-1 922 1.5
Anxiety/Depression: Yes 397 1.6
Anxiety/Depression: No 1033 1.6
Bleeding Symp: Yes 1172 1.6
Bleeding Symp: No 258 1.6
Pelvic Pain: Yes 897 1.7
Pelvic Pain: No 533 1.6
Total QoL: ≤ 35 493 1.6
QoL: > 35 937 1.7
Severity: < 30 226 1.6
Severity: 30-69 746 1.6
Severity: ≥ 70 458 1.5
EQ5D VAS: < 70 498 1.6
EQ5D VAS: ≥ 70 932 1.7
Time since Diag: 0-1 396 1.7
Time since Diag: 2+ 1034 1.6
Uterine Volume: ≤ 750 984 1.7
Uterine Volume: > 750 446 1.6
Obesity: Yes 597 1.6
Obesity: No 833 1.7
Gynec. Disorders: Yes 223 1.5
Gynec. Disorders: No 1207 1.7
Finacial Toxicity: Yes 398 1.7
Finacial Toxicity: No 1032 1.6
Route: Abdominal 466 1.7
Route: Minimally Invasive 964 1.6
Overall 1430 1.6
Subgroup
Age
Race: Caucasian
Race: African American
Race: Other
Ethnicity
Insurance
2+ Children
Prior procedure
Bleeding Symptoms
Pelvic Pain
Anxiety
UFS-QOL: Total
UFS-QOL: Concern
UFS-QOL: Activity
UFS-QOL: Self-conscious
UFS-QOL: Control
UFS-QOL: Energy/Mood
UFS-QOL: Sexual function
EQ5D: VAS
Time since diagnosis
Uterine Volume
Confounder
Absolute Standardized Mean Difference
ASMD ≤ 0.10
ASMD 0.10 - 0.15
ASMD 0.15 - 0.25
ASMD > 0.25

