Supplementary Material for:

Performance Evaluation of Regression Splines for Propensity Score Adjustment in Post-Market Safety Analysis with Multiple Treatments

1 Notation

For the $i = 1, \cdots, n$ individuals in the observed data, let $T_i \in \{0, 1, 2\}$ denote the treatment variable with observed value t_i, Y_i denote the binary outcome variable with observed value y_i, and $\boldsymbol{x}_i = (x_{i,1} \cdots x_{i,p})$ denote a vector of p pretreatment baseline covariates. The propensity score (PS) has three components indicating probability of assignment to each treatment: $\boldsymbol{e}_i = (e_{i,0}, e_{i,1}, e_{i,2})$, where $e_{i,t} = \Pr(T_i = t|\boldsymbol{x}_i), t = 0, 1, 2$. Because the three PS components sum to 1, when we use the PS directly in the outcome model regression we drop the first component $e_{i,0}$ (similar to using a scalar PS for two treatments). We use $\boldsymbol{e^*} = (e_1^*, e_2^*) = (\text{logit}(e_1), \text{logit}(e_2))$ to denote the logit of the PS.

2 Marginal Relative Risk Estimator

Let our logistic outcome model be $\text{logit}(\mathbb{E}(Y_i)) = \beta_0 + \beta_1 * I(t_i = 1) + \beta_2 * I(t_i = 2) + \boldsymbol{x}_{i,r}\boldsymbol{\beta}_r$, where I is an indicator function for the treatments, β_1 and β_2 are the effect sizes for treatments 1 and 2, and $\boldsymbol{x}_{i,r}$ represents a $(k-3)$-dimensional vector of all other outcome model components. When using IPTW, we fit a weighted outcome model with only treatment as covariates, so $\boldsymbol{x}_{i,r}$ is null. When using direct outcome model regression, $\boldsymbol{x}_{i,r} = \boldsymbol{x}_i$. When using spline methods, $\boldsymbol{x}_{i,r}$ are the values of $\boldsymbol{e}_i^*$ under the corresponding spline basis functions.

Let $\boldsymbol{\beta}$ be the k-dimensional vector of outome model parameters $(\beta_0, \beta_1, \beta_2, \boldsymbol{\beta}_r)^T$ and $\hat{\boldsymbol{\beta}}$ be its maximum likelihood estimate. Let $\boldsymbol{X}_0, \boldsymbol{X}_1, \boldsymbol{X}_2$ be the n x k dimensional matrix

of outcome model parameter values $(1, I(t_i = 1), I(t_i = 2), \boldsymbol{x}_{i,r})$ when all patients are on treatments 0, 1, and 2, respectively. Also, define $\boldsymbol{X} = (\boldsymbol{X}_0^T, \boldsymbol{X}_1^T, \boldsymbol{X}_2^T)^T$ to be a 3n x k matrix of all outcome model parameter values for all treatments.

Using treatment 0 as the reference, the marginal risk estimators for treatments 1 and 2 are respectively

$$RR_1 = \frac{\bar{p}_1}{\bar{p}_0}, RR_2 = \frac{\bar{p}_2}{\bar{p}_0} \tag{1}$$

where $\bar{p}_k = \frac{1}{n}\sum_{i=1}^{n} \hat{p}_{i,k}$, $k = 0, 1, 2$ is the average outcome when all patients are on the same treatment k. We express the relative risk estimators as functions of these average outcomes, as in $RR_1 = \bar{p}_1/\bar{p}_0 \equiv f(\bar{p}_0, \bar{p}_1)$. Using the multivariate Delta method, the variance of the RR_1 estimator is

$$\begin{aligned}\mathbb{V}(f(\bar{p}_0, \bar{p}_1)) &= [\frac{\partial f(\bar{p}_0, \bar{p}_1)}{\partial \bar{p}_0}]^2 \mathbb{V}(\bar{p}_0) + [\frac{\partial f(\bar{p}_0, \bar{p}_1)}{\partial \bar{p}_1}]^2 \mathbb{V}(\bar{p}_1) + 2\frac{\partial f(\bar{p}_0, \bar{p}_1)}{\partial \bar{p}_0}\frac{\partial f(\bar{p}_0, \bar{p}_1)}{\partial \bar{p}_1}\mathrm{Cov}(\bar{p}_0, \bar{p}_1) \\ &= \frac{\bar{p}_1^2}{\bar{p}_0^4}\mathbb{V}(\bar{p}_0) + \frac{1}{\bar{p}_0^2}\mathbb{V}(\bar{p}_1) - 2\frac{\bar{p}_1}{\bar{p}_0^3}\mathrm{Cov}(\bar{p}_0, \bar{p}_1)\end{aligned} \tag{2}$$

We obtain the variances and covariances of $\bar{p}_0$ and $\bar{p}_1$ using the estimated variance-covariace matrix of $\hat{\boldsymbol{\beta}}$ and another application of the Delta method, which we detail below.

Let $\hat{\boldsymbol{p}}_0 = (\hat{p}_{1,0}, \cdots, \hat{p}_{n,0})$ be the n-dimensional vector of predicted outcomes under treatment 0, with corresponding definitions for $\hat{\boldsymbol{p}}_1$ and $\hat{\boldsymbol{p}}_2$. The 3n-dimensional vector $\hat{\boldsymbol{p}} = (\hat{\boldsymbol{p}}_0, \hat{\boldsymbol{p}}_1, \hat{\boldsymbol{p}}_2)$ has expected value $\hat{\boldsymbol{p}} = h(\boldsymbol{X}\hat{\boldsymbol{\beta}})$, where $h(\theta)$ is the inverse logit function of linear predictor θ.

We assume $\hat{\boldsymbol{\beta}}$ is asymptotically normal, and use its estimated variance-covariance matrix $\mathbb{V}(\hat{\boldsymbol{\beta}}) = \Sigma$ and the Delta method to find the variance of the the predicted outcomes $\hat{\boldsymbol{p}}$:

$$\begin{aligned}\mathbb{V}(\hat{\boldsymbol{p}}) &= h'(\boldsymbol{X}\hat{\boldsymbol{\beta}})\boldsymbol{X}\mathbb{V}(\hat{\boldsymbol{\beta}})\boldsymbol{X}^T h'(\boldsymbol{X}\hat{\boldsymbol{\beta}})^T \\ &= h'(\boldsymbol{X}\hat{\boldsymbol{\beta}})\boldsymbol{X}\Sigma\boldsymbol{X}^T h'(\boldsymbol{X}\hat{\boldsymbol{\beta}})^T\end{aligned} \tag{3}$$

where $h'(\boldsymbol{X}\hat{\boldsymbol{\beta}})$ is a 3n x 3n diagonal matrix of inverse logit derivatives. We need the

variance of $\bar{p}_0$, which can be written as a function of $\hat{\boldsymbol{p}}$:

$$\bar{p}_0 = \frac{1}{n}\hat{\boldsymbol{p}}^T[\mathbb{1}, \mathbb{0}, \mathbb{0}] \tag{4}$$

where $\mathbb{0}$ and $\mathbb{1}$ are n-dimensional vectors of 0 and 1, respectively. The variance is then

$$\begin{aligned}
\mathbb{V}(\bar{p}_0) &= \frac{1}{n^2}[\mathbb{1}, \mathbb{0}, \mathbb{0}]^T \mathbb{V}(\hat{p})[\mathbb{1}, \mathbb{0}, \mathbb{0}] \\
&= \frac{1}{n^2}[\mathbb{1}, \mathbb{0}, \mathbb{0}]^T h'(\boldsymbol{X}\hat{\boldsymbol{\beta}})\boldsymbol{X}\Sigma\boldsymbol{X}^T h'(\boldsymbol{X}\hat{\boldsymbol{\beta}})^T[\mathbb{1}, \mathbb{0}, \mathbb{0}] \\
&= \frac{1}{n^2}\mathbb{1}^T h'(\boldsymbol{X}_0\hat{\boldsymbol{\beta}})\boldsymbol{X}_0\Sigma\boldsymbol{X}_0^T h'(\boldsymbol{X}_0\hat{\boldsymbol{\beta}})^T\mathbb{1}
\end{aligned} \tag{5}$$

Likewise, $\bar{p}_1$ is the second component of $\hat{\boldsymbol{p}}$, $\bar{p}_1 = \frac{1}{n}[\mathbb{0}\mathbb{1}\mathbb{0}]\hat{\boldsymbol{p}}$, and its variance is

$$\begin{aligned}
\mathbb{V}(\bar{p}_1) &= \frac{1}{n^2}[\mathbb{0}, \mathbb{1}, \mathbb{0}]^T \mathbb{V}(\hat{\boldsymbol{p}})[\mathbb{0}, \mathbb{1}, \mathbb{0}] \\
&= \frac{1}{n^2}[\mathbb{0}, \mathbb{1}, \mathbb{0}]^T h'(\boldsymbol{X}\hat{\boldsymbol{\beta}})\boldsymbol{X}\Sigma\boldsymbol{X}^T h'(\boldsymbol{X}\hat{\boldsymbol{\beta}})^T[\mathbb{0}, \mathbb{1}, \mathbb{0}] \\
&= \frac{1}{n^2}\mathbb{1}^T h'(\boldsymbol{X}_1\hat{\boldsymbol{\beta}})\boldsymbol{X}_1\Sigma\boldsymbol{X}_1^T h'(\boldsymbol{X}_1\hat{\boldsymbol{\beta}})^T\mathbb{1}
\end{aligned}$$

Finally, the covariance of $\bar{p}_0$ and $\bar{p}_1$ is the off diagonal component of $\mathbb{V}(\hat{\boldsymbol{p}})$:

$$\begin{aligned}
\mathrm{Cov}(\bar{p}_0, \bar{p}_1) &= \frac{1}{n^2}[\mathbb{1}, \mathbb{0}, \mathbb{0}]^T \mathbb{V}(\hat{\boldsymbol{p}})[\mathbb{0}, \mathbb{1}, \mathbb{0}] \\
&= \frac{1}{n^2}\mathbb{1}^T h'(\boldsymbol{X}_0\hat{\boldsymbol{\beta}})\boldsymbol{X}_0\Sigma\boldsymbol{X}_1^T h'(\boldsymbol{X}_1\hat{\boldsymbol{\beta}})^T\mathbb{1}
\end{aligned}$$

3 Simulation Design

Our simulation design follows partly that of (Yoshida et al., 2017), which in turn follows that of (Franklin et al., 2014).

3.1 Covariate Generation

We simulate ten covariates with the following distributions:

Variable	Generation Process
X_1	Normal$(0, 1^2)$
X_2	Log-Normal$(0, 0.5^2)$
X_3	Normal$(0, 2^2)$
X_4	Bernoulli$(p = \exp(2X_1)/\exp(1 + \exp(2X_1)))$
X_5	Bernoulli$(p = 0.2)$
X_6	Multinomial$(p = (0.5, 0.3, 0.1, 0.05, 0.05)^T)$
X_7	$\sin(X_1)$
X_8	X_2^2
X_9	$X_3 * X_4$
X_{10}	$X_4 * X_5$

3.2 Treatment Generation

Each subject's treatment is generated from a multinomial distribution with probabilities $(e_{i,0}, e_{i,1}, e_{i,2})$. The probabilities represent propensity scores $e_{i,t} = \Pr(T_i = t|\boldsymbol{x_i})$, $t = 0, 1, 2$, and are derived from a multinomial logistic model. We define the linear predictors for each treatment as:

$$\begin{aligned} \eta_{i,0} &= \boldsymbol{x}_i \boldsymbol{\alpha}_0 \\ \eta_{i,1} &= \boldsymbol{x}_i \boldsymbol{\alpha}_1 \\ \eta_{i,2} &= \boldsymbol{x}_i \boldsymbol{\alpha}_2 \end{aligned} \tag{6}$$

where $\boldsymbol{x}$ includes an intercept term $x_{i,0} = 1$ and the ten covariates $x_{i,j}, j = 1, \cdots 10$. Each patient's treatment probabilities for the three treatments are:

$$\begin{aligned} q_i &= \exp(\eta_{i,0}) + \exp(\eta_{i,1}) + \exp(\eta_{i,2}) \\ e_{i,0} &= \Pr(z_i = 0|\boldsymbol{x}_i) = \exp(\eta_{i,0})/q_i \\ e_{i,1} &= \Pr(z_i = 1|\boldsymbol{x}_i) = \exp(\eta_{i,1})/q_i \\ e_{i,2} &= \Pr(z_i = 2|\boldsymbol{x}_i) = \exp(\eta_{i,2})/q_i \end{aligned} \tag{7}$$

We select $\boldsymbol{\alpha}_1$, $\boldsymbol{\alpha}_2$ and $\boldsymbol{\alpha}_3$ to give desired treatment proportions and propensity score distributions. We simulate under 3 propensity score distributions that we judge to be "good", "fair", and "poor", and we simulate under two treatment prevalence proportions, equal(33:33:33) and unequal(10:45:45). For each of these 3 x 2 = 6 treatment generating

distributions, we present the coefficients and the resultant PS distributions for a sample of 5000 patients. The following pages detail each distribution.

3.2.1 Equal (33:33:33) treatment prevalence, "good" PS overlap

$$\boldsymbol{\alpha}_0 = \{0.300, -0.027, 0.158, 0.188, -0.290, -0.300, -0.130, 0.140, 0.170, -0.019, -0.338\}$$
$$\boldsymbol{\alpha}_1 = \{0.162, -0.040, 0.075, 0.181, -0.303, -0.048, 0.156, 0.051, -0.054, 0.395, -0.053\}$$
$$\boldsymbol{\alpha}_2 = \{-0.401, -0.121, -0.222, 0.206, 0.360, 0.122, 0.311, -0.207, 0.146, 0.006, -0.249\}$$

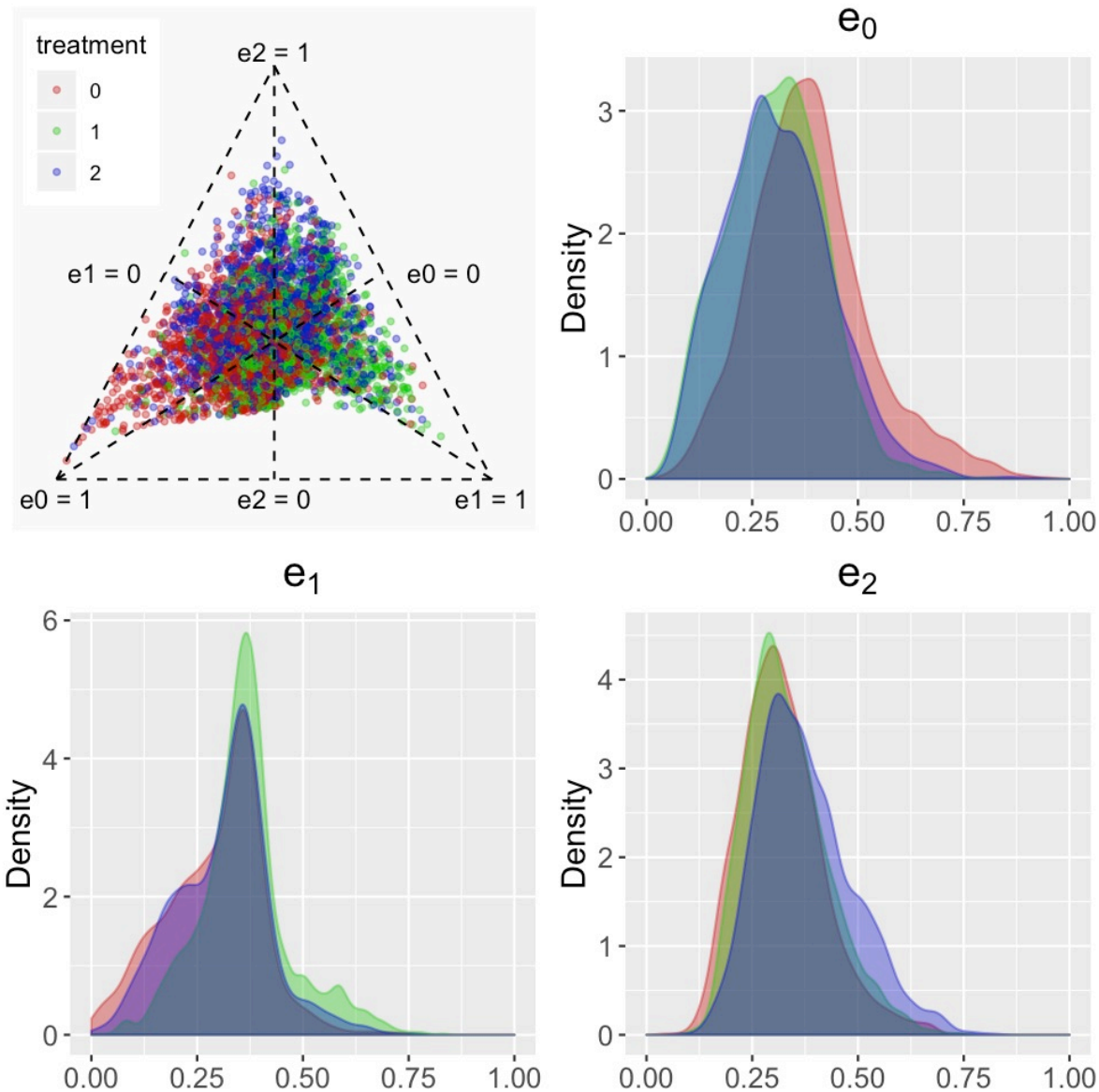


Figure 1: PS distribution for equal (33:33:33) treatment prevalence and good PS overlap. The three PS components (e_0, e_1, e_2) lie on a two-dimensional equilateral triangle in the plane $e_0 + e_1 + e_2 = 1$ and constrained by $0 < e_0, e_1, e_2 < 1$.