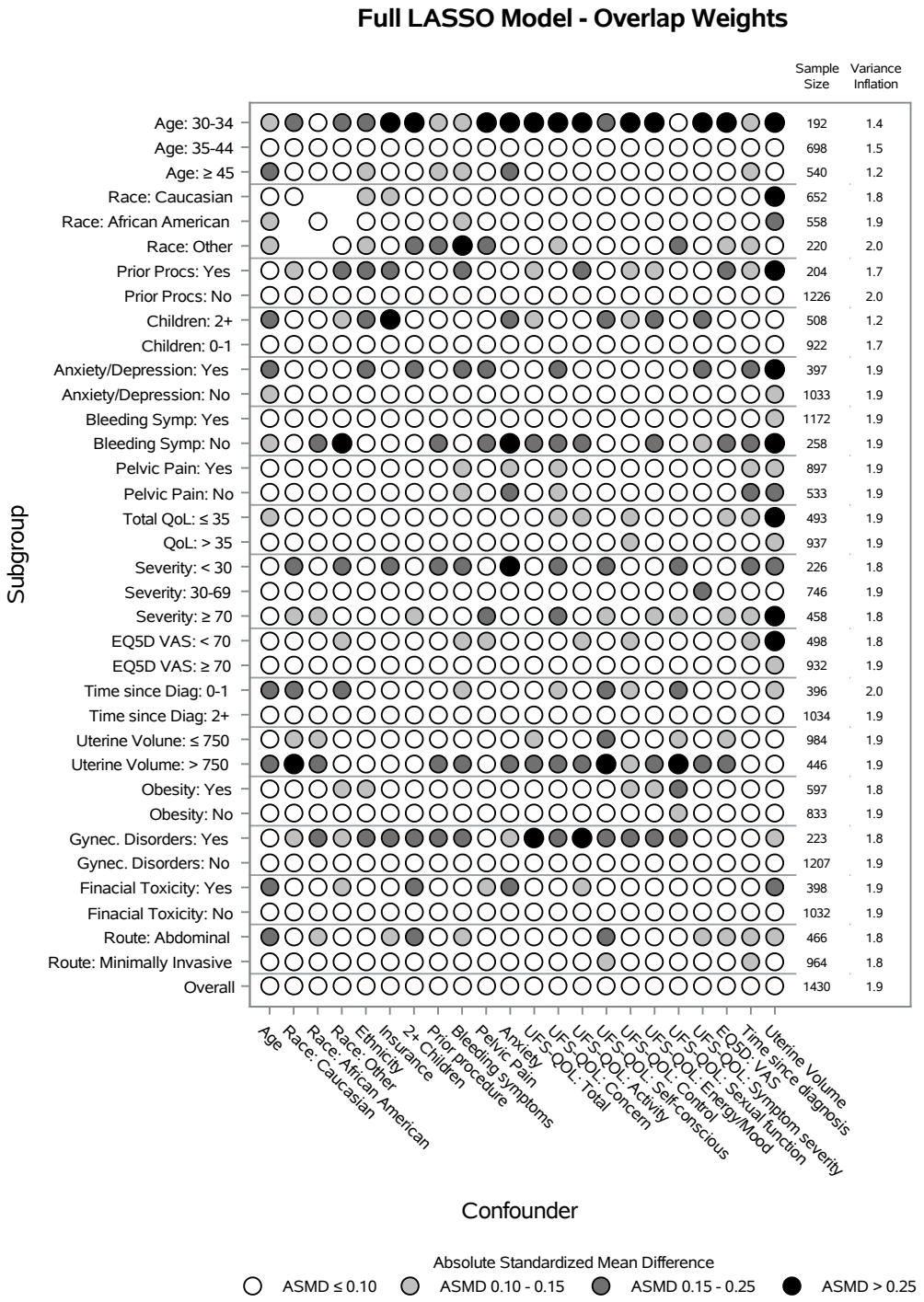

Generalized Boosted Model - IP Weights
Sample Size
Variance Inflation
Age: 30-34 192 1.6
Age: 35-44 698 1.5
Age: ≥ 45 540 1.4
Race: Caucasian 652 1.9
Race: African American 558 1.5
Race: Other 220 1.7
Prior Procs: Yes 204 1.6
Prior Procs: No 1226 1.7
Children: 2+ 508 1.3
Children: 0-1 922 1.5
Anxiety/Depression: Yes 397 1.8
Anxiety/Depression: No 1033 1.7
Bleeding Symp: Yes 1172 1.7
Bleeding Symp: No 258 1.6
Pelvic Pain: Yes 897 1.6
Pelvic Pain: No 533 1.8
Total QoL: ≤ 35 493 1.6
QoL: > 35 937 1.8
Severity: < 30 226 1.5
Severity: 30-69 746 1.8
Severity: ≥ 70 458 1.7
EQ5D VAS: < 70 498 1.6
EQ5D VAS: ≥ 70 932 1.8
Time since Diag: 0-1 396 1.8
Time since Diag: 2+ 1034 1.7
Uterine Volume: ≤ 750 984 1.8
Uterine Volume: > 750 446 1.5
Obesity: Yes 597 1.5
Obesity: No 833 1.8
Gynec. Disorders: Yes 223 1.7
Gynec. Disorders: No 1207 1.7
Finacial Toxicity: Yes 398 1.4
Finacial Toxicity: No 1032 1.8
Route: Abdominal 466 1.6
Route: Minimally Invasive 964 1.8
Overall 1430 1.7
Subgroup
Age
Race: Caucasian
Race: African American
Race: Other
Ethnicity
Insurance
2+ Children
Prior procedure
Bleeding Symptoms
Pelvic Pain
Anxiety
UFS-QOL: Total
UFS-QOL: Concern
UFS-QOL: Activity
UFS-QOL: Self-conscious
UFS-QOL: Control
UFS-QOL: Energy/Mood
UFS-QOL: Sexual function
EQ5D: VAS
UFS-QOL: Symptom severity
Time since diagnosis
Uterine Volume
Confounder
Absolute Standardized Mean Difference
ASMD ≤ 0.10
ASMD 0.10 - 0.15
ASMD 0.15 - 0.25
ASMD > 0.25


Random Forest - Overlap Weights
Sample Size
Variance Inflation
Age: 30-34 192 1.2
Age: 35-44 698 1.3
Age: ≥ 45 540 1.1
Race: Caucasian 652 1.4
Race: African American 558 1.5
Race: Other 220 1.3
Prior Procs: Yes 204 1.3
Prior Procs: No 1226 1.5
Children: 2+ 508 1.1
Children: 0-1 922 1.3
Anxiety/Depression: Yes 397 1.4
Anxiety/Depression: No 1033 1.4
Bleeding Symp: Yes 1172 1.4
Bleeding Symp: No 258 1.4
Pelvic Pain: Yes 897 1.5
Pelvic Pain: No 533 1.4
Total QoL: ≤ 35 493 1.4
QoL: > 35 937 1.4
Severity: < 30 226 1.3
Severity: 30-69 746 1.4
Severity: ≥ 70 458 1.4
EQ5D VAS: < 70 498 1.4
EQ5D VAS: ≥ 70 932 1.5
Time since Diag: 0-1 396 1.5
Time since Diag: 2+ 1034 1.4
Uterine Volume: ≤ 750 984 1.4
Uterine Volume: > 750 446 1.4
Obesity: Yes 597 1.4
Obesity: No 833 1.4
Gynec. Disorders: Yes 223 1.3
Gynec. Disorders: No 1207 1.4
Finacial Toxicity: Yes 398 1.4
Finacial Toxicity: No 1032 1.4
Route: Abdominal 466 1.4
Route: Minimally Invasive 964 1.4
Overall 1430 1.4
Subgroup
Age
Race: Caucasian
Race: African American
Race: Other
Ethnicity
Insurance
2+ Children
Prior procedure
Bleeding Symptoms
Pelvic Pain
Anxiety
UFS-QOL: Total
UFS-QOL: Concern
UFS-QOL: Activity
UFS-QOL: Self-conscious
UFS-QOL: Control
UFS-QOL: Energy/Mood
UFS-QOL: Sexual function
EQ5D: VAS
UFS-QOL: Symptom severity
Time since diagnosis
Uterine Volume
Confounder
Absolute Standardized Mean Difference
ASMD ≤ 0.10
ASMD 0.10 - 0.15
ASMD 0.15 - 0.25
ASMD > 0.25

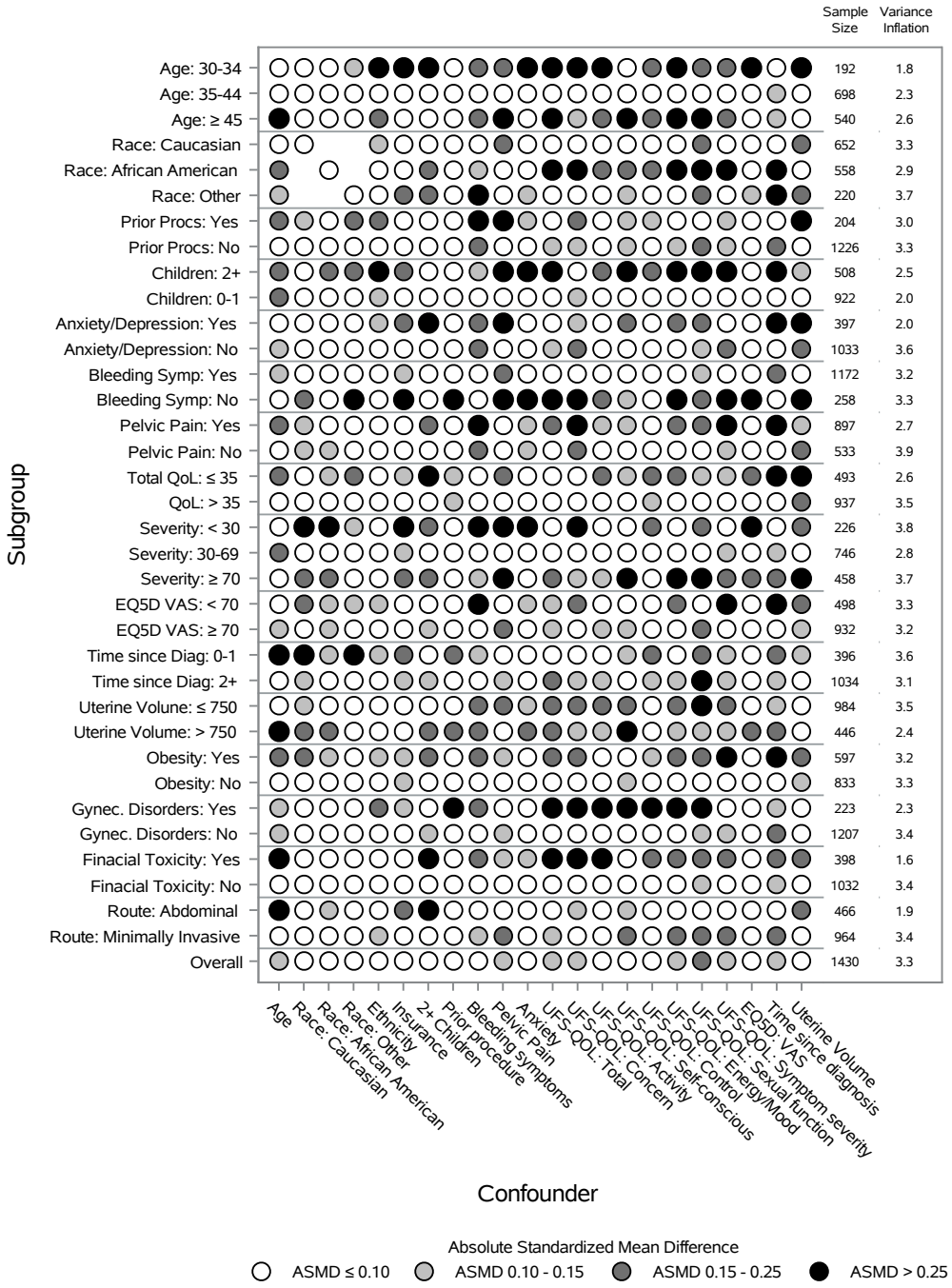

Random Forest - IP Weights
Sample Size
Variance Inflation
Age: 30-34 192 1.4
Age: 35-44 698 1.5
Age: ≥ 45 540 1.6
Race: Caucasian 652 1.8
Race: African American 558 1.6
Race: Other 220 1.6
Prior Procs: Yes 204 1.4
Prior Procs: No 1226 1.7
Children: 2+ 508 1.4
Children: 0-1 922 1.4
Anxiety/Depression: Yes 397 1.5
Anxiety/Depression: No 1033 1.8
Bleeding Symp: Yes 1172 1.8
Bleeding Symp: No 258 1.4
Pelvic Pain: Yes 897 1.6
Pelvic Pain: No 533 1.8
Total QoL: ≤ 35 493 1.6
QoL: > 35 937 1.8
Severity: < 30 226 1.4
Severity: 30-69 746 1.7
Severity: ≥ 70 458 1.8
EQ5D VAS: < 70 498 1.6
EQ5D VAS: ≥ 70 932 1.8
Time since Diag: 0-1 396 1.6
Time since Diag: 2+ 1034 1.7
Uterine Volume: ≤ 750 984 1.8
Uterine Volume: > 750 446 1.5
Obesity: Yes 597 1.5
Obesity: No 833 1.7
Gynec. Disorders: Yes 223 1.4
Gynec. Disorders: No 1207 1.8
Finacial Toxicity: Yes 398 1.4
Finacial Toxicity: No 1032 1.8
Route: Abdominal 466 1.5
Route: Minimally Invasive 964 1.8
Overall 1430 1.7
Subgroup
Age
Race: Caucasian
Race: African American
Race: Other
Ethnicity
Insurance
2+ Children
Prior procedure
Bleeding Symptoms
Pelvic Pain
Anxiety
UFS-QOL: Total
UFS-QOL: Concern
UFS-QOL: Activity
UFS-QOL: Self-conscious
UFS-QOL: Control
UFS-QOL: Energy/Mood
UFS-QOL: Sexual function
EQ5D: VAS
UFS-QOL: Symptom severity
Time since diagnosis
Uterine Volume
Confounder
Absolute Standardized Mean Difference
ASMD ≤ 0.10
ASMD 0.10 - 0.15
ASMD 0.15 - 0.25
ASMD > 0.25