3.2.2 Equal (33:33:33) treatment prevalence, "fair" PS overlap

$$\boldsymbol{\alpha}_0 = \{0.599, -0.054, 0.316, 0.375, -0.580, -0.599, -0.261, 0.281, 0.339, -0.038, -0.676\}$$
$$\boldsymbol{\alpha}_1 = \{0.323, -0.080, 0.150, 0.362, -0.606, -0.096, 0.312, 0.102, -0.107, 0.790, -0.106\}$$
$$\boldsymbol{\alpha}_2 = \{-0.803, -0.242, -0.445, 0.411, 0.720, 0.244, 0.622, -0.415, 0.293, 0.011, -0.498\}$$

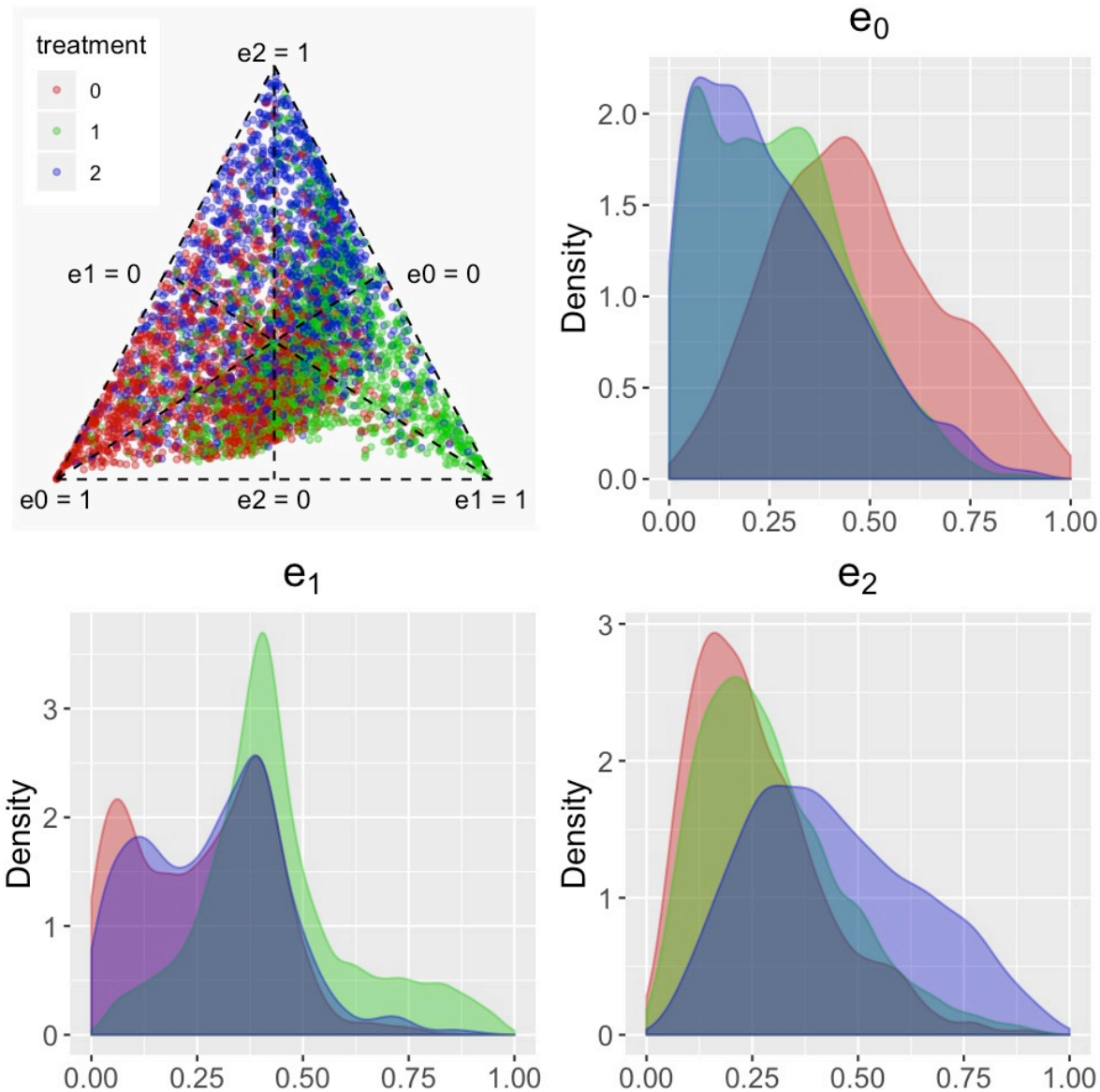


Figure 2: PS distribution for equal (33:33:33) treatment prevalence and fair PS overlap. The three PS components (e_0, e_1, e_2) lie on a two-dimensional equilateral triangle in the plane $e_0 + e_1 + e_2 = 1$ and constrained by $0 < e_0, e_1, e_2 < 1$.

3.2.3 Equal (33:33:33) treatment prevalence, "poor" PS overlap

$$\boldsymbol{\alpha}_0 = \{1.199, -0.107, 0.633, 0.751, -1.160, -1.198, -0.522, 0.561, 0.679, -0.077, -1.351\}$$
$$\boldsymbol{\alpha}_1 = \{0.646, -0.160, 0.301, 0.724, -1.211, -0.191, 0.623, 0.203, -0.215, 1.579, -0.212\}$$
$$\boldsymbol{\alpha}_2 = \{-1.605, -0.483, -0.889, 0.822, 1.439, 0.488, 1.244, -0.829, 0.586, 0.023, -0.997\}$$

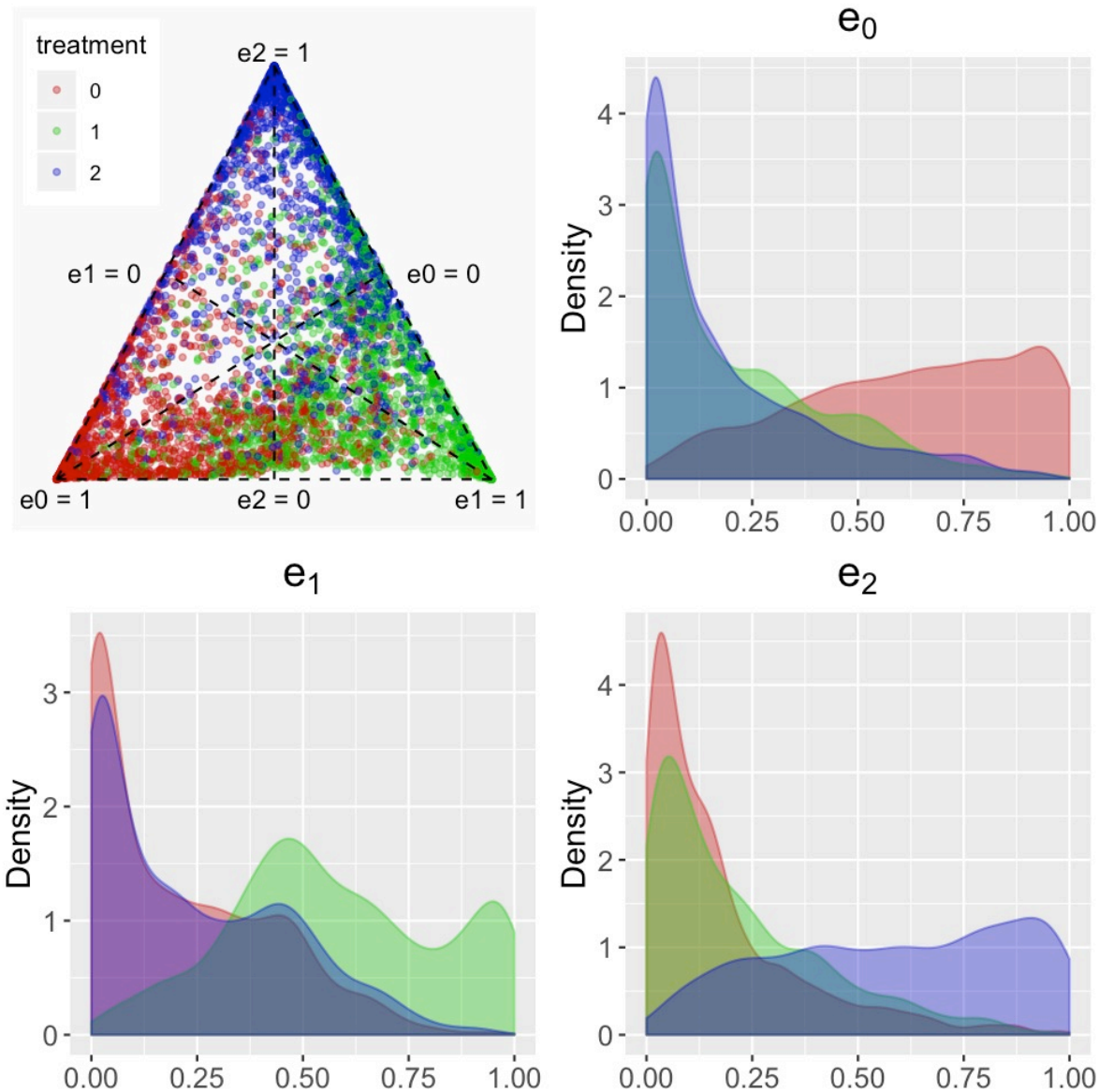


Figure 3: PS distribution for equal (33:33:33) treatment prevalence and poor PS overlap. The three PS components (e_0, e_1, e_2) lie on a two-dimensional equilateral triangle in the plane $e_0 + e_1 + e_2 = 1$ and constrained by $0 < e_0, e_1, e_2 < 1$.

3.2.4 Unequal (10:45:45) treatment prevalence, "good" PS overlap

$$\boldsymbol{\alpha}_0 = \{0.300, -0.027, 0.158, 0.188, -0.290, -0.300, -0.130, 0.140, 0.170, -0.019, -0.338\}$$
$$\boldsymbol{\alpha}_1 = \{1.762, -0.040, 0.075, 0.181, -0.303, -0.048, 0.156, 0.051, -0.054, 0.395, -0.053\}$$
$$\boldsymbol{\alpha}_2 = \{1.199, -0.121, -0.222, 0.206, 0.360, 0.122, 0.311, -0.207, 0.146, 0.006, -0.249\}$$

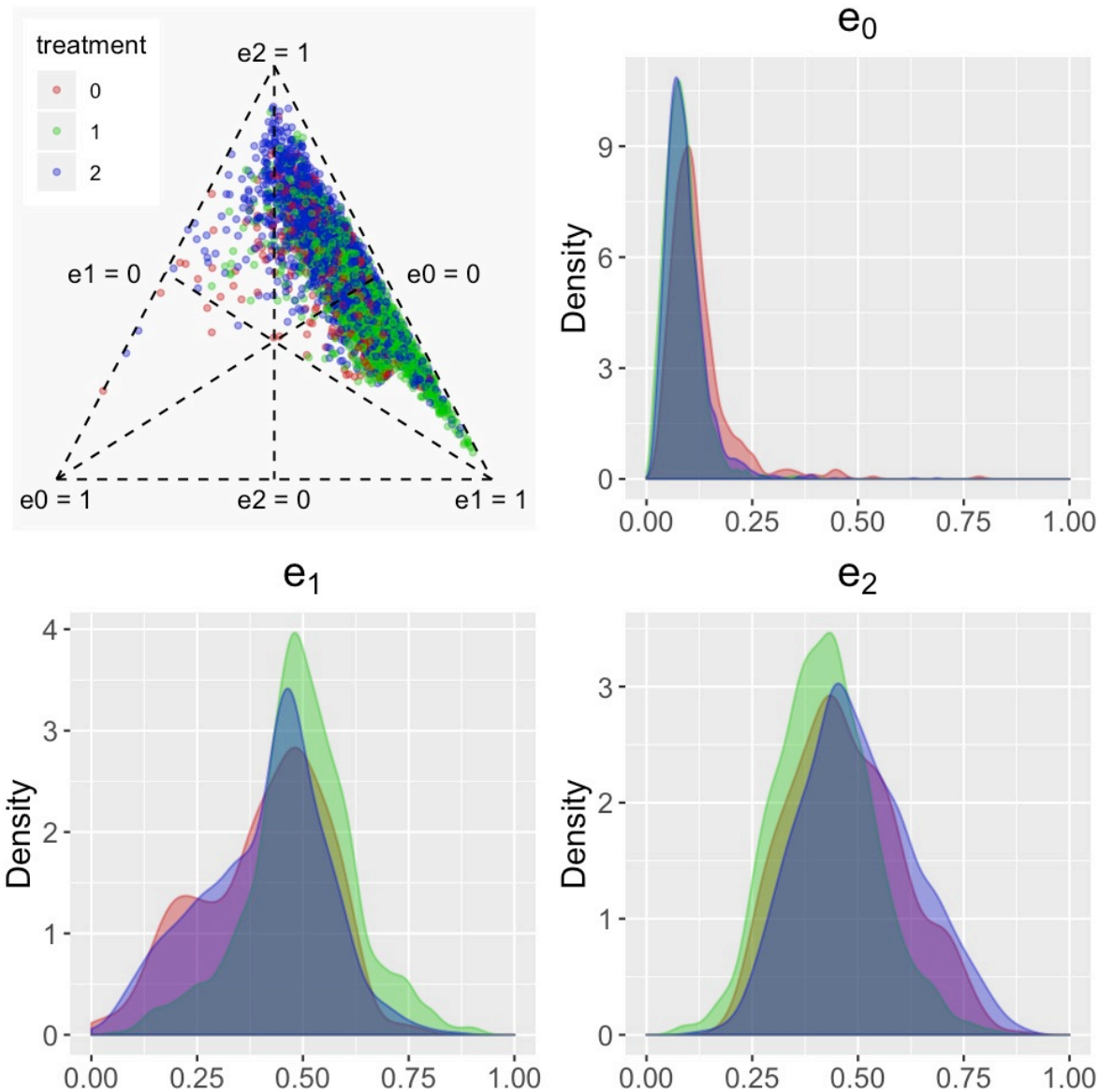


Figure 4: PS distribution for unequal (10:45:45) treatment prevalence and good PS overlap. The three PS components (e_0, e_1, e_2) lie on a two-dimensional equilateral triangle in the plane $e_0 + e_1 + e_2 = 1$ and constrained by $0 < e_0, e_1, e_2 < 1$.

3.2.5 Unequal (10:45:45) treatment prevalence, "fair" PS overlap

$$\boldsymbol{\alpha}_0 = \{0.599, -0.054, 0.316, 0.375, -0.580, -0.599, -0.261, 0.281, 0.339, -0.038, -0.676\}$$
$$\boldsymbol{\alpha}_1 = \{2.173, -0.080, 0.150, 0.362, -0.606, -0.096, 0.312, 0.102, -0.107, 0.790, -0.106\}$$
$$\boldsymbol{\alpha}_2 = \{1.047, -0.242, -0.445, 0.411, 0.720, 0.244, 0.622, -0.415, 0.293, 0.011, -0.498\}$$

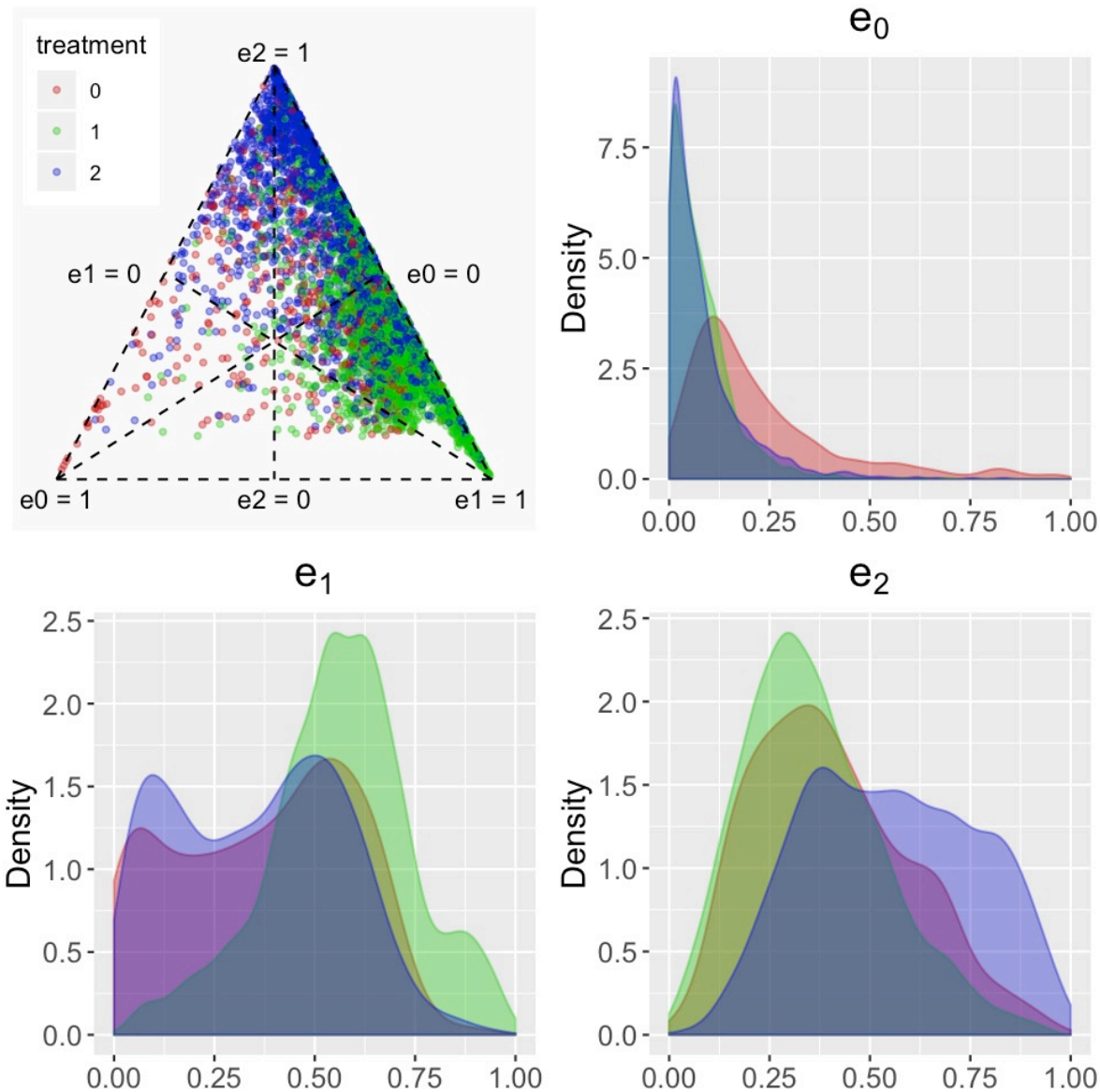


Figure 5: PS distribution for unequal (10:45:45) treatment prevalence and fair PS overlap. The three PS components (e_0, e_1, e_2) lie on a two-dimensional equilateral triangle in the plane $e_0 + e_1 + e_2 = 1$ and constrained by $0 < e_0, e_1, e_2 < 1$.