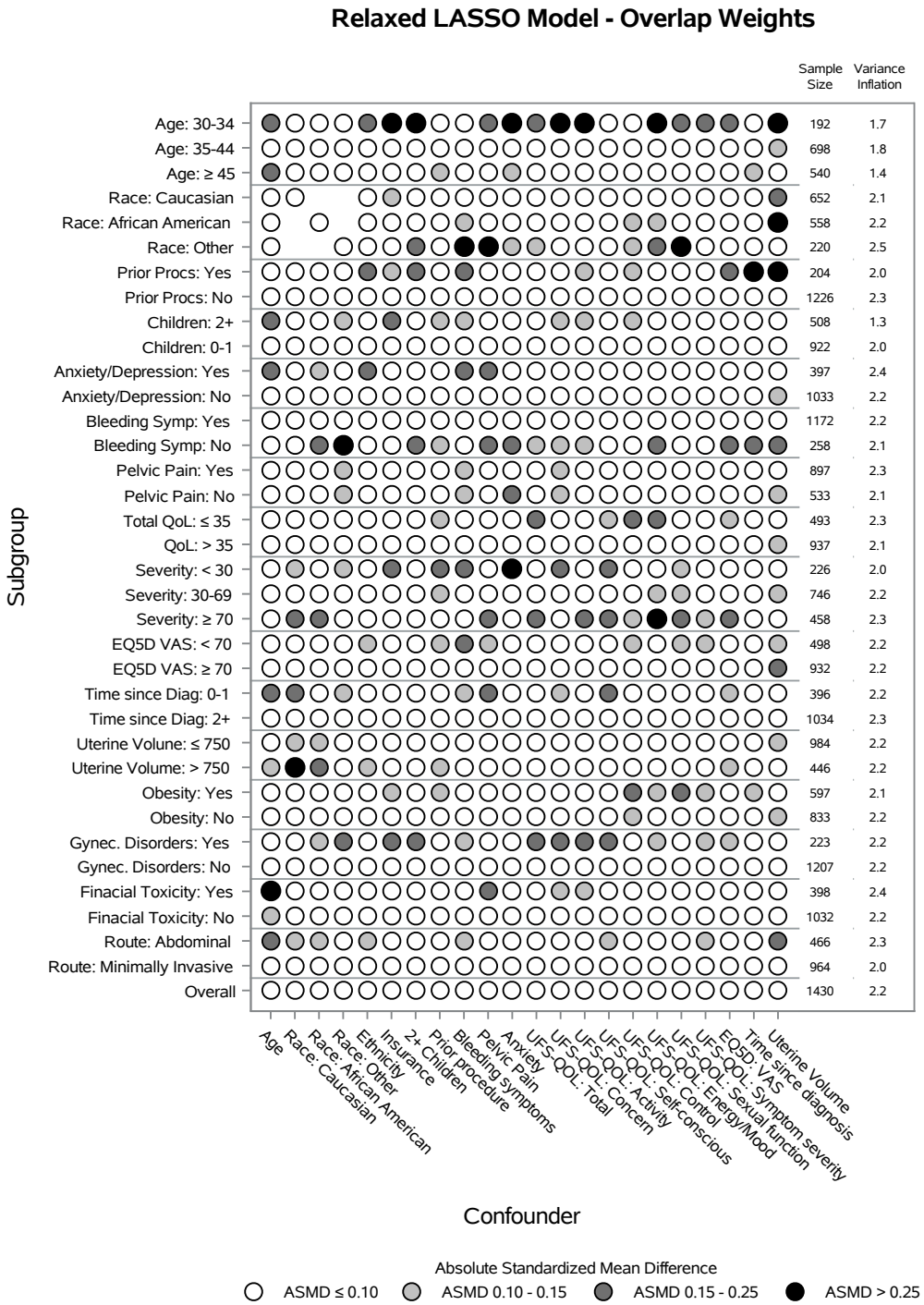

Bayesian Additive Regression Trees - Overlap Weights
Sample Size
Variance Inflation
Age: 30-34 192 1.2
Age: 35-44 698 1.4
Age: ≥ 45 540 1.2
Race: Caucasian 652 1.7
Race: African American 558 1.7
Race: Other 220 1.6
Prior Procs: Yes 204 1.5
Prior Procs: No 1226 1.8
Children: 2+ 508 1.2
Children: 0-1 922 1.5
Anxiety/Depression: Yes 397 1.7
Anxiety/Depression: No 1033 1.8
Bleeding Symp: Yes 1172 1.7
Bleeding Symp: No 258 1.7
Pelvic Pain: Yes 897 1.8
Pelvic Pain: No 533 1.7
Total QoL: ≤ 35 493 1.6
QoL: > 35 937 1.7
Severity: < 30 226 1.7
Severity: 30-69 746 1.7
Severity: ≥ 70 458 1.6
EQ5D VAS: < 70 498 1.6
EQ5D VAS: ≥ 70 932 1.8
Time since Diag: 0-1 396 1.8
Time since Diag: 2+ 1034 1.7
Uterine Volume: ≤ 750 984 1.7
Uterine Volume: > 750 446 1.7
Obesity: Yes 597 1.7
Obesity: No 833 1.8
Gynec. Disorders: Yes 223 1.6
Gynec. Disorders: No 1207 1.8
Finacial Toxicity: Yes 398 1.7
Finacial Toxicity: No 1032 1.7
Route: Abdominal 466 1.7
Route: Minimally Invasive 964 1.7
Overall 1430 1.7
Subgroup
Age
Race: Caucasian
Race: African American
Race: Other
Ethnicity
Insurance
2+ Children
Prior procedure
Bleeding Symptoms
Pelvic Pain
Anxiety
UFS-QOL: Total
UFS-QOL: Concern
UFS-QOL: Activity
UFS-QOL: Self-conscious
UFS-QOL: Control
UFS-QOL: Energy/Mood
UFS-QOL: Sexual function
EQ5D: VAS
Time since diagnosis
Uterine Volume
Confounder
Absolute Standardized Mean Difference
ASMD ≤ 0.10
ASMD 0.10 - 0.15
ASMD 0.15 - 0.25
ASMD > 0.25

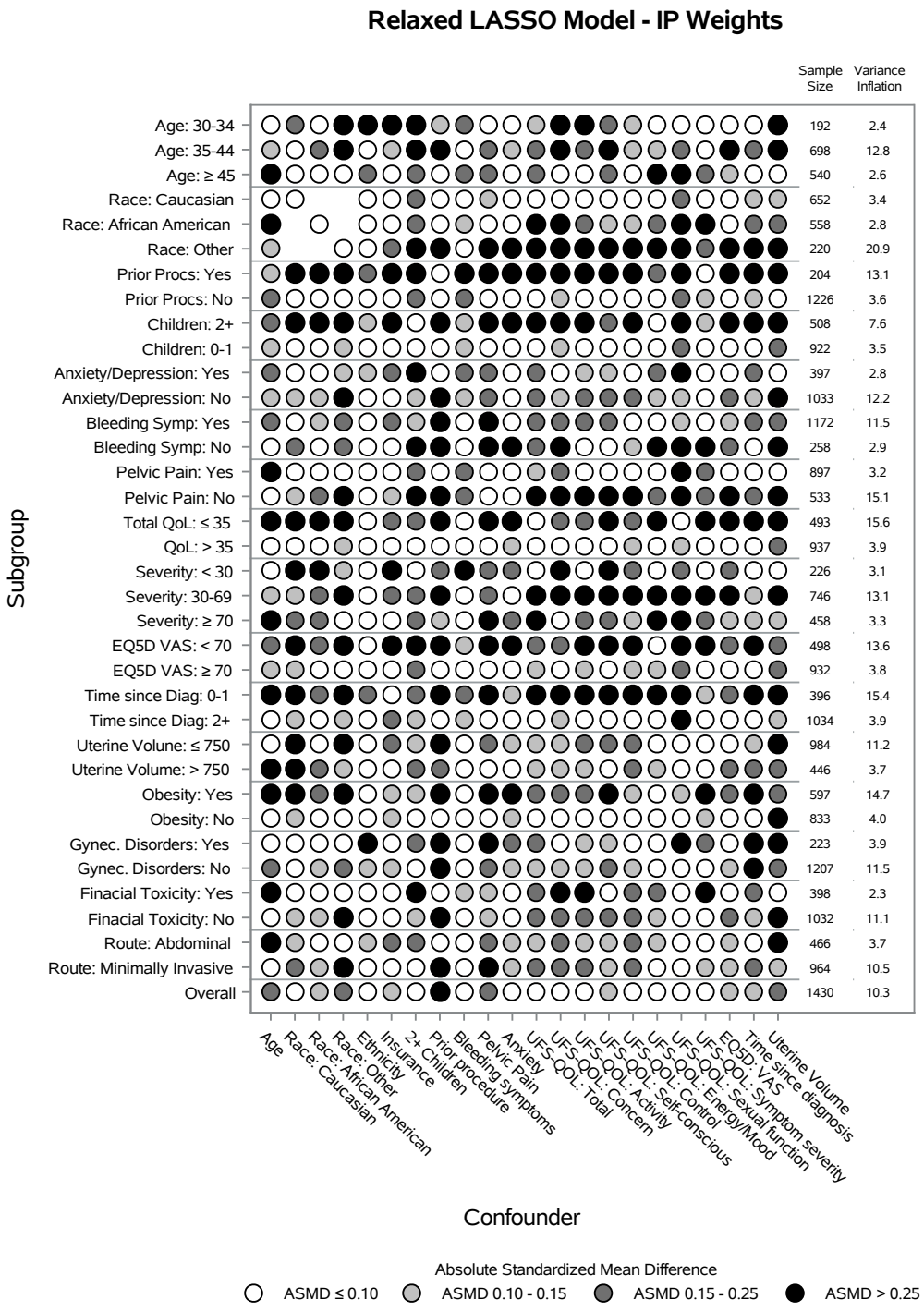

Bayesian Additive Regression Trees - IP Weights
Sample Size
Variance Inflation
Age: 30-34 192 1.4
Age: 35-44 698 1.7
Age: ≥ 45 540 2.1
Race: Caucasian 652 2.6
Race: African American 558 1.6
Race: Other 220 1.9
Prior Procs: Yes 204 1.9
Prior Procs: No 1226 2.2
Children: 2+ 508 1.8
Children: 0-1 922 1.5
Anxiety/Depression: Yes 397 1.8
Anxiety/Depression: No 1033 2.4
Bleeding Symp: Yes 1172 2.3
Bleeding Symp: No 258 1.8
Pelvic Pain: Yes 897 1.9
Pelvic Pain: No 533 2.6
Total QoL: ≤ 35 493 1.7
QoL: > 35 937 2.4
Severity: < 30 226 1.7
Severity: 30-69 746 2.5
Severity: ≥ 70 458 1.8
EQ5D VAS: < 70 498 1.9
EQ5D VAS: ≥ 70 932 2.4
Time since Diag: 0-1 396 2.0
Time since Diag: 2+ 1034 2.3
Uterine Volume: ≤ 750 984 2.5
Uterine Volume: > 750 446 1.5
Obesity: Yes 597 1.7
Obesity: No 833 2.3
Gynec. Disorders: Yes 223 1.8
Gynec. Disorders: No 1207 2.3
Finacial Toxicity: Yes 398 1.4
Finacial Toxicity: No 1032 2.4
Route: Abdominal 466 1.6
Route: Minimally Invasive 964 2.5
Overall 1430 2.2
Subgroup
Age
Race: Caucasian
Race: African American
Race: Other
Ethnicity
Insurance
2+ Children
Prior procedure
Bleeding Symptoms
Pelvic Pain
Anxiety
UFS-QOL: Total
UFS-QOL: Concern
UFS-QOL: Activity
UFS-QOL: Self-conscious
UFS-QOL: Control
UFS-QOL: Energy/Mood
UFS-QOL: Sexual function
EQ5D: VAS
UFS-QOL: Symptom severity
Time since diagnosis
Uterine Volume
Confounder
Absolute Standardized Mean Difference
ASMD ≤ 0.10
ASMD 0.10 - 0.15
ASMD 0.15 - 0.25
ASMD > 0.25