3.2.6 Unequal (10:45:45) treatment prevalence, "poor" PS overlap

$$\boldsymbol{\alpha}_0 = \{1.199, -0.107, 0.633, 0.751, -1.160, -1.198, -0.522, 0.561, 0.679, -0.077, -1.351\}$$
$$\boldsymbol{\alpha}_1 = \{3.146, -0.160, 0.301, 0.724, -1.211, -0.191, 0.623, 0.203, -0.215, 1.579, -0.212\}$$
$$\boldsymbol{\alpha}_2 = \{0.995, -0.483, -0.889, 0.822, 1.439, 0.488, 1.244, -0.829, 0.586, 0.023, -0.997\}$$

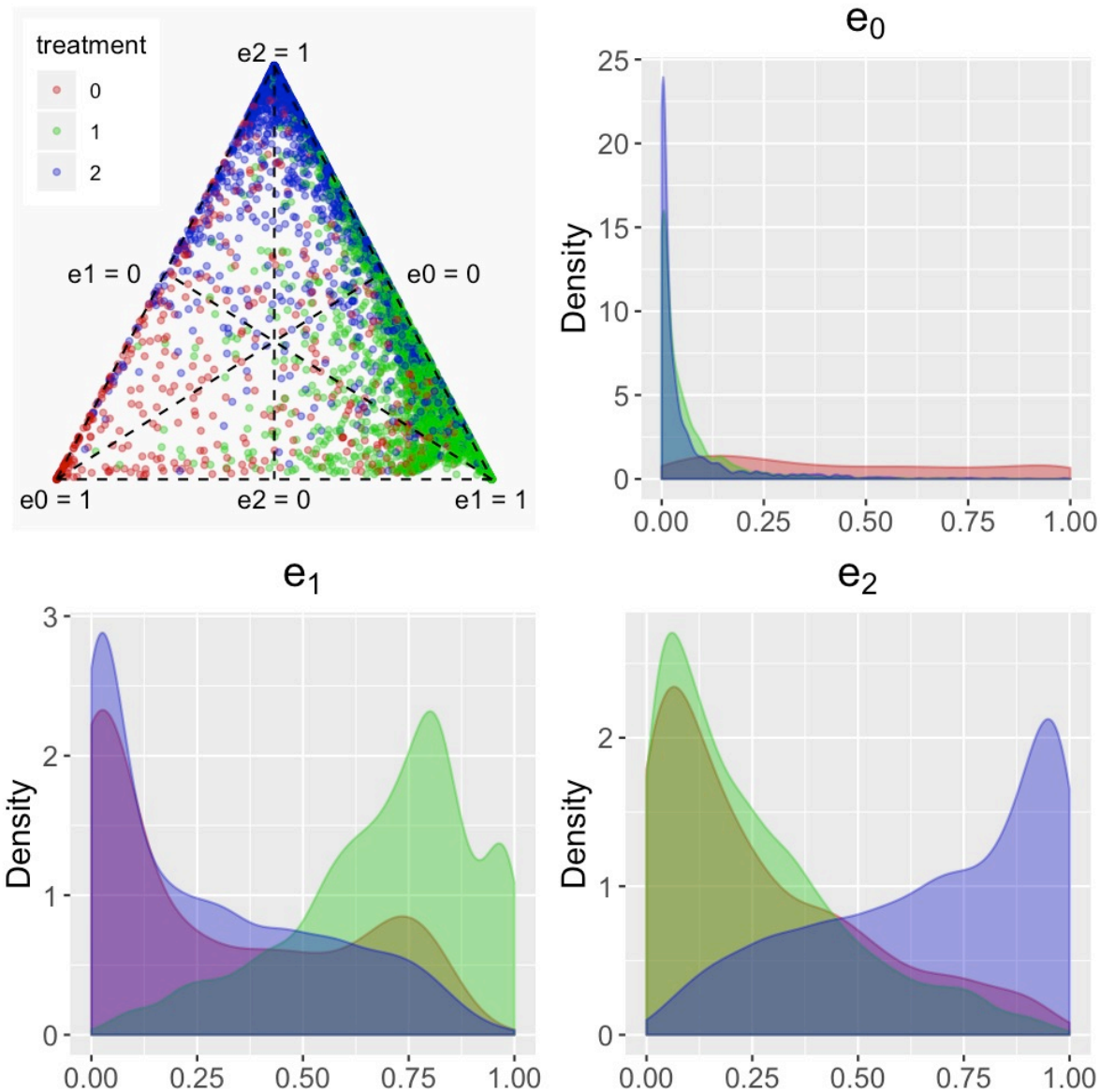


Figure 6: PS distribution for unequal (10:45:45) treatment prevalence and poor PS overlap. The three PS components (e_0, e_1, e_2) lie on a two-dimensional equilateral triangle in the plane $e_0 + e_1 + e_2 = 1$ and constrained by $0 < e_0, e_1, e_2 < 1$.

3.3 Outcome Generation

We use a binomial logistic model to generate outcomes, and use the 10 covariates and treatment as linear predictors. The outcome probability is:

$$\text{logit}(\Pr(Y_i = 1)) = \beta_0 + \boldsymbol{x}_i^T \boldsymbol{\beta}_X + \beta_1 I(t_i = 1) + \beta_2 I(t_i = 2) + \beta_{I,1} x_{i,4} I(t_i = 1) + \beta_{I,2} x_{i,4} I(t_i = 2) \tag{8}$$

where:

- β_0 is intercept term that we adjust to set outcome prevalence
- $\boldsymbol{\beta}_X$ are effects of the ten covariates
- β_1 and β_2 are effects of treatments 1 and 2 compared to treatment 0
- $\beta_{I,1}$ and $\beta_{I,2}$ are effect of interaction of covariate X_4 and treatments 1 or 2, respectively. Only used when simulating treatment effect heterogeneity

We select specific values of β based on desired simulation outcome prevalence, treatment effects, and covariate effects:

Outcome Prevalence: We simulate under two different values of β_0:

1. $\beta_0 = \log(0.02)$ corresponds roughly to a "rare" outcome prevalence of 2%
2. $\beta_0 = \log(0.10)$ corresponds roughly to a "common" outcome prevalence of 10%

Main Effects: For all simulations, we use the following covariate effects for the 10 covariates:

$$\boldsymbol{\beta}_X = (-0.15, 0.20, 0.10, -0.30, 0.30, 0.15, -0.50, -0.20, 0.40, -0.10)$$

Treatment Effects: We simulate under null and non-null true marginal relative risks, with treatment 0 as the reference:

1. Null: $\text{RR}_1 = \text{RR}_2 = 1$

2. Non-null: $\text{RR}_1 = 0.8$, $\text{RR}_2 = 0.6$

However, these are marginal relative risks, that are different from the conditional log odds ratios β_1 and β_2. When simulating, we empirically find the β_1 and β_2 that provide the desired effect sizes.

Heterogeneity: We simulate with and without heterogeneity, modeled as an interaction of treatment and covariate X_4:

1. No heterogeneity: $\beta_{I,1} = \beta_{I,2} = 0$

2. Heterogeneity: $\beta_{I,1} = \log(0.7)$, $\beta_{I,2} = \log(0.5)$

3.4 Simulation Parameters

We simulate 24 "scenarios" that represent all combinations of different simulations settings for treatment prevalence (equal, unequal), PS overlap (good, fair, poor), outcome prevalence (rare, approximately 2%; common, approximately 10%), and true treatment effect (null, non-null). We simulate 1000 times with a sample size of 5000. We estimate the two treatment effects RR_1 and RR_2 relative to reference treatment 0.

4 Propensity Score Methods

4.1 Propensity Score Estimation

Although we know the true PS for each patient, we use an estimated PS for PS adjustment. Our PS estimation model is a multinomial logistic regression fit on the ten covariates X_1 through X_{10}, with an intercept term.

4.2 Propensity Score Adjustment

We compare 10 PS adjustment methods that we list below. All outcome models include intercept estimation. We use s generally to represent spline functions as described in each method. We use the logit of the PS, e^* for the PS splines. For the purpose of simplicity in

these model equations, we use $Z_1 = I(T_i = 1)$ and $Z_2 = I(T_i = 2)$ as indicator variables for treatments 1 and 2, respectively.

1. IPTW: regress outcome on treatment indicators using stabilized IPTW weights (Xu et al., 2010)

$$w_i = p_{e_i}/e_{i,t_i}$$

   where p_k, $k = 0, 1, 2$ are the marginal treatment probabilities.

2. IPTW Truncated: truncate stabilized weights that lie outside the range $[0.1, 10]$ to 0.1 or 10, as appropriate.

3. IPTW Truncated %: truncate stabilized weights only on high end that exceed 99th percentile of weights among same treatment groups.

4. Outcome model: regress outcome on covariates directly without using the PS. Continuous covariates are modeled with natural cubic splines with four interior knots placed at evenly spaced quantiles. The splines are not smoothed.

$$Y \sim Z_1 + Z_2 + s(X_1) + s(X_2) + s(X_3) + X_4 + X_5 + X_6 + s(X_7) + s(X_8) + s(X_9) + X_{10}$$

5. Cubic 1: regress outcome on treatments and natural cubic splines of PS that have one internal knot located at the median. The splines are not smoothed.

$$Y \sim Z_1 + Z_2 + s(e_1^*) + s(e_2^*)$$

6. Cubic 4: regress outcome on treatments and natural cubic splines of PS that have four internal knots located at evenly spaced quantiles. The splines are not smoothed.

$$Y \sim Z_1 + Z_2 + s(e_1^*) + s(e_2^*)$$

7. TPRS 1D: regress outcome on treatments and separate thin plate regression splines for each PS component. We use the default parameters in the "mgcv" R package that uses 10 degrees of freedom for a 1-dimensional TPRS.

$$Y \sim Z_1 + Z_2 + s(e_1^*) + s(e_2^*)$$

8. TPRS 2D: regress outcome of treatments and joint thin plate (tp) regression spline on both PS components. We use the default parameters in the "mgcv" R package that uses 27 degrees of freedom for a 2-dimensional TPRS.

$$Y \sim Z_1 + Z_2 + s(e_1^*, e_2^*)$$

9. Cubic 4 + interaction: similar to Cubic 4, with interactions of treatment and PS splines.

$$Y \sim Z_1 + Z_2 + s(e_1^*) + s(e_2^*) + Z_1 * s(e_1^*) + Z_1 * s(e_2^*) + Z_2 * s(e_1^*) + Z_2 * s(e_2^*)$$

10. TPRS 2D + interaction: similar to TPRS 2D, with interactions of treatment and PS spline:

$$Y \sim Z_1 + Z_2 + s(e_1^*, e_2^*) + Z_1 * s(e_1^*, e_2^*) + Z_2 * s(e_1^*, e_2^*)$$

For methods using cubic splines, we use the estimated frequentist covariance matrix of model parameters in our variance calculation for the marginal relative risk. For methods using IPTW, we use a sandwich variance estimator that is the default approach in the "sandwich" R package. For methods using thin plate regression splines, we use the Bayesian covariance matrix calculated by the default "gam{mgcv}" function for our variance calculation, that has been shown to provide more accurate confidence interval estimates than the frequentist covariance matrix in the presence of smoothing (Marra and Wood, 2012).

For each scenario, we fit methods under four "settings":

1. Heterogeneity: added treatment-covariate X_4 interaction in outcome generating model
2. Trimming: based on PS, and applied to study population before PS adjustment
3. PS misspecification: intentional removal of covariate X_9 from PS estimation process
4. Standard: no heterogeneity, trimming, or PS misspecification

PS misspecification is accomplished by intentionally removing covariate X_9 from the PS estimation model. We perform percentile based trimming (Stürmer et al., 2014) using a 1% threshold. That is, for each of the three treatments $k = 0, 1, 2$, we find the 1st percentile of $\{e_{i,k} : t_i = k\}$, then all patients of any treatment with $e_{i,k}$ smaller than this threshold are trimmed.

5 IPTW Diagnostics

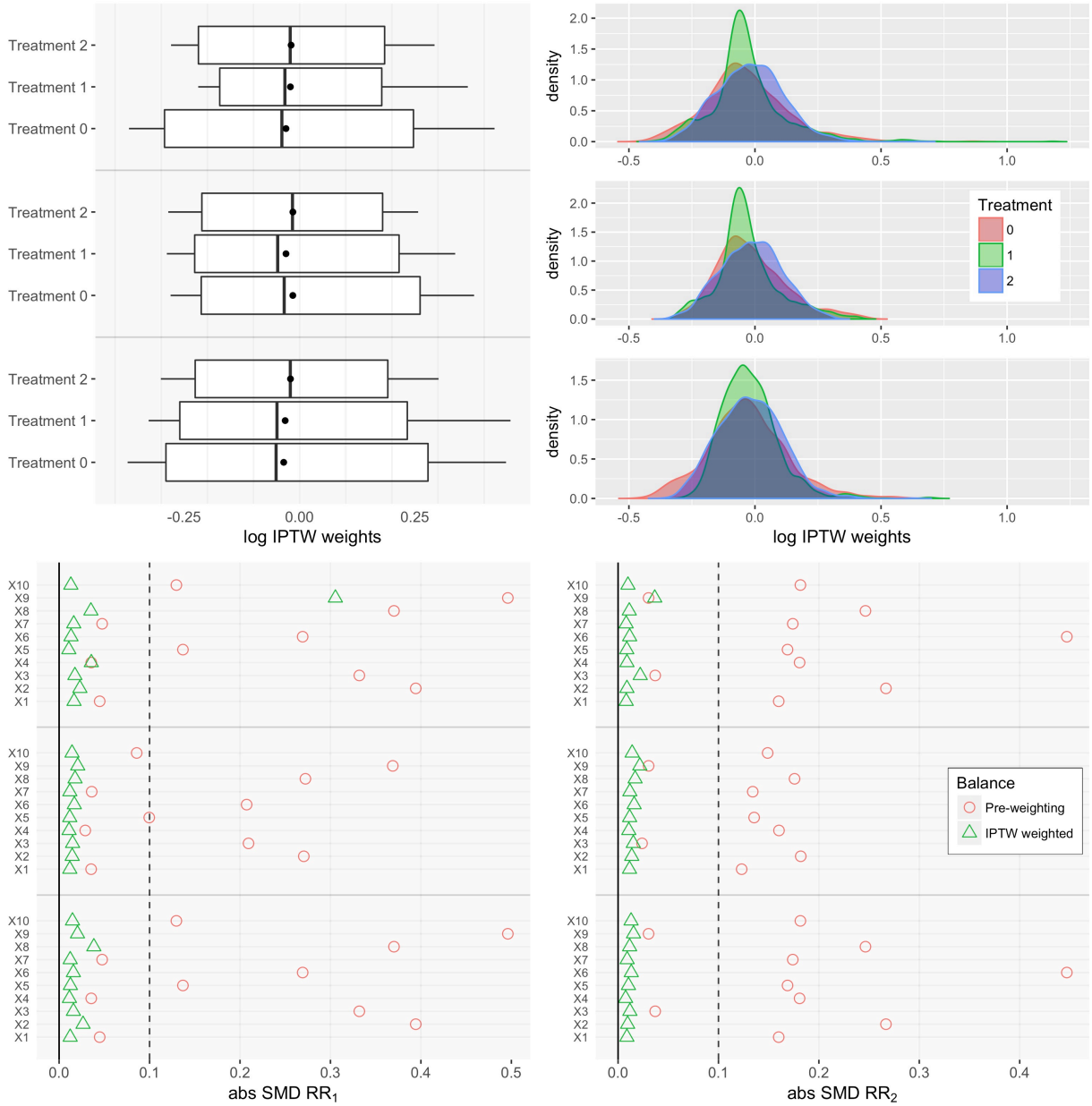


Figure 7: IPTW diagnostics for simulations under equal treatment prevalence and good PS overlap. Top left graph: Log base 10 percentiles of IPTW weights. Tips are 1st and 99th percentile, box spans 5th to 95th percentile, middle line is median, dot is mean. Top right graph: density of log IPTW weights. Bottom left graph: Absolute standardized mean differences between treatment groups 1 and 0 before and after IPTW weighting.. Bottom right graph: Absolute standardized mean differences between treatment groups 2 and 0 before and after IPTW weighting. Top row of each graph: misspecified PS setting. Middle row of each graph: trimmed setting. Bottom row of each graph: standard and treatment heterogeneity setting.

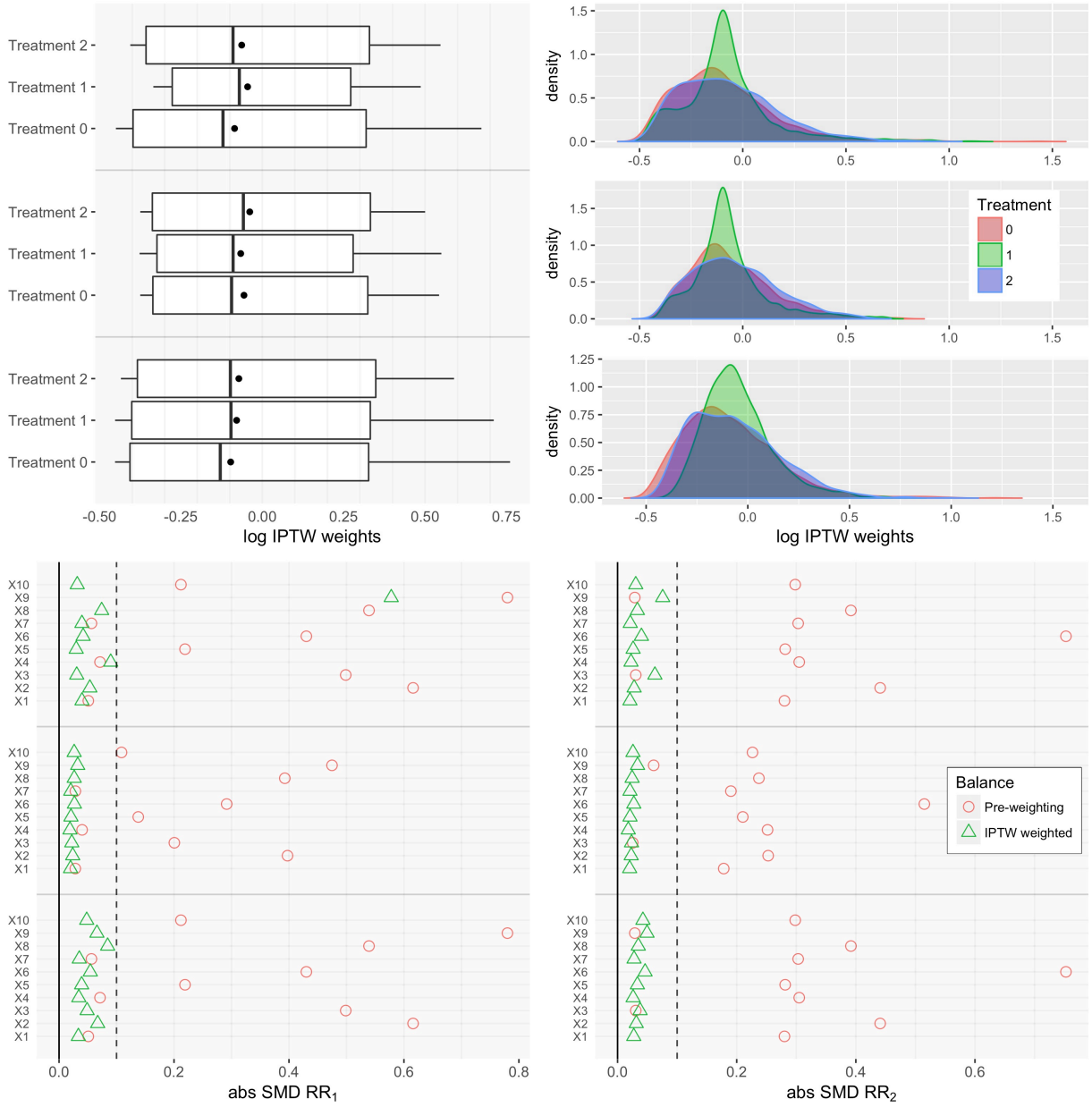


Figure 8: IPTW diagnostics for simulations under equal treatment prevalence and fair PS overlap. Top left graph: Log base 10 percentiles of IPTW weights. Tips are 1st and 99th percentile, box spans 5th to 95th percentile, middle line is median, dot is mean. Top right graph: density of log IPTW weights. Bottom left graph: Absolute standardized mean differences between treatment groups 1 and 0 before and after IPTW weighting.. Bottom right graph: Absolute standardized mean differences between treatment groups 2 and 0 before and after IPTW weighting. Top row of each graph: misspecified PS setting. Middle row of each graph: trimmed setting. Bottom row of each graph: standard and treatment heterogeneity setting.

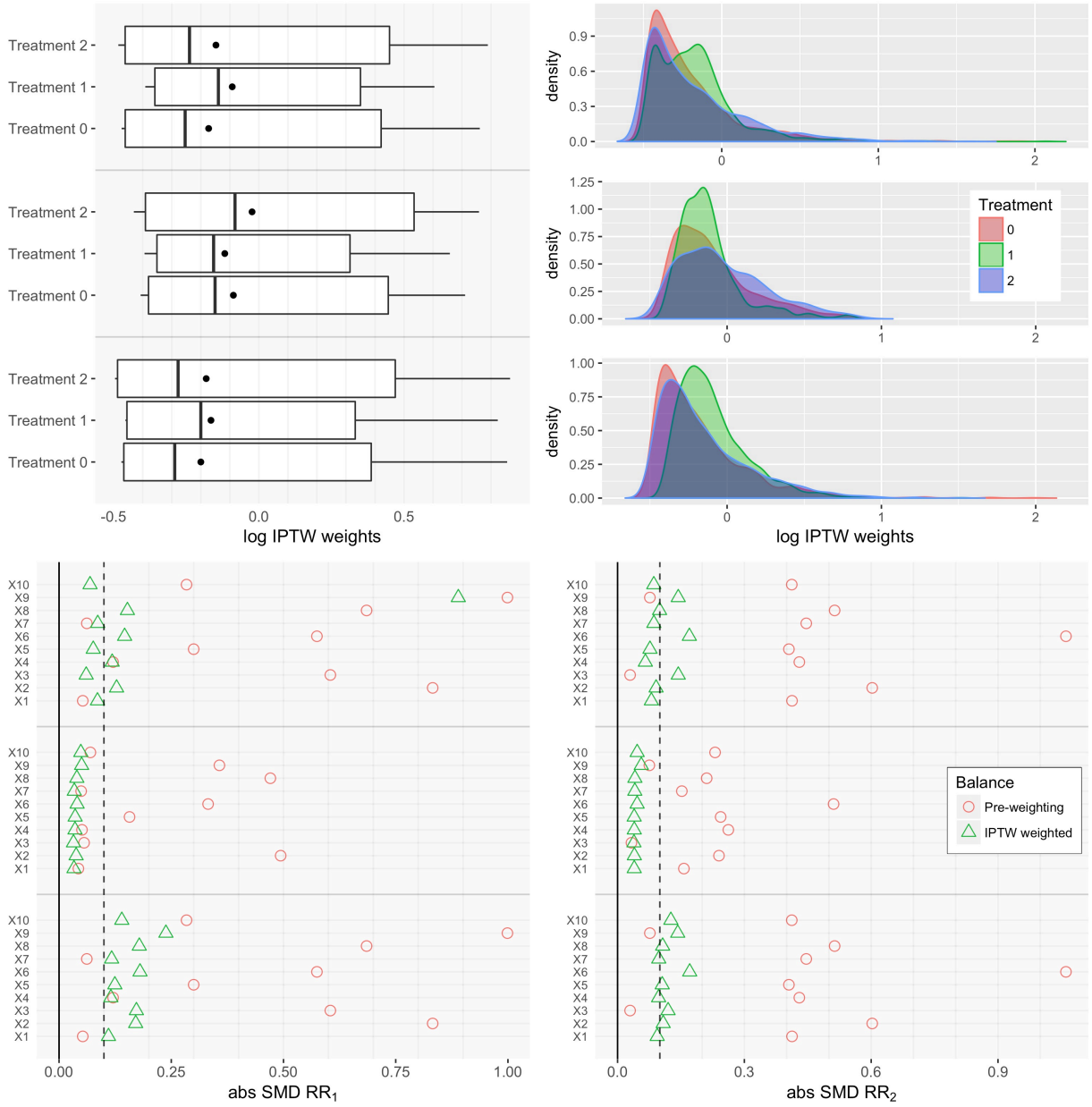


Figure 9: IPTW diagnostics for simulations under equal treatment prevalence and poor PS overlap. Top left graph: Log base 10 percentiles of IPTW weights. Tips are 1st and 99th percentile, box spans 5th to 95th percentile, middle line is median, dot is mean. Top right graph: density of log IPTW weights. Bottom left graph: Absolute standardized mean differences between treatment groups 1 and 0 before and after IPTW weighting.. Bottom right graph: Absolute standardized mean differences between treatment groups 2 and 0 before and after IPTW weighting. Top row of each graph: misspecified PS setting. Middle row of each graph: trimmed setting. Bottom row of each graph: standard and treatment heterogeneity setting.

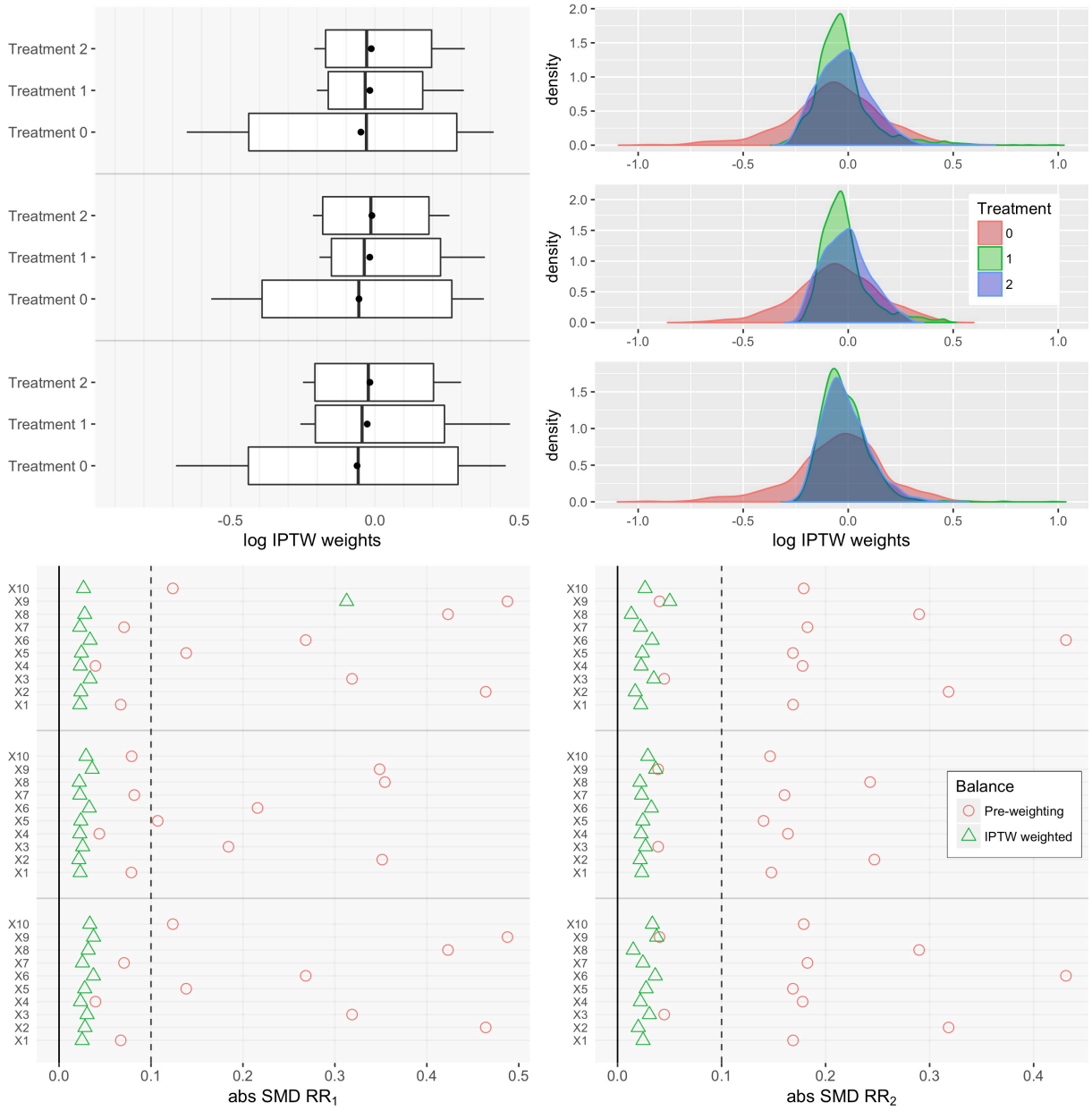


Figure 10: IPTW diagnostics for simulations under unequal treatment prevalence and good PS overlap. Top left graph: Log base 10 percentiles of IPTW weights. Tips are 1st and 99th percentile, box spans 5th to 95th percentile, middle line is median, dot is mean. Top right graph: density of log IPTW weights. Bottom left graph: Absolute standardized mean differences between treatment groups 1 and 0 before and after IPTW weighting.. Bottom right graph: Absolute standardized mean differences between treatment groups 2 and 0 before and after IPTW weighting. Top row of each graph: misspecified PS setting. Middle row of each graph: trimmed setting. Bottom row of each graph: standard and treatment heterogeneity setting.

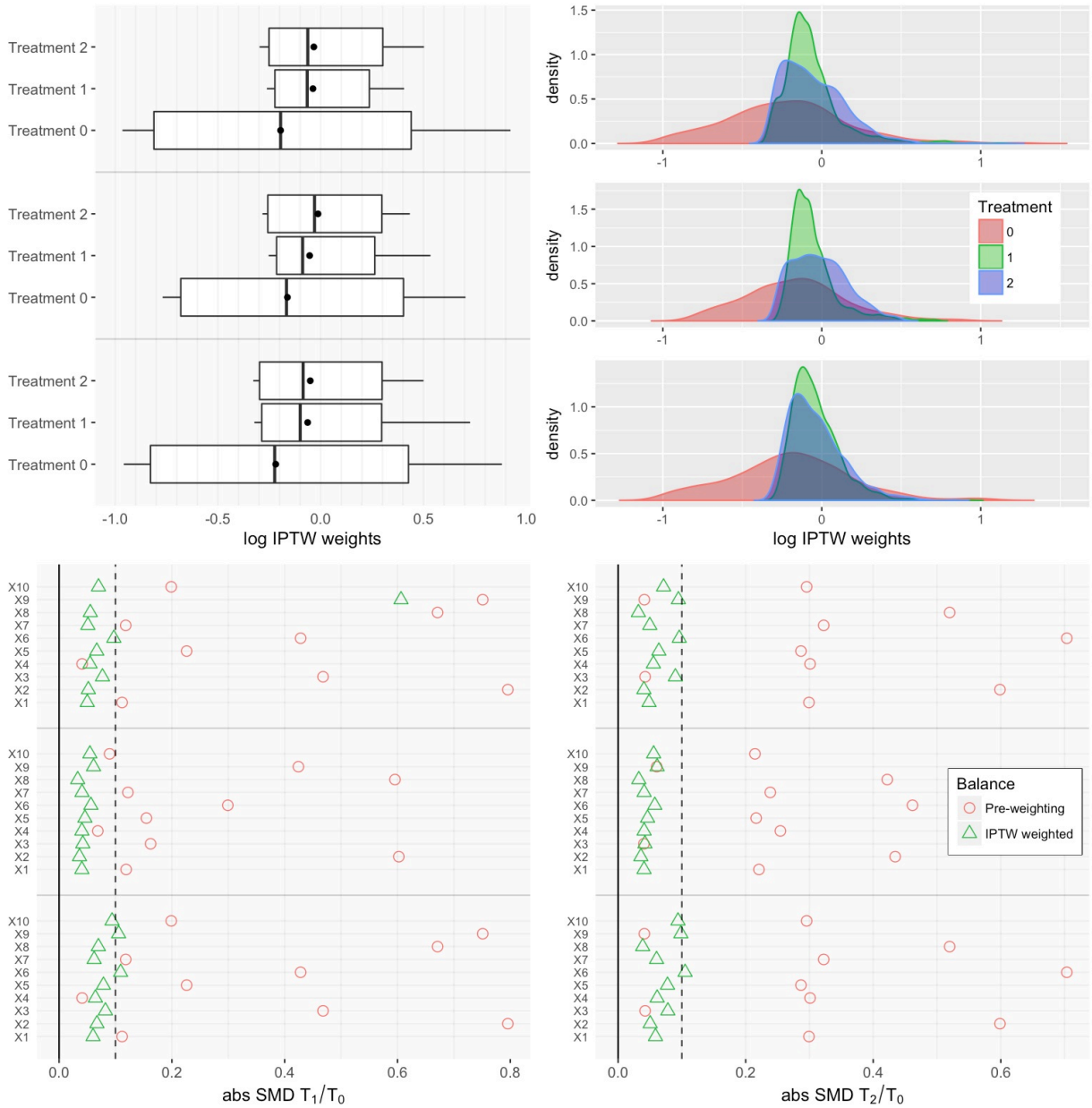


Figure 11: IPTW diagnostics for simulations under unequal treatment prevalence and fair PS overlap. Top left graph: Log base 10 percentiles of IPTW weights. Tips are 1st and 99th percentile, box spans 5th to 95th percentile, middle line is median, dot is mean. Top right graph: density of log IPTW weights. Bottom left graph: Absolute standardized mean differences between treatment groups 1 and 0 before and after IPTW weighting.. Bottom right graph: Absolute standardized mean differences between treatment groups 2 and 0 before and after IPTW weighting. Top row of each graph: misspecified PS setting. Middle row of each graph: trimmed setting. Bottom row of each graph: standard and treatment heterogeneity setting.

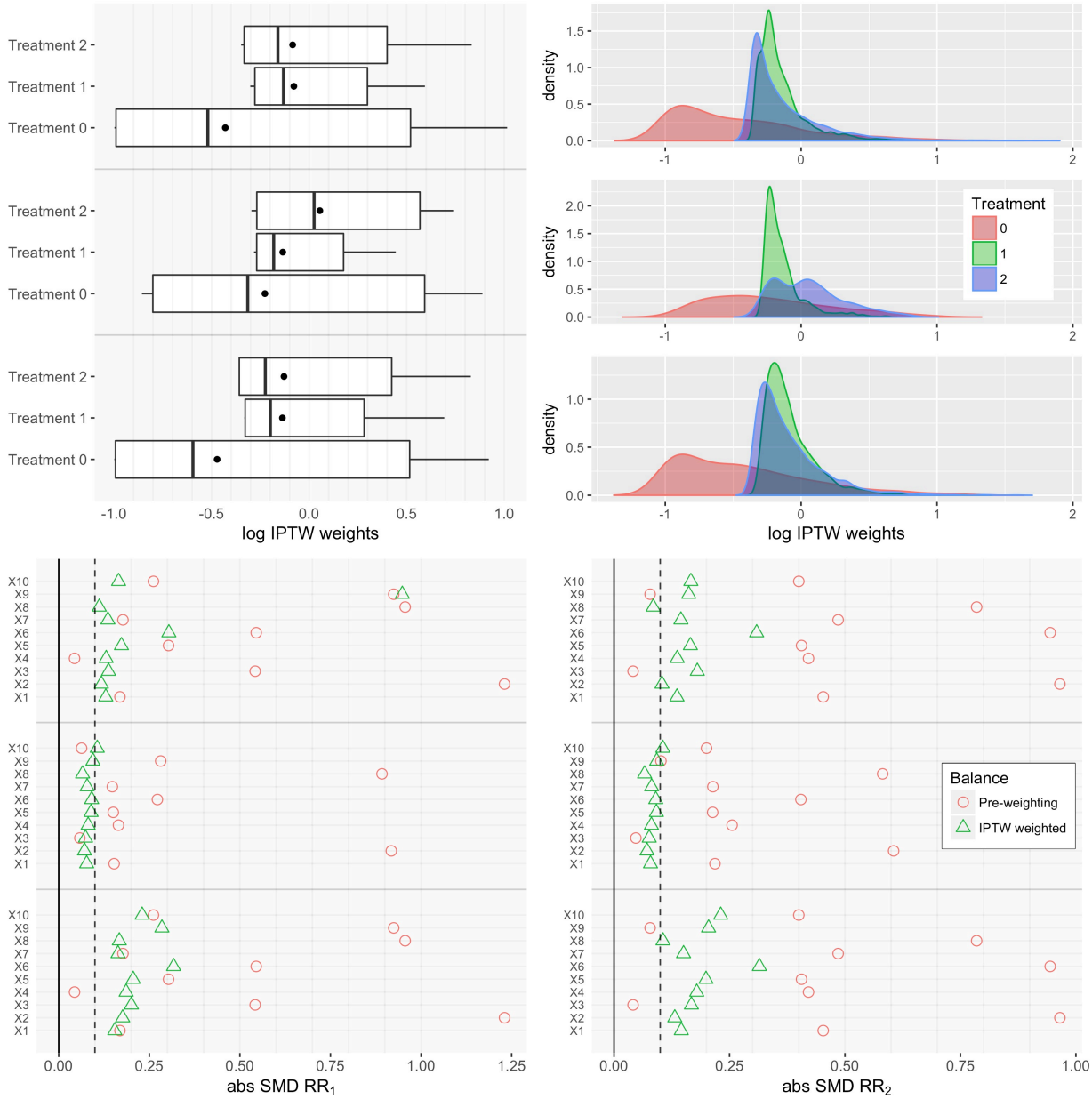


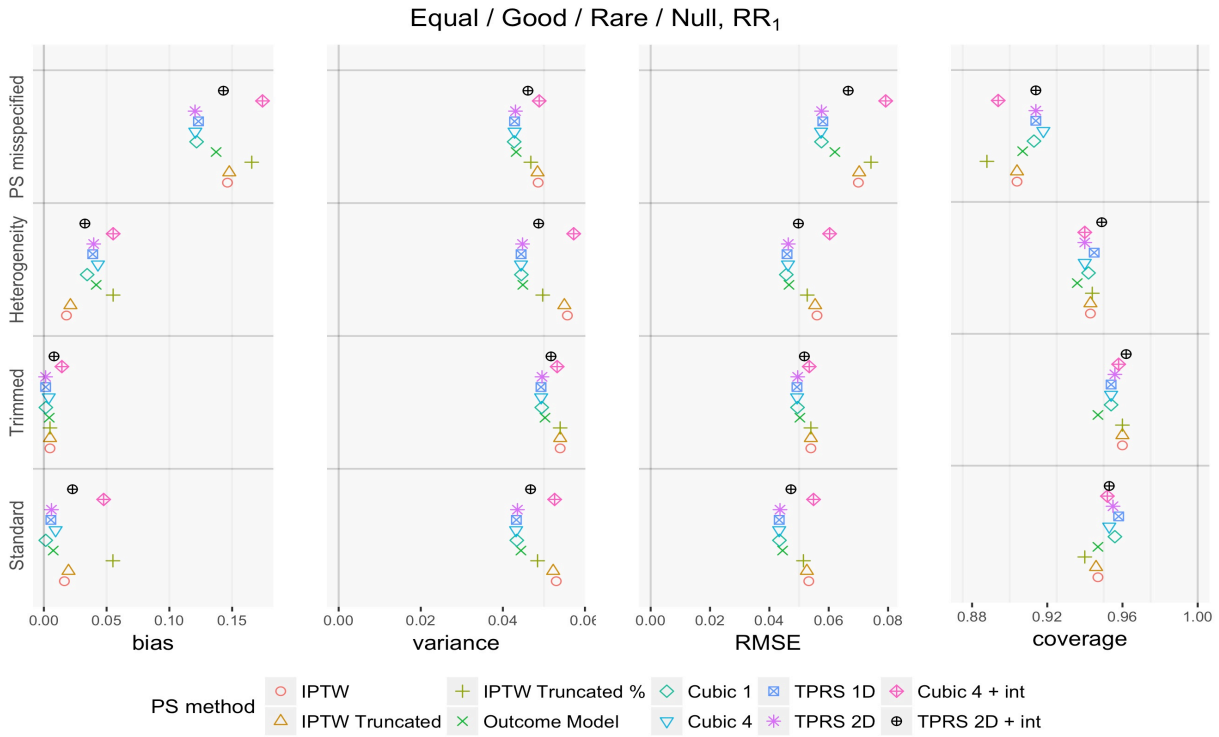
Figure 12: IPTW diagnostics for simulations under unequal treatment prevalence and poor PS overlap. Top left graph: Log base 10 percentiles of IPTW weights. Tips are 1st and 99th percentile, box spans 5th to 95th percentile, middle line is median, dot is mean. Top right graph: density of log IPTW weights. Bottom left graph: Absolute standardized mean differences between treatment groups 1 and 0 before and after IPTW weighting.. Bottom right graph: Absolute standardized mean differences between treatment groups 2 and 0 before and after IPTW weighting. Top row of each graph: misspecified PS setting. Middle row of each graph: trimmed setting. Bottom row of each graph: standard and treatment heterogeneity setting.