Full LASSO Model - Overlap Weights
Sample Size
Variance Inflation
Age: 30-34 192 1.4
Age: 35-44 698 1.5
Age: ≥ 45 540 1.2
Race: Caucasian 652 1.8
Race: African American 558 1.9
Race: Other 220 2.0
Prior Procs: Yes 204 1.7
Prior Procs: No 1226 2.0
Children: 2+ 508 1.2
Children: 0-1 922 1.7
Anxiety/Depression: Yes 397 1.9
Anxiety/Depression: No 1033 1.9
Bleeding Symp: Yes 1172 1.9
Bleeding Symp: No 258 1.9
Pelvic Pain: Yes 897 1.9
Pelvic Pain: No 533 1.9
Total QoL: ≤ 35 493 1.9
QoL: > 35 937 1.9
Severity: < 30 226 1.8
Severity: 30-69 746 1.9
Severity: ≥ 70 458 1.8
EQ5D VAS: < 70 498 1.8
EQ5D VAS: ≥ 70 932 1.9
Time since Diag: 0-1 396 2.0
Time since Diag: 2+ 1034 1.9
Uterine Volume: ≤ 750 984 1.9
Uterine Volume: > 750 446 1.9
Obesity: Yes 597 1.8
Obesity: No 833 1.9
Gynec. Disorders: Yes 223 1.8
Gynec. Disorders: No 1207 1.9
Finacial Toxicity: Yes 398 1.9
Finacial Toxicity: No 1032 1.9
Route: Abdominal 466 1.8
Route: Minimally Invasive 964 1.8
Overall 1430 1.9
Subgroup
Age
Race: Caucasian
Race: African American
Race: Other
Ethnicity
Insurance
2+ Children
Prior procedure
Bleeding Symptoms
Pelvic Pain
Anxiety
UFS-QOL: Total
UFS-QOL: Concern
UFS-QOL: Activity
UFS-QOL: Self-conscious
UFS-QOL: Control
UFS-QOL: Energy/Mood
UFS-QOL: Sexual function
EQ5D: VAS
Time since diagnosis
Uterine Volume
Confounder
Absolute Standardized Mean Difference
ASMD ≤ 0.10
ASMD 0.10 - 0.15
ASMD 0.15 - 0.25
ASMD > 0.25