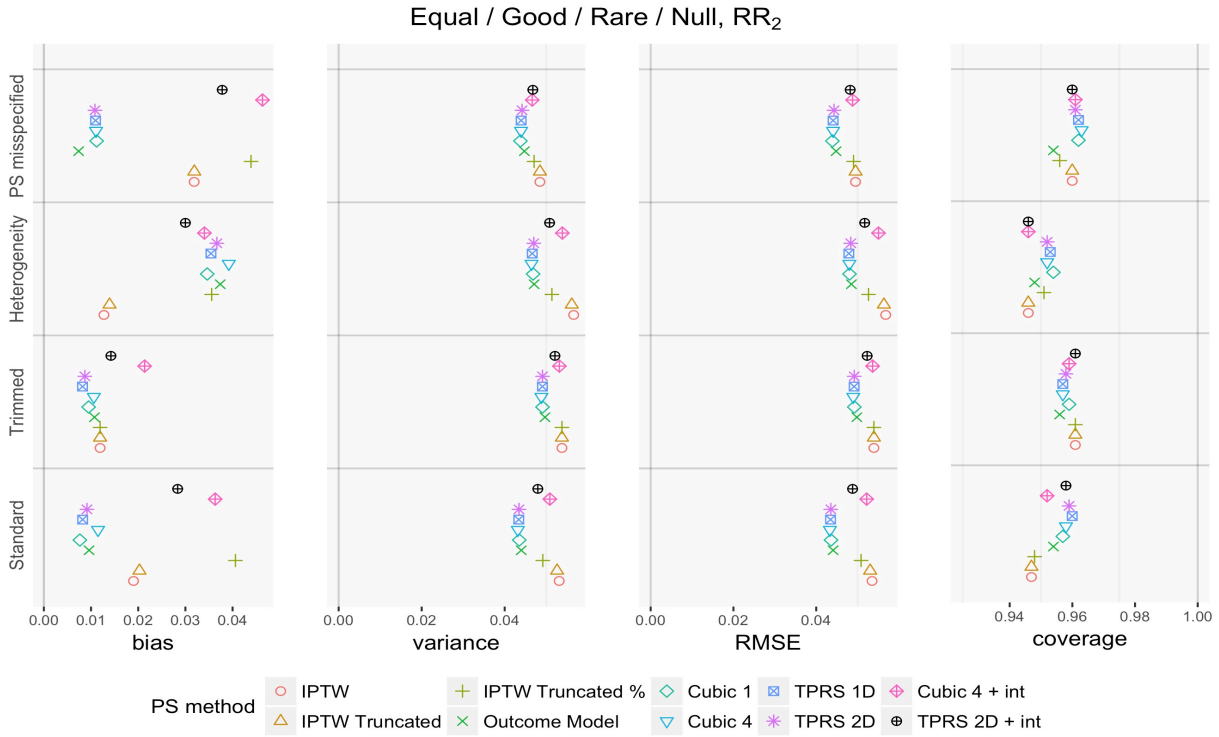
6 Simulation Results

In the following simulation results, the scenarios are labeled as (ex: "Equal / Good / Common / Null, RR_1"), where

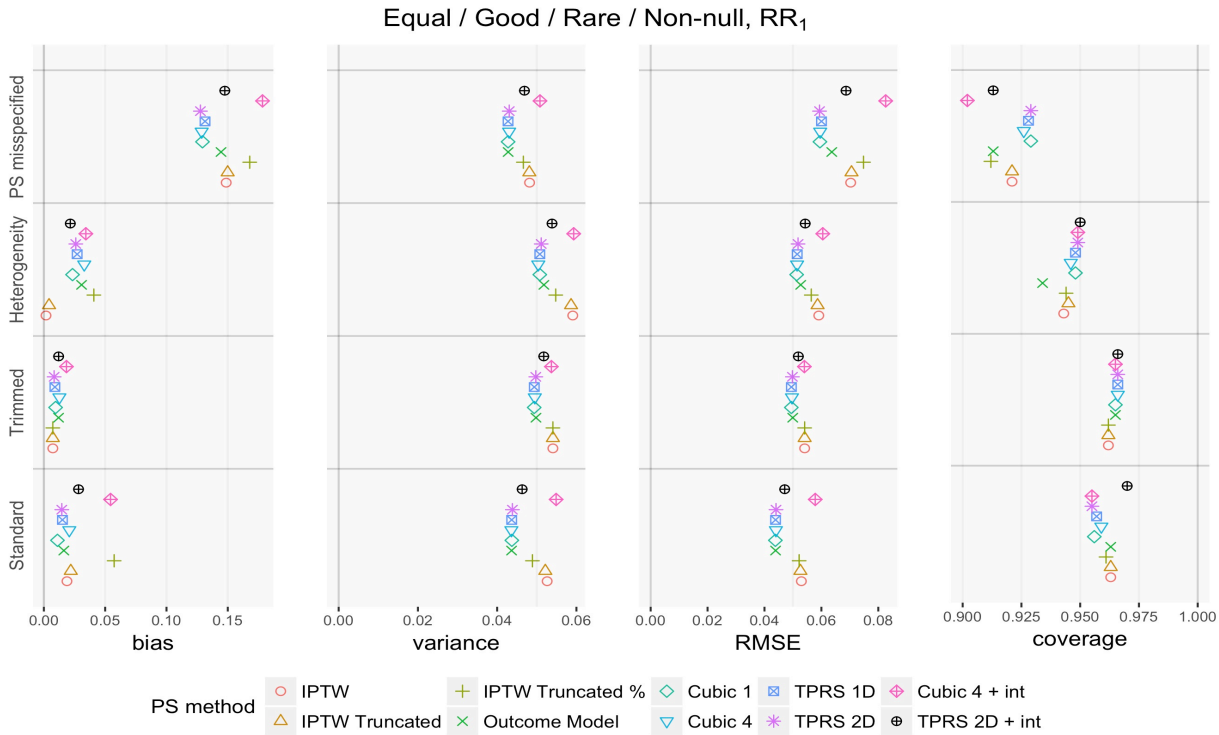
- The first parameter refers to treatment prevalence (Equal, Nonequal)
- The second parameter refers to PS overlap (Good, Fair, Poor)
- The third parameter refers to outcome prevalence (Rare, Common)
- The fourth parameter refers to true treatment effect (Null, Non-null)
- The last parameter refers to which of two marginal relative risks ($RR_1 = 0.8, RR_2 = 0.6$)

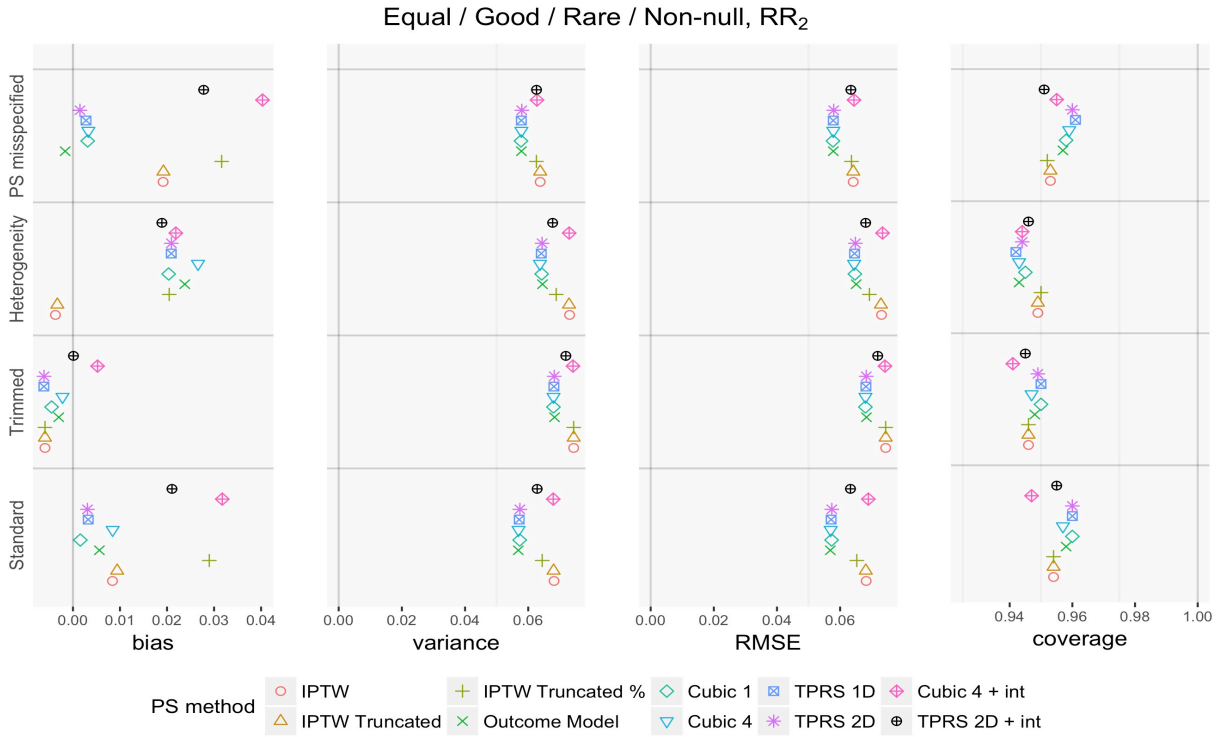
6.1 Equal / Good / Rare / Null



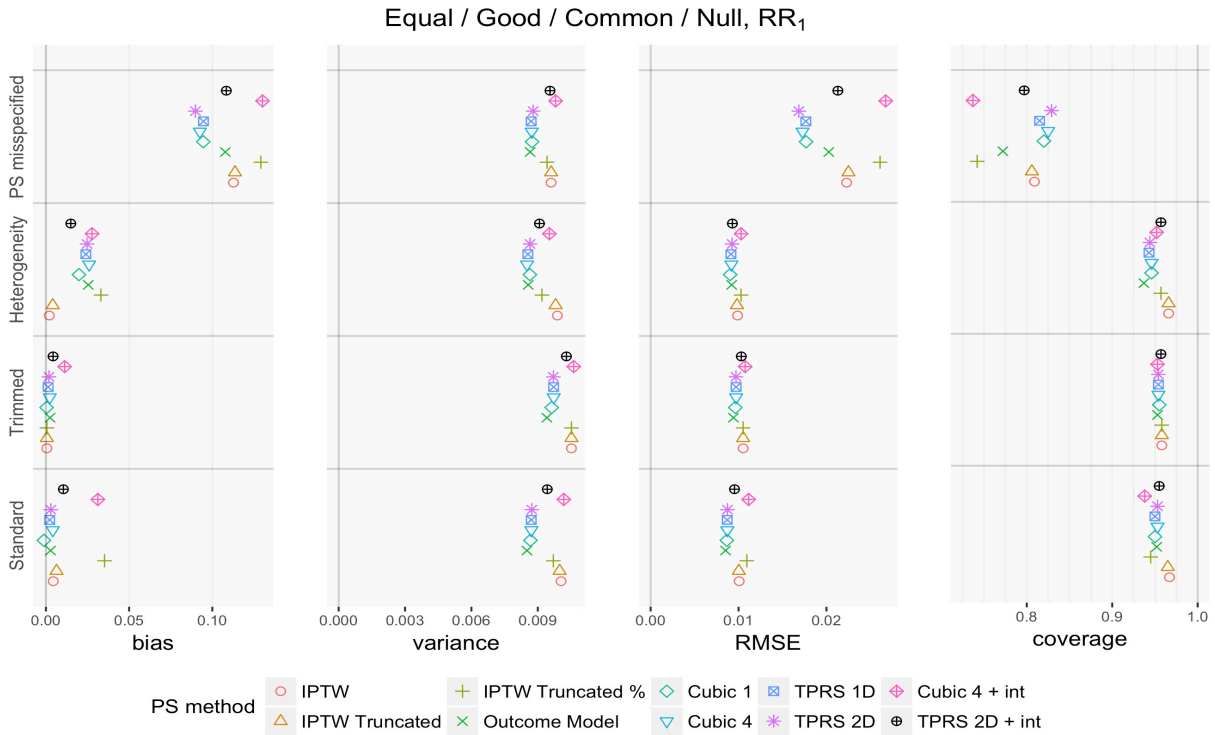


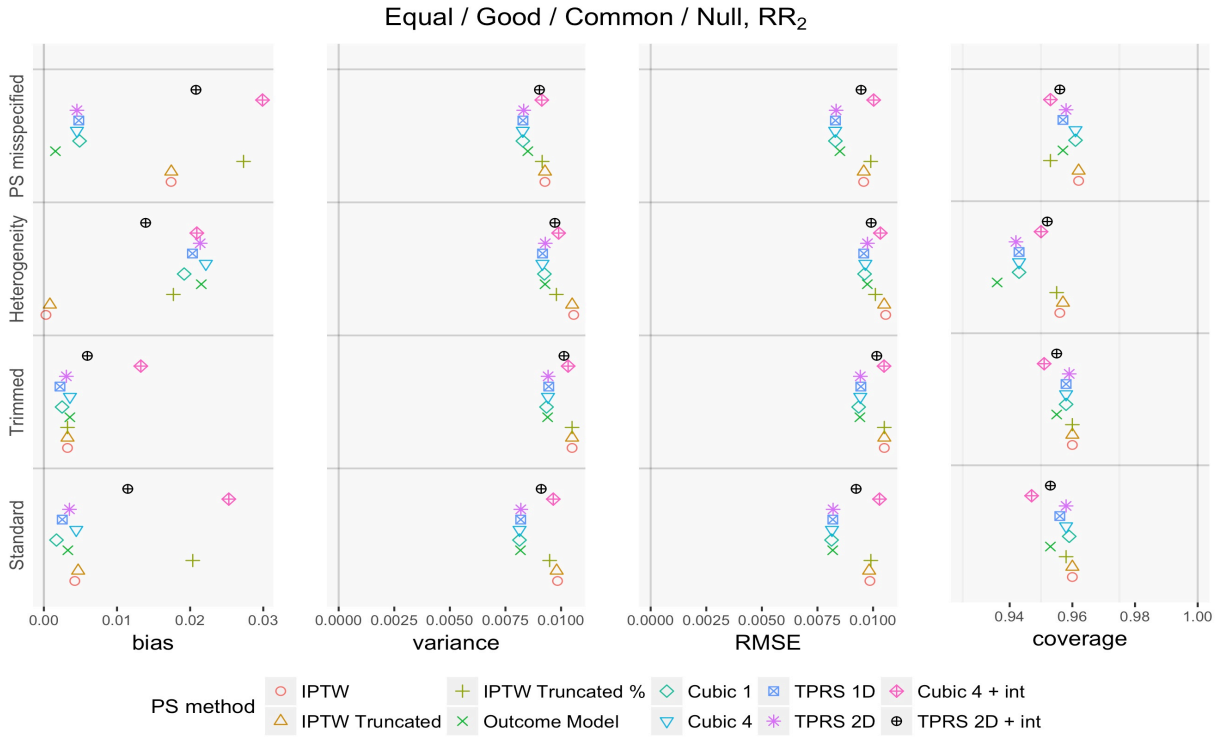
6.2 Equal / Good / Rare / Non-null



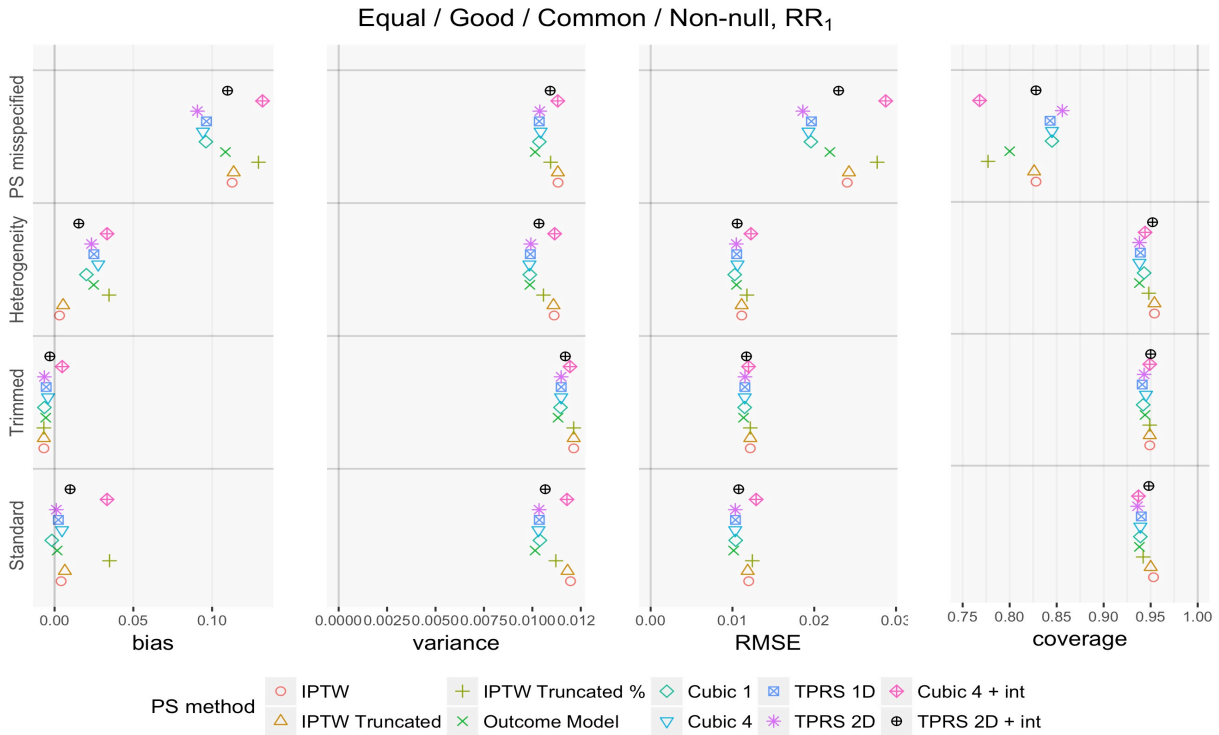


6.3 Equal / Good / Common / Null

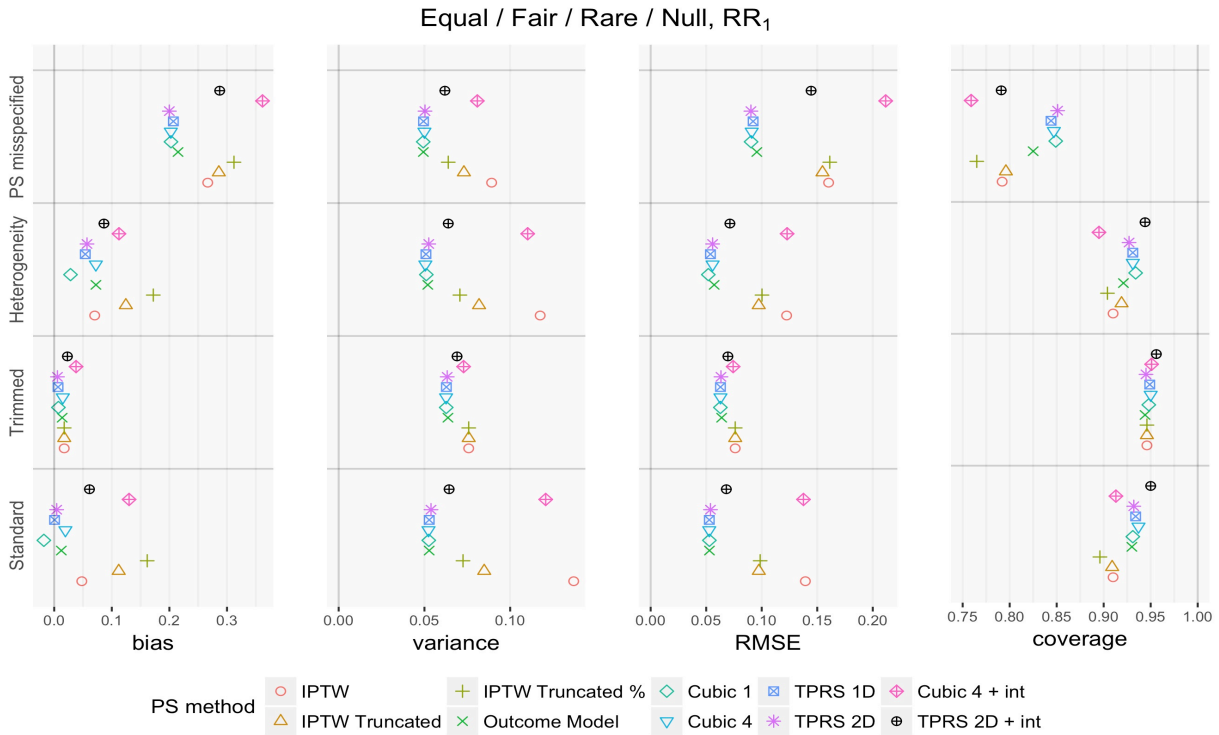


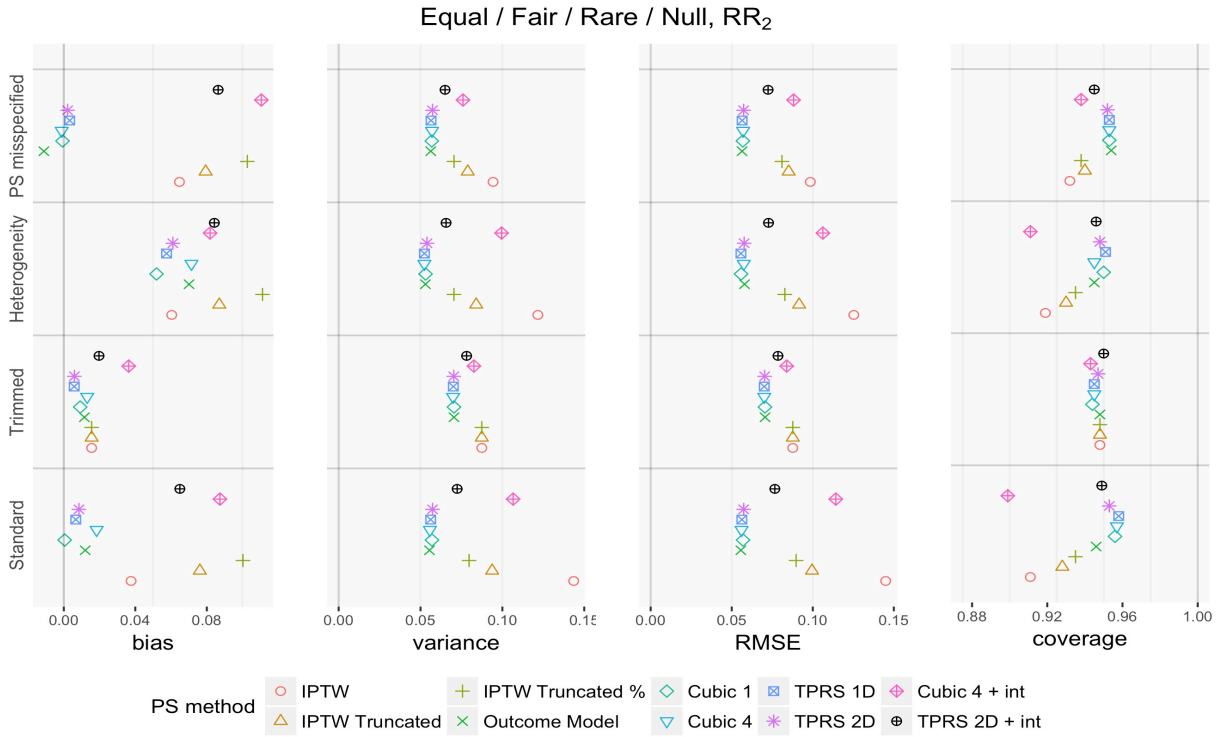


6.4 Equal / Good / Common / Non-null

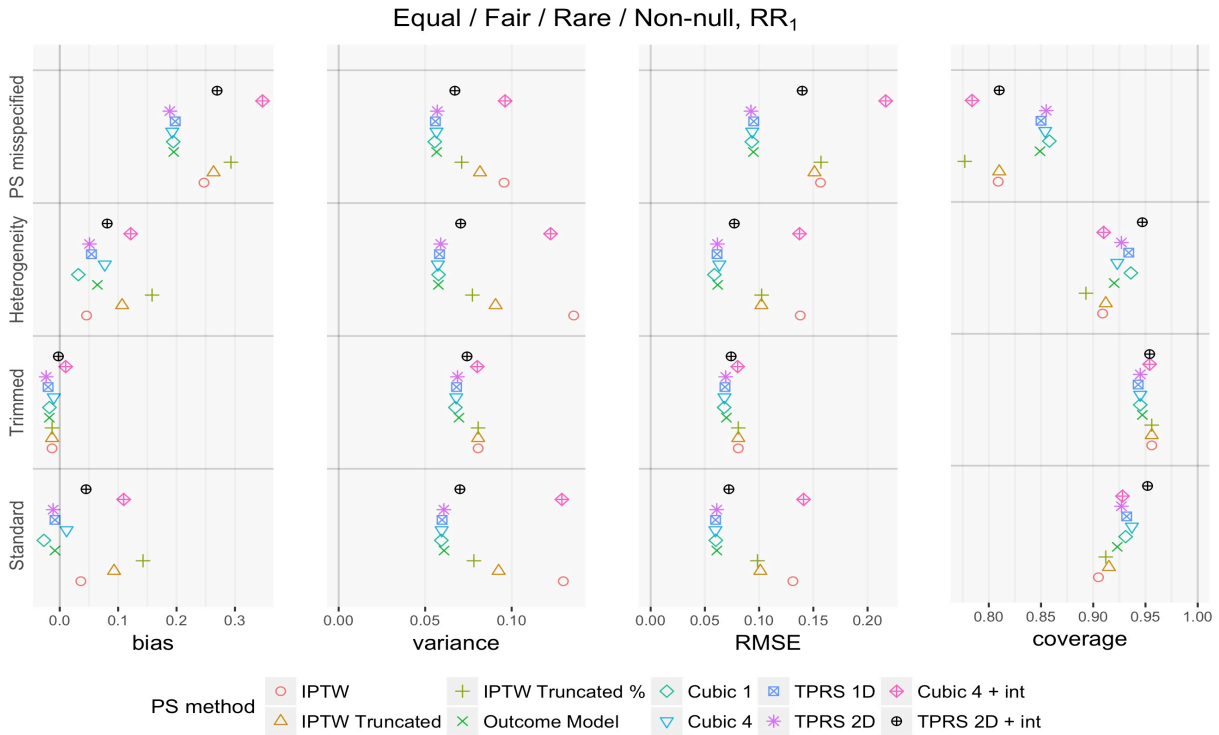


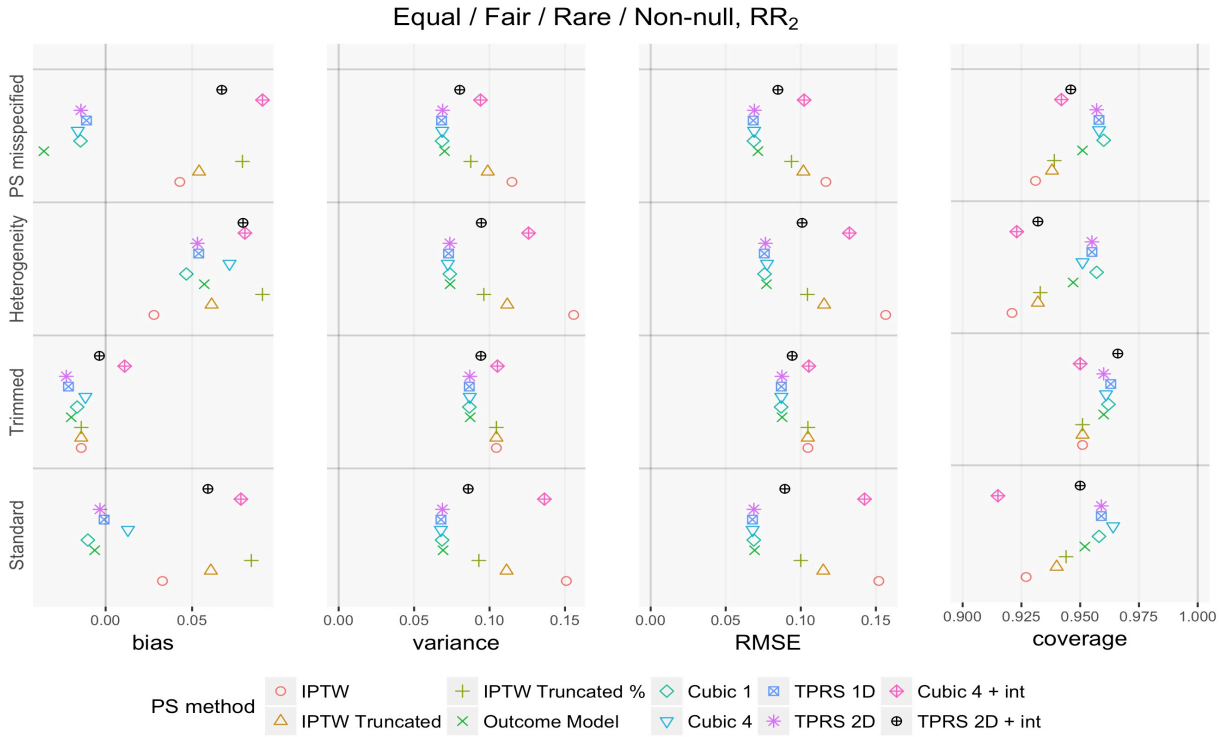