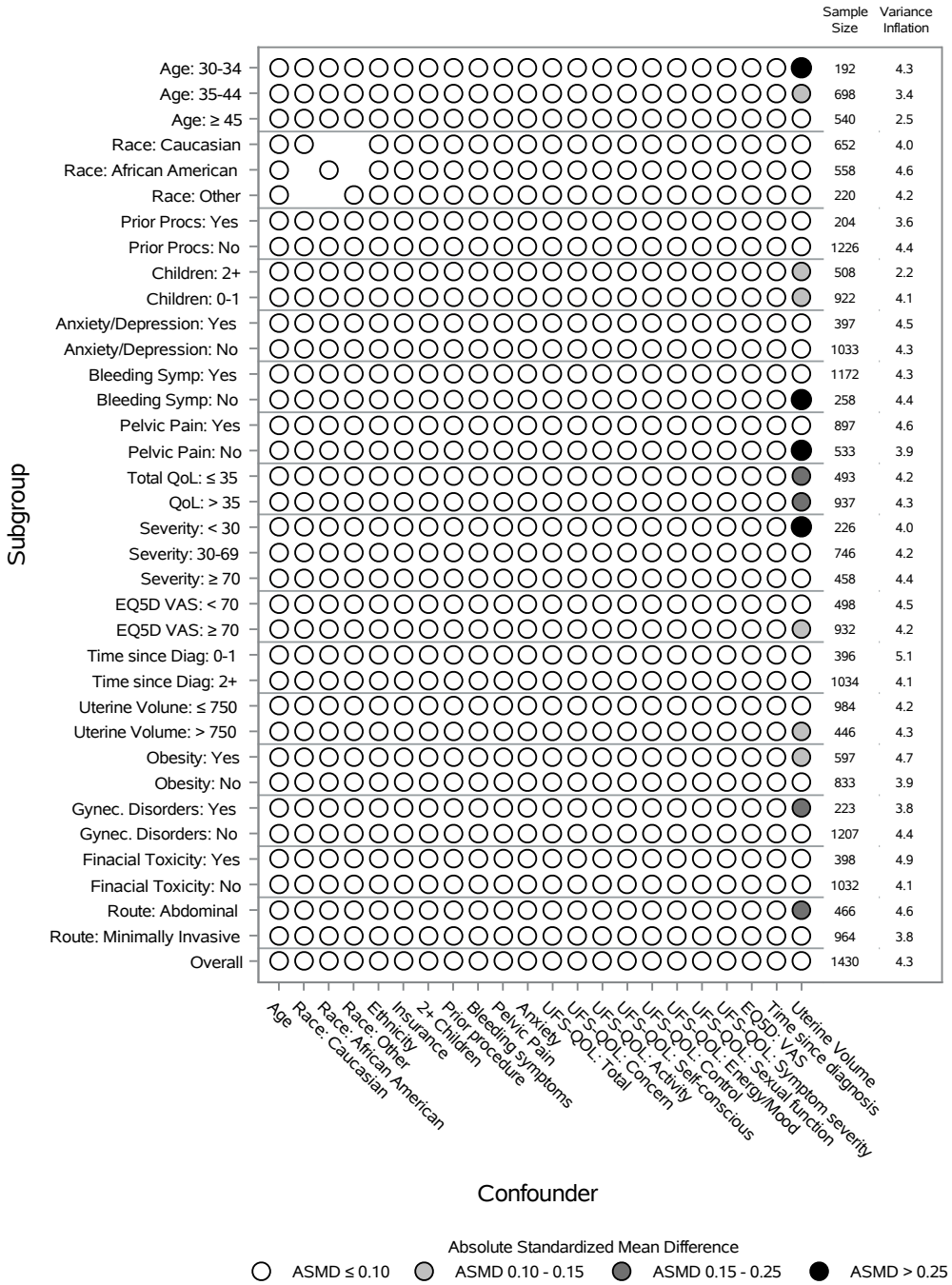
Full LASSO Model- IP Weights

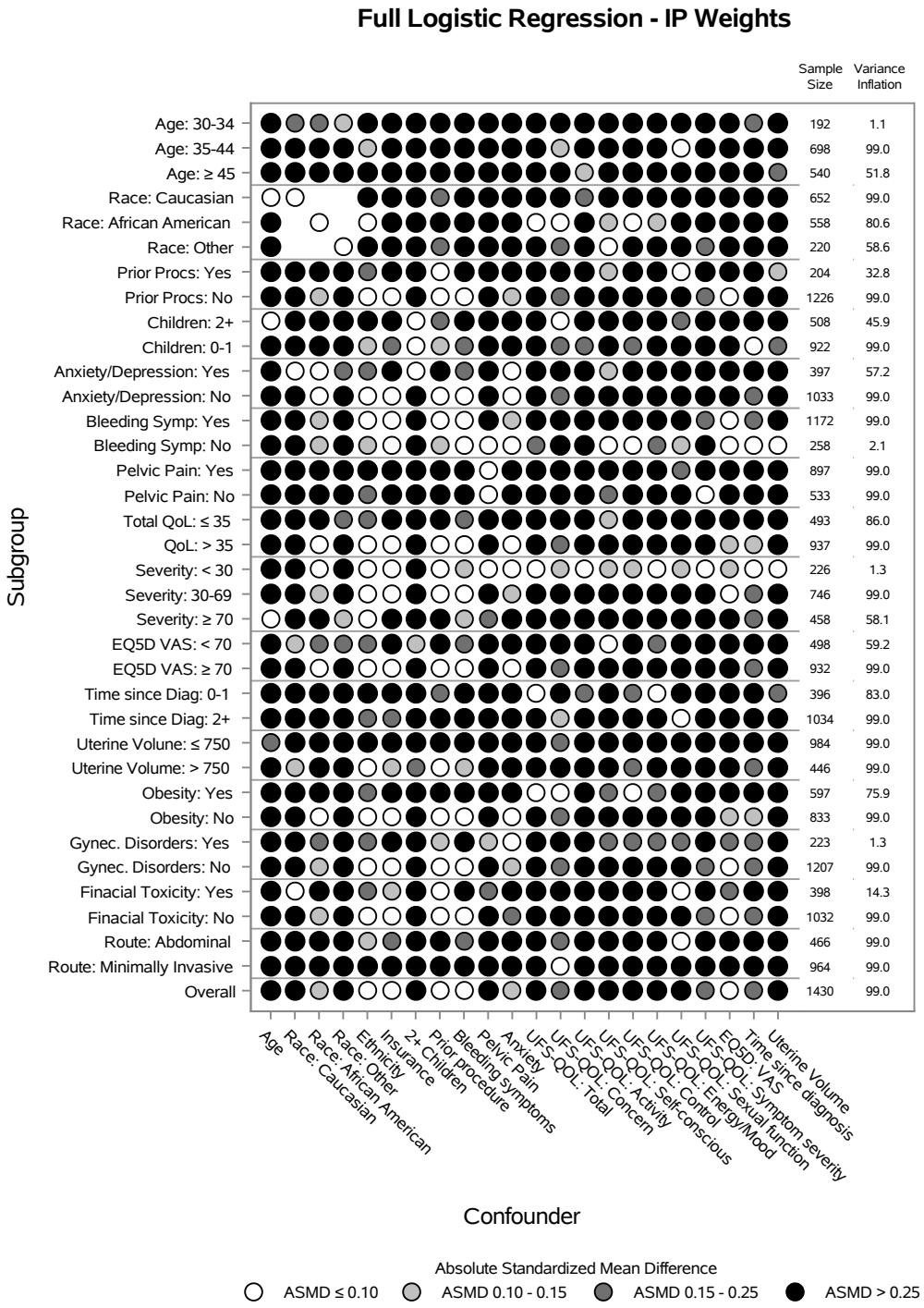



Relaxed LASSO Model - Overlap Weights
Sample Size
Variance Inflation
Age: 30-34 192 1.7
Age: 35-44 698 1.8
Age: ≥ 45 540 1.4
Race: Caucasian 652 2.1
Race: African American 558 2.2
Race: Other 220 2.5
Prior Procs: Yes 204 2.0
Prior Procs: No 1226 2.3
Children: 2+ 508 1.3
Children: 0-1 922 2.0
Anxiety/Depression: Yes 397 2.4
Anxiety/Depression: No 1033 2.2
Bleeding Symp: Yes 1172 2.2
Bleeding Symp: No 258 2.1
Pelvic Pain: Yes 897 2.3
Pelvic Pain: No 533 2.1
Total QoL: ≤ 35 493 2.3
QoL: > 35 937 2.1
Severity: < 30 226 2.0
Severity: 30-69 746 2.2
Severity: ≥ 70 458 2.3
EQ5D VAS: < 70 498 2.2
EQ5D VAS: ≥ 70 932 2.2
Time since Diag: 0-1 396 2.2
Time since Diag: 2+ 1034 2.3
Uterine Volume: ≤ 750 984 2.2
Uterine Volume: > 750 446 2.2
Obesity: Yes 597 2.1
Obesity: No 833 2.2
Gynec. Disorders: Yes 223 2.2
Gynec. Disorders: No 1207 2.2
Finacial Toxicity: Yes 398 2.4
Finacial Toxicity: No 1032 2.2
Route: Abdominal 466 2.3
Route: Minimally Invasive 964 2.0
Overall 1430 2.2
Subgroup
Age
Race: Caucasian
Race: African American
Race: Other
Ethnicity
Insurance
2+ Children
Prior procedure
Bleeding Symptoms
Pelvic Pain
Anxiety
UFS-QOL: Total
UFS-QOL: Concern
UFS-QOL: Activity
UFS-QOL: Self-conscious
UFS-QOL: Control
UFS-QOL: Energy/Mood
UFS-QOL: Sexual function
EQ5D: VAS
Time since diagnosis
Uterine Volume
Confounder
Absolute Standardized Mean Difference
ASMD ≤ 0.10
ASMD 0.10 - 0.15
ASMD 0.15 - 0.25
ASMD > 0.25