6.5 Equal / Fair / Rare / Null



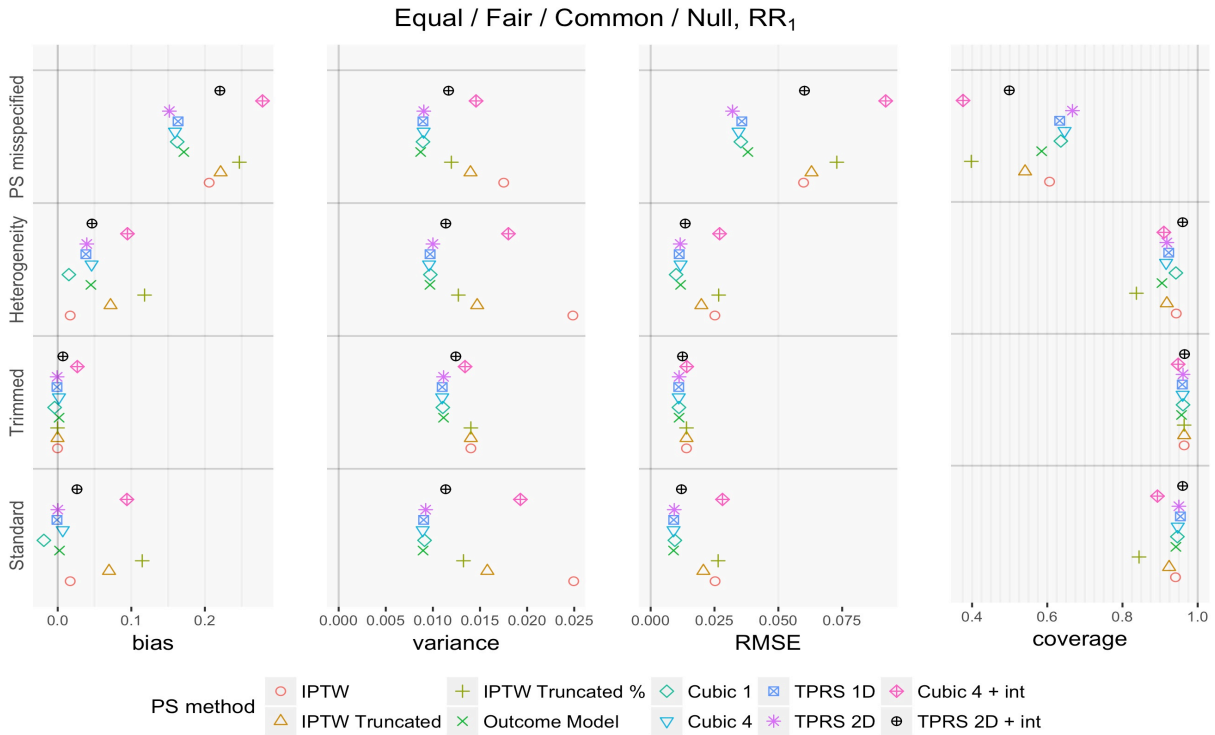


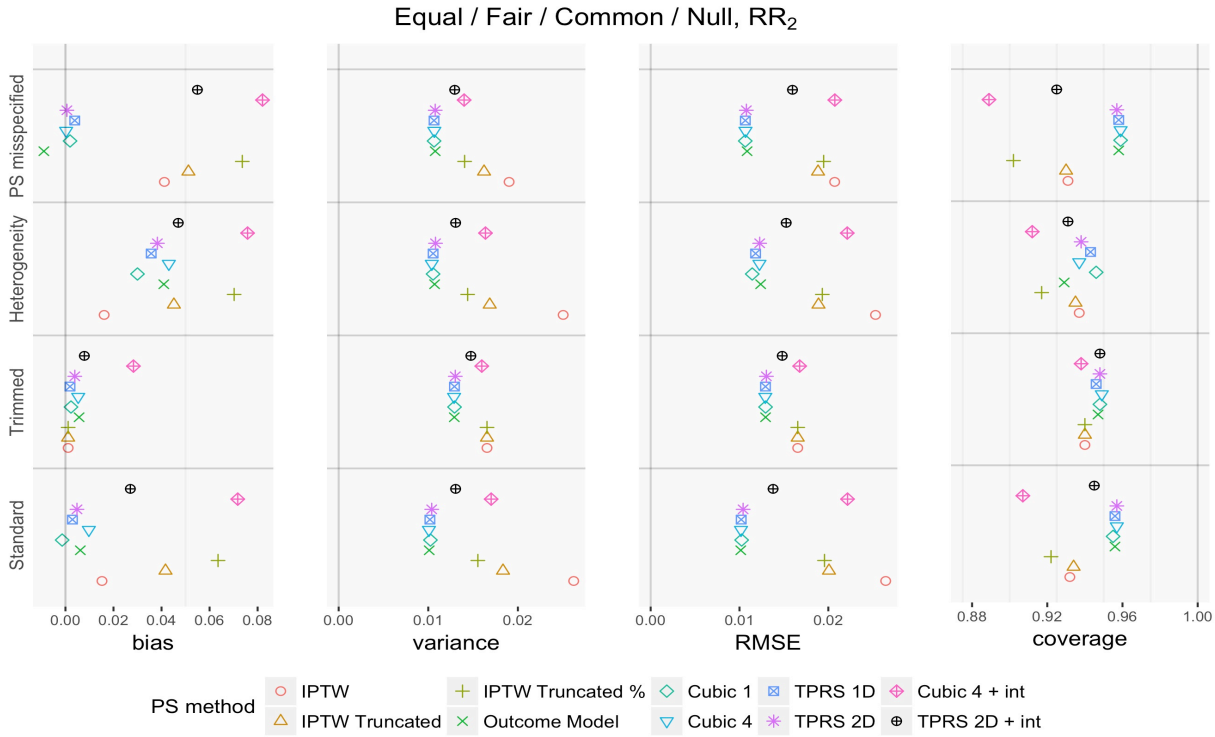
6.6 Equal / Fair / Rare / Non-null



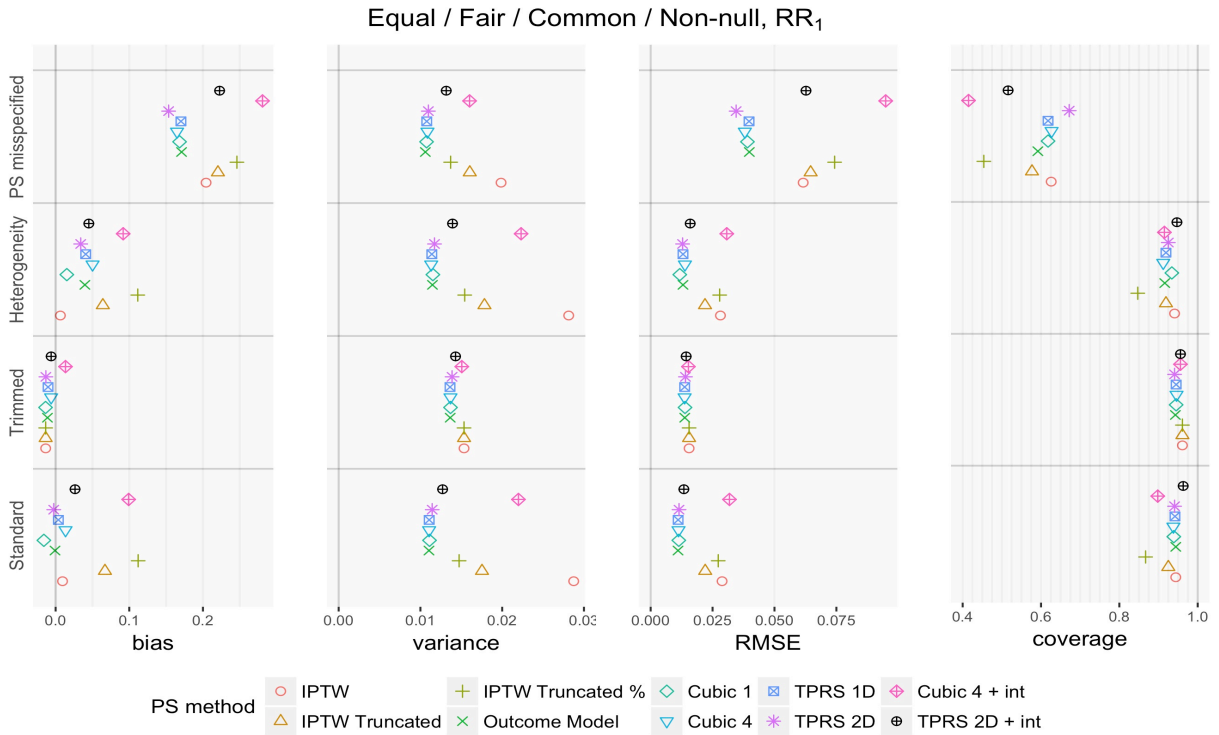


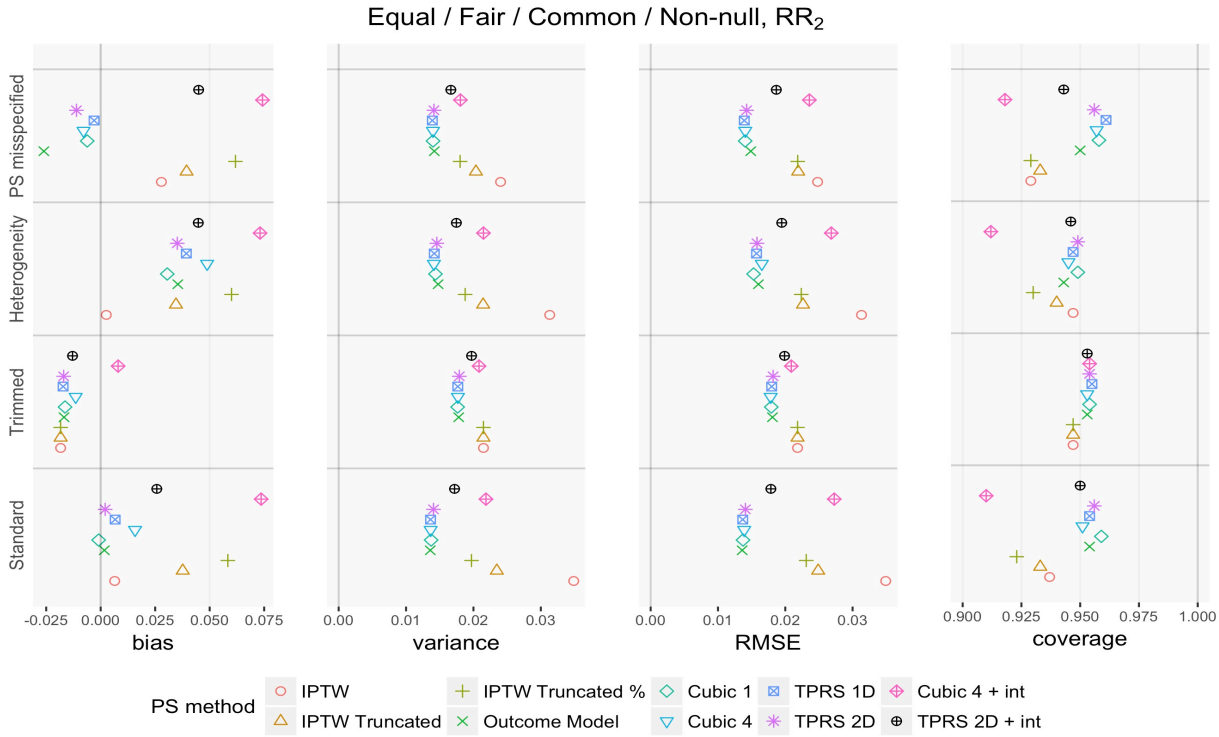
6.7 Equal / Fair / Common / Null



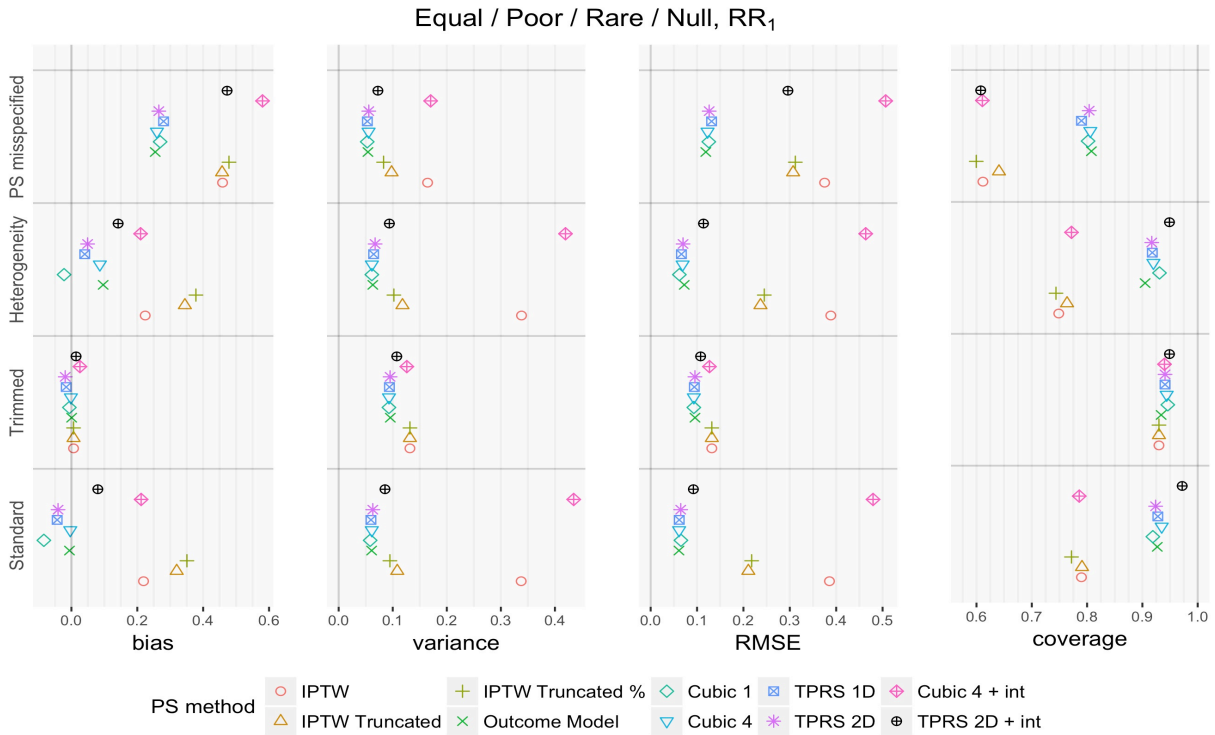


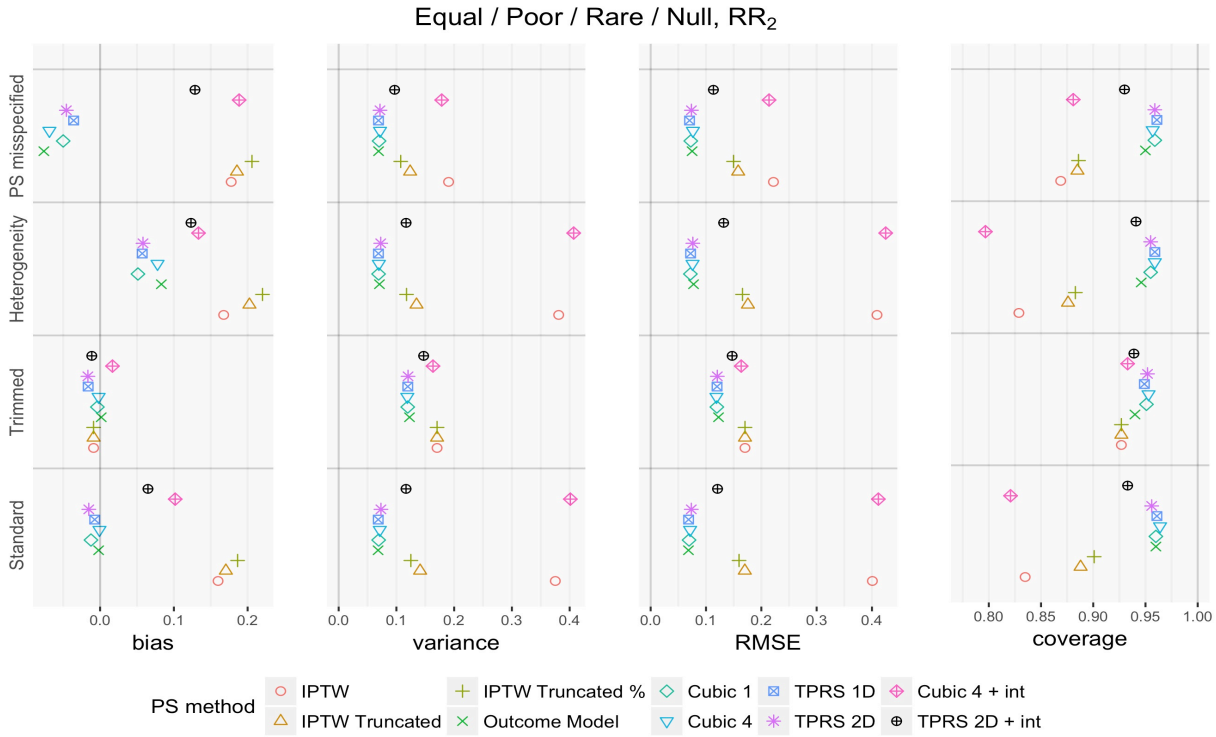
6.8 Equal / Fair / Common / Non-null



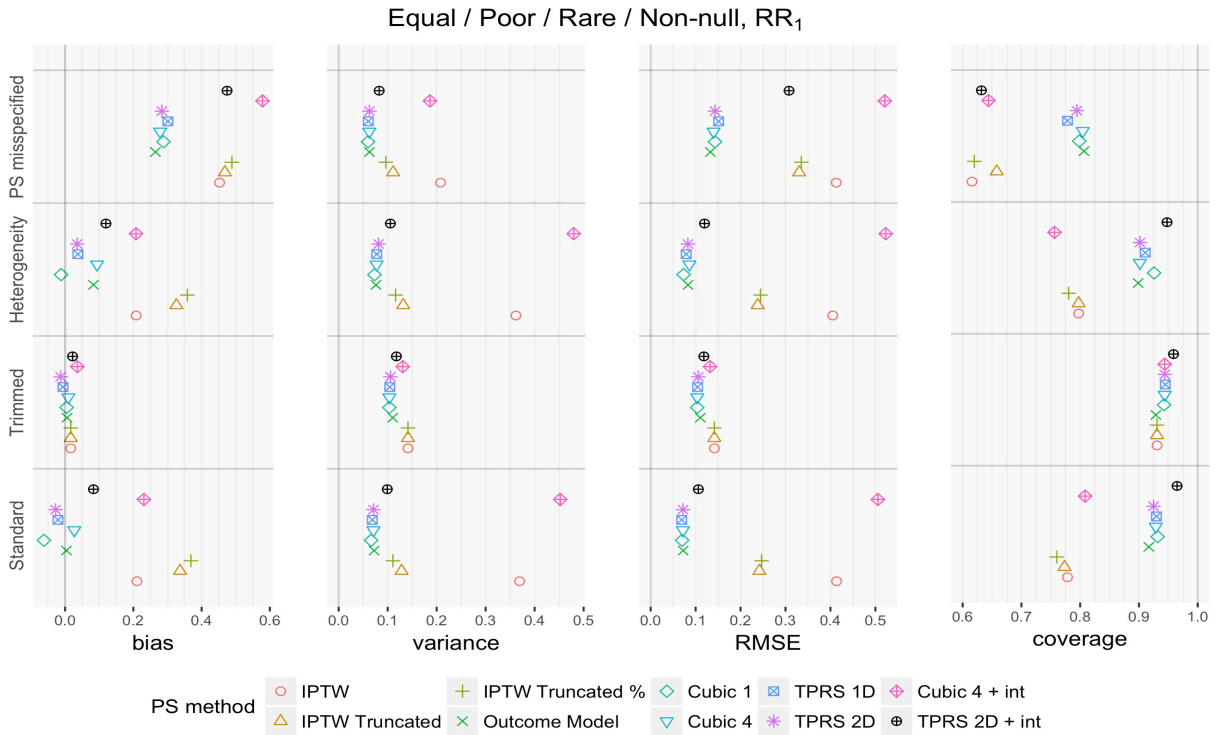


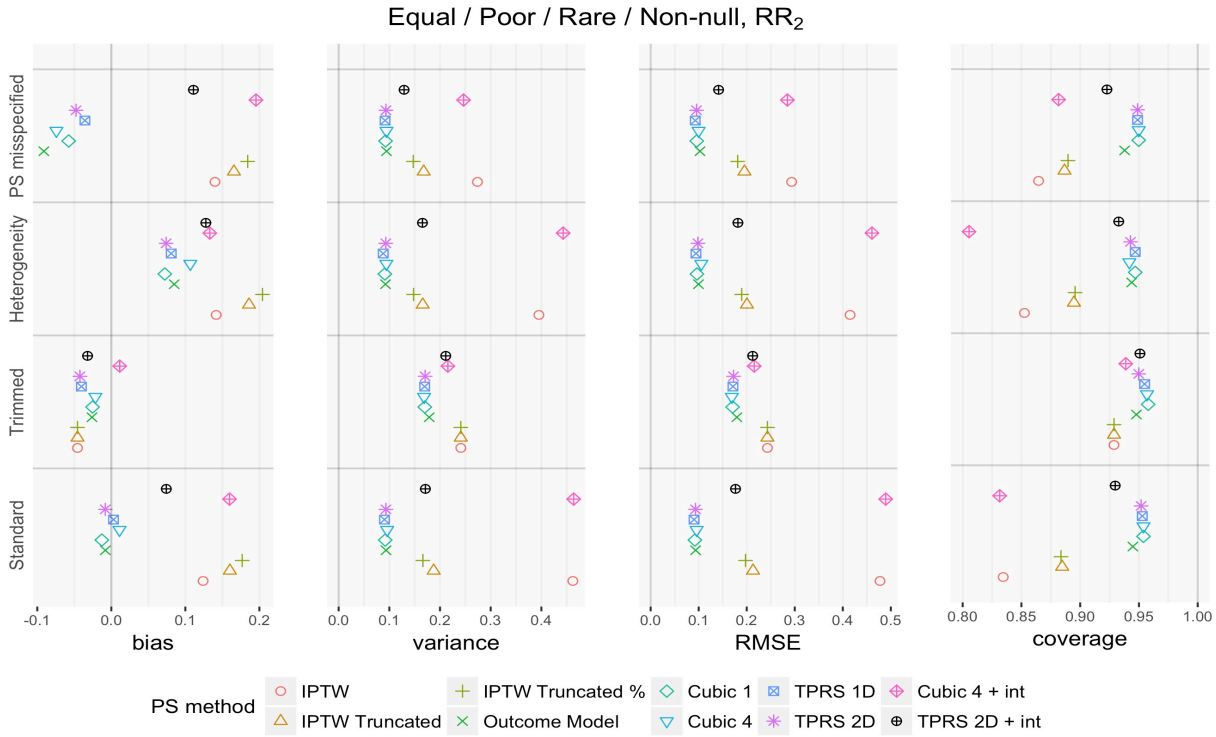
6.9 Equal / Poor / Rare / Null



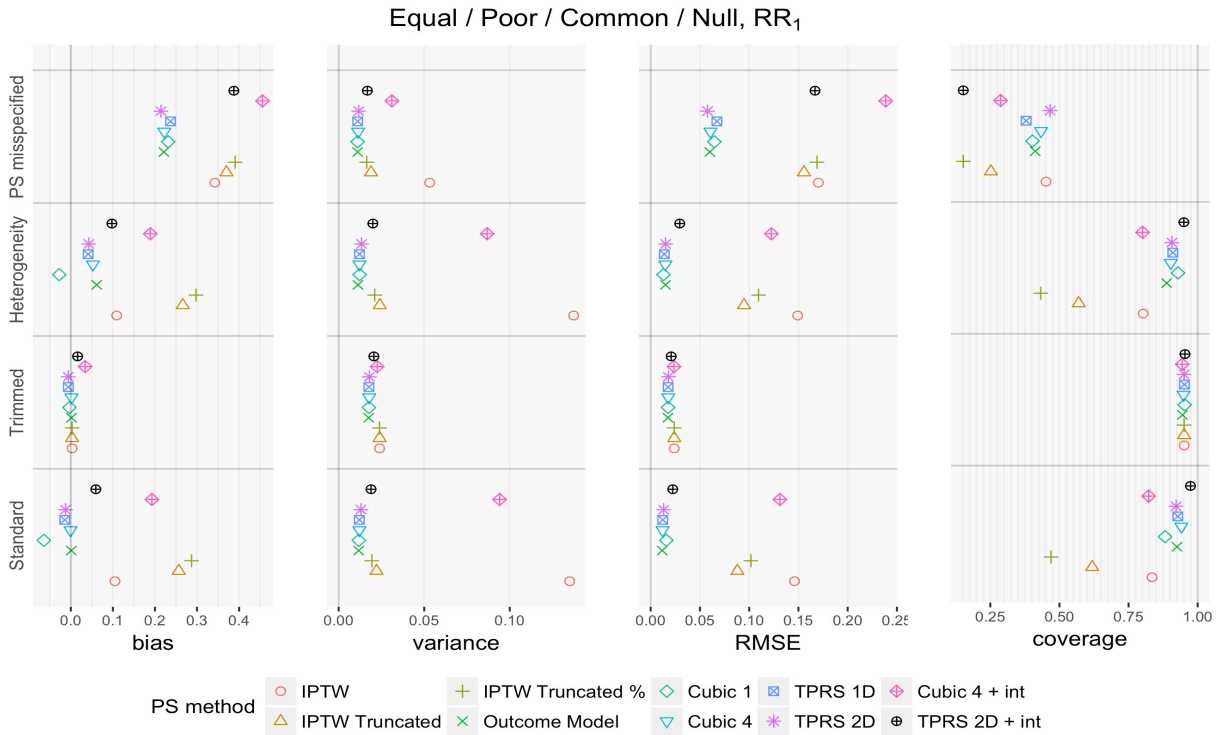


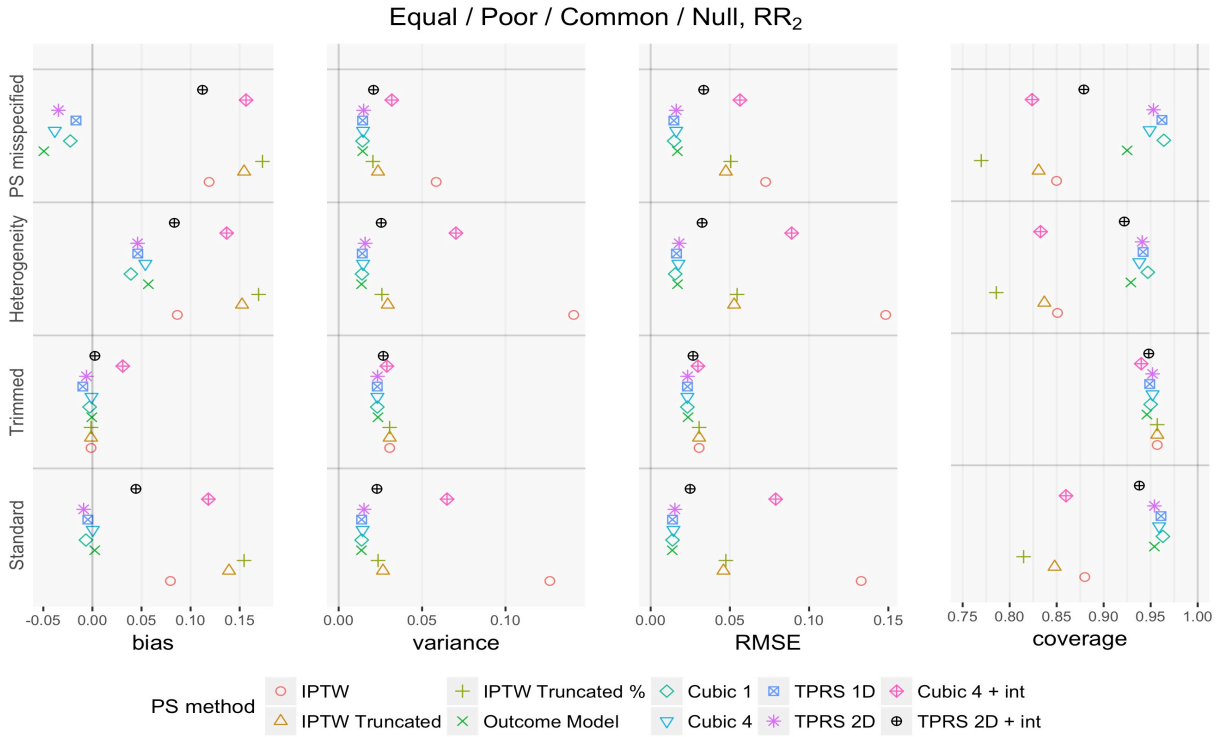
6.10 Equal / Poor / Rare / Non-null



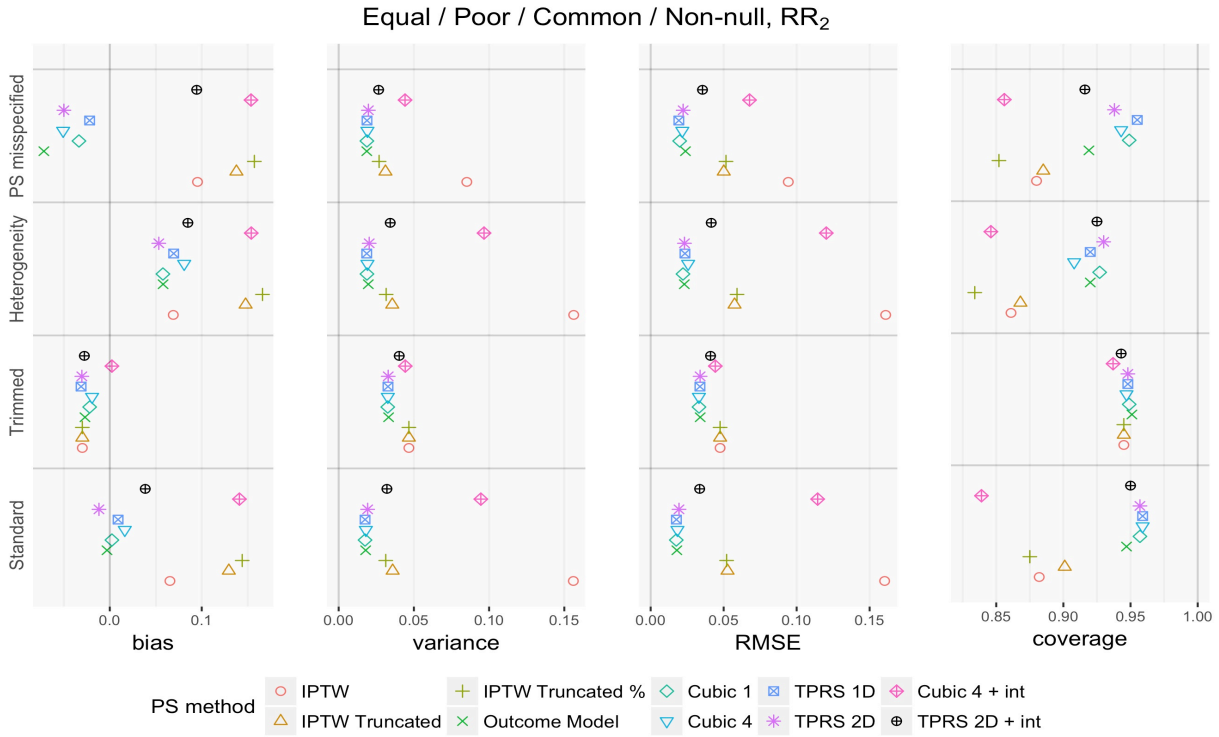


6.11 Equal / Poor / Common / Null

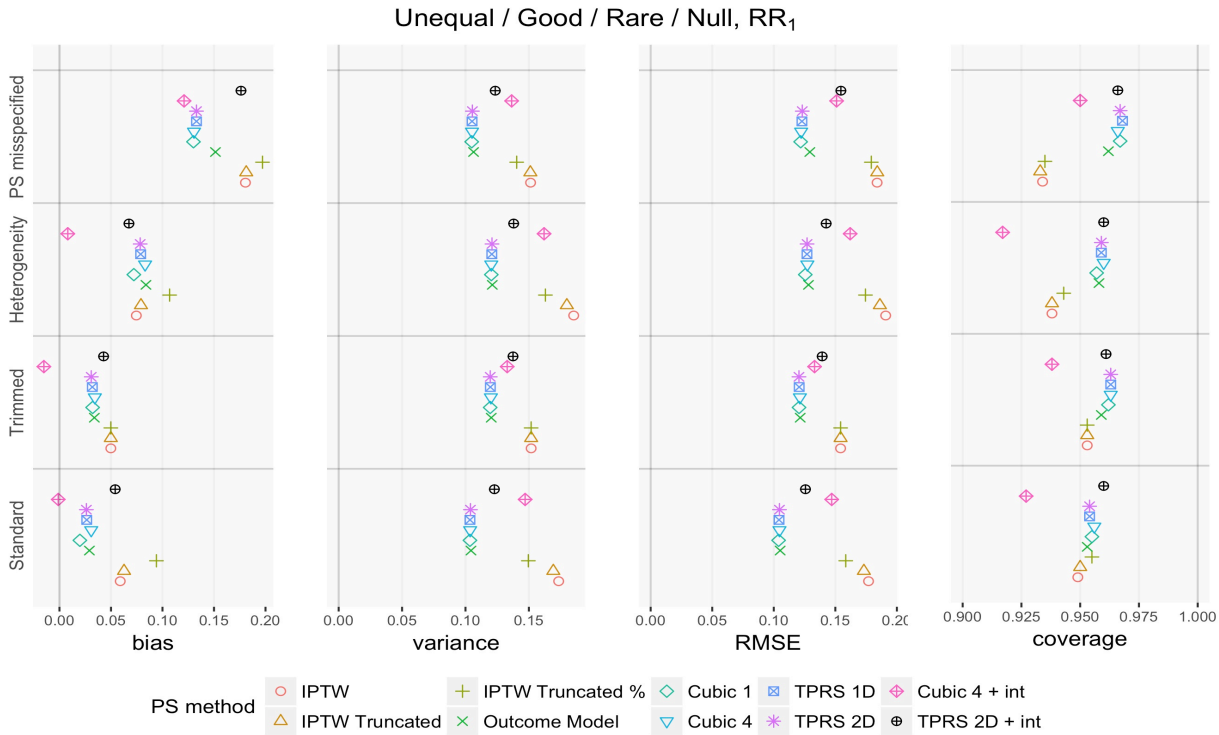


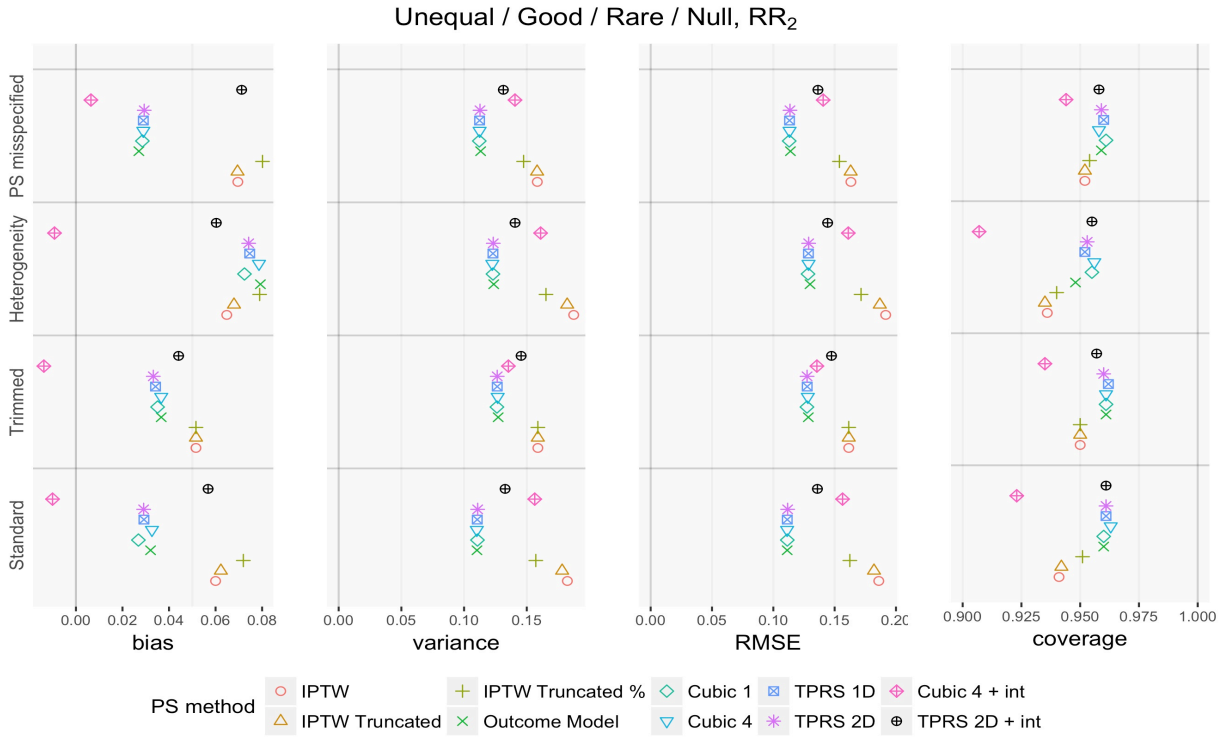


6.12 Equal / Poor / Common / Non-null

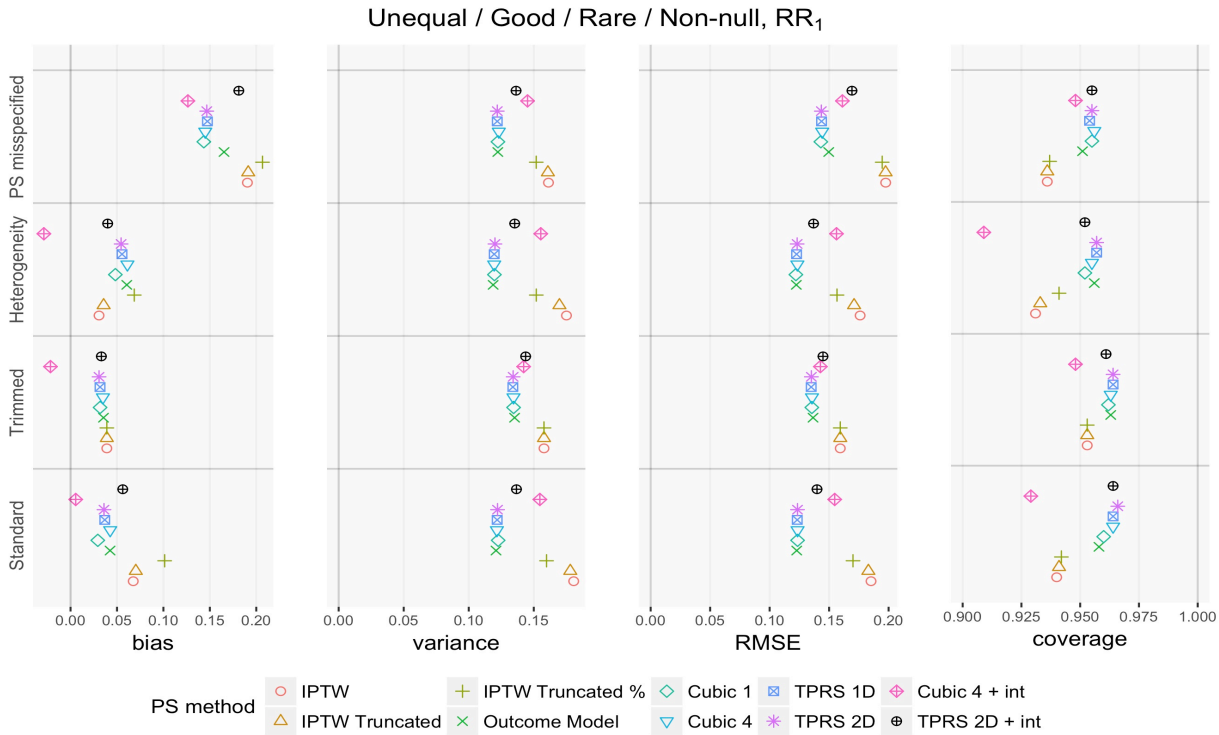


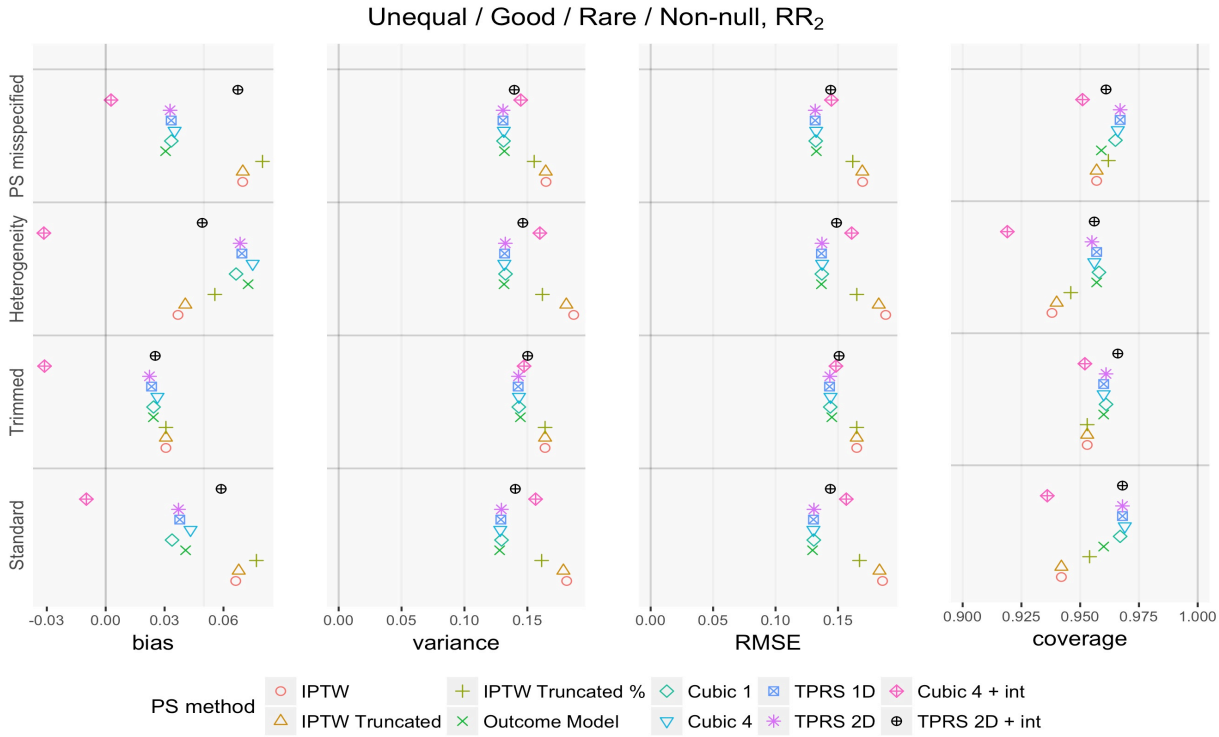
6.13 Unequal / Good / Rare / Null



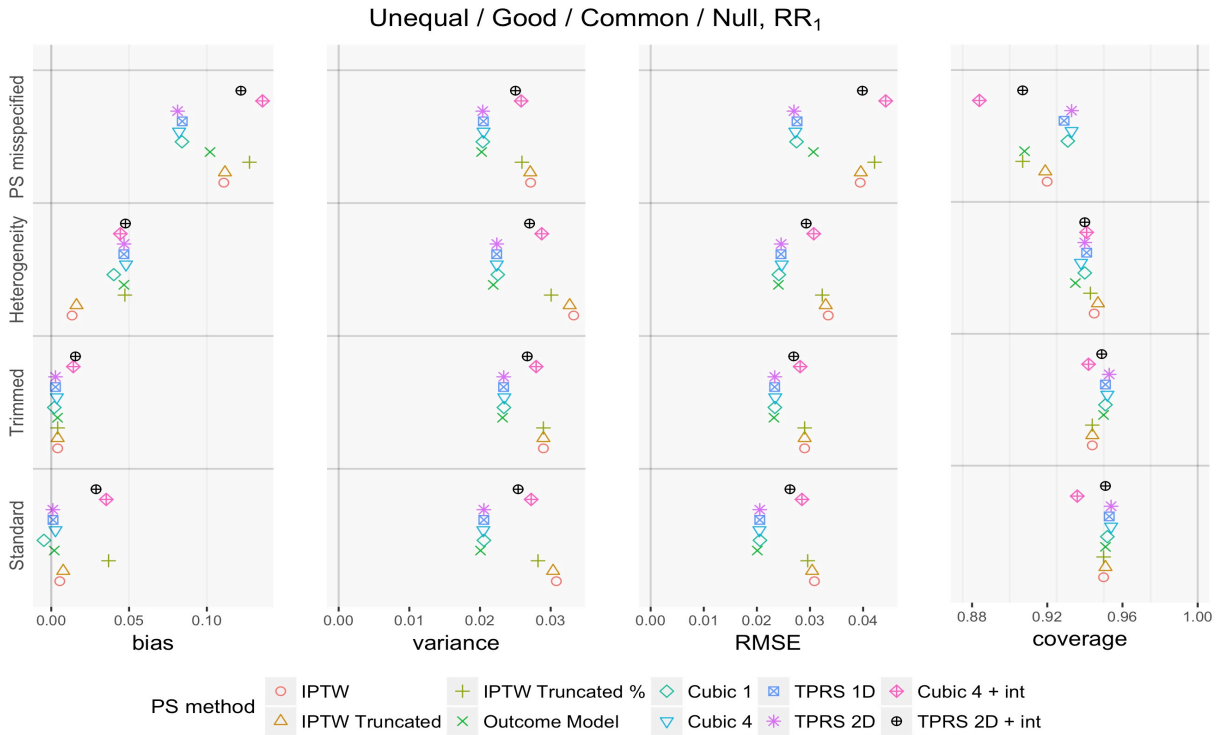