Relaxed LASSO Model - IP Weights
Sample Size
Variance Inflation
Age: 30-34 192 2.4
Age: 35-44 698 12.8
Age: ≥ 45 540 2.6
Race: Caucasian 652 3.4
Race: African American 558 2.8
Race: Other 220 20.9
Prior Procs: Yes 204 13.1
Prior Procs: No 1226 3.6
Children: 2+ 508 7.6
Children: 0-1 922 3.5
Anxiety/Depression: Yes 397 2.8
Anxiety/Depression: No 1033 12.2
Bleeding Symp: Yes 1172 11.5
Bleeding Symp: No 258 2.9
Pelvic Pain: Yes 897 3.2
Pelvic Pain: No 533 15.1
Total QoL: ≤ 35 493 15.6
QoL: > 35 937 3.9
Severity: < 30 226 3.1
Severity: 30-69 746 13.1
Severity: ≥ 70 458 3.3
EQ5D VAS: < 70 498 13.6
EQ5D VAS: ≥ 70 932 3.8
Time since Diag: 0-1 396 15.4
Time since Diag: 2+ 1034 3.9
Uterine Volume: ≤ 750 984 11.2
Uterine Volume: > 750 446 3.7
Obesity: Yes 597 14.7
Obesity: No 833 4.0
Gynec. Disorders: Yes 223 3.9
Gynec. Disorders: No 1207 11.5
Finacial Toxicity: Yes 398 2.3
Finacial Toxicity: No 1032 11.1
Route: Abdominal 466 3.7
Route: Minimally Invasive 964 10.5
Overall 1430 10.3
Subgroup
Age
Race: Caucasian
Race: African American
Race: Other
Ethnicity
Insurance
2+ Children
Prior procedure
Bleeding Symptoms
Pelvic Pain
Anxiety
UFS-QOL: Total
UFS-QOL: Concern
UFS-QOL: Activity
UFS-QOL: Self-conscious
UFS-QOL: Control
UFS-QOL: Energy/Mood
UFS-QOL: Sexual function
EQ5D: VAS
Time since diagnosis
Uterine Volume
Confounder
Absolute Standardized Mean Difference
ASMD ≤ 0.10
ASMD 0.10 - 0.15
ASMD 0.15 - 0.25
ASMD > 0.25

Full Logistic Regression - Overlap Weights




Full Logistic Regression - IP Weights
Sample Size
Variance Inflation
Age: 30-34 192 1.1
Age: 35-44 698 99.0
Age: ≥ 45 540 51.8
Race: Caucasian 652 99.0
Race: African American 558 80.6
Race: Other 220 58.6
Prior Procs: Yes 204 32.8
Prior Procs: No 1226 99.0
Children: 2+ 508 45.9
Children: 0-1 922 99.0
Anxiety/Depression: Yes 397 57.2
Anxiety/Depression: No 1033 99.0
Bleeding Symp: Yes 1172 99.0
Bleeding Symp: No 258 2.1
Pelvic Pain: Yes 897 99.0
Pelvic Pain: No 533 99.0
Total QoL: ≤ 35 493 86.0
QoL: > 35 937 99.0
Severity: < 30 226 1.3
Severity: 30-69 746 99.0
Severity: ≥ 70 458 58.1
EQ5D VAS: < 70 498 59.2
EQ5D VAS: ≥ 70 932 99.0
Time since Diag: 0-1 396 83.0
Time since Diag: 2+ 1034 99.0
Uterine Volume: ≤ 750 984 99.0
Uterine Volume: > 750 446 99.0
Obesity: Yes 597 75.9
Obesity: No 833 99.0
Gynec. Disorders: Yes 223 1.3
Gynec. Disorders: No 1207 99.0
Finacial Toxicity: Yes 398 14.3
Finacial Toxicity: No 1032 99.0
Route: Abdominal 466 99.0
Route: Minimally Invasive 964 99.0
Overall 1430 99.0
Subgroup
Age
Race: Caucasian
Race: African American
Race: Other
Ethnicity
Insurance
2+ Children
Prior procedure
Bleeding Symptoms
Pelvic Pain
Anxiety
UFS-QOL: Total
UFS-QOL: Concern
UFS-QOL: Activity
UFS-QOL: Self-conscious
UFS-QOL: Control
UFS-QOL: Energy/Mood
UFS-QOL: Sexual function
EQ5D: VAS
Time since diagnosis
Uterine Volume
Confounder
Absolute Standardized Mean Difference
ASMD ≤ 0.10
ASMD 0.10 - 0.15
ASMD 0.15 - 0.25
ASMD > 0.25

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.

4. Ho DE, Imai K, King G, Stuart EA. Matching as nonparametric preprocessing for reducing model dependence in parametric causal inference. *Political analysis* 2007; 15(3): 199–236.

5. Hirano K, Imbens GW, Ridder G. Efficient Estimation of Average Treatment Effects Using the Estimated Propensity Score. *Econometrica* 2003; 71(4): 1161-1189.

6. Lunceford JK, Davidian M. Stratification and weighting via the propensity score in estimation of causal treatment effects: a comparative study. *Statistics in medicine* 2004; 23(19): 2937-2960.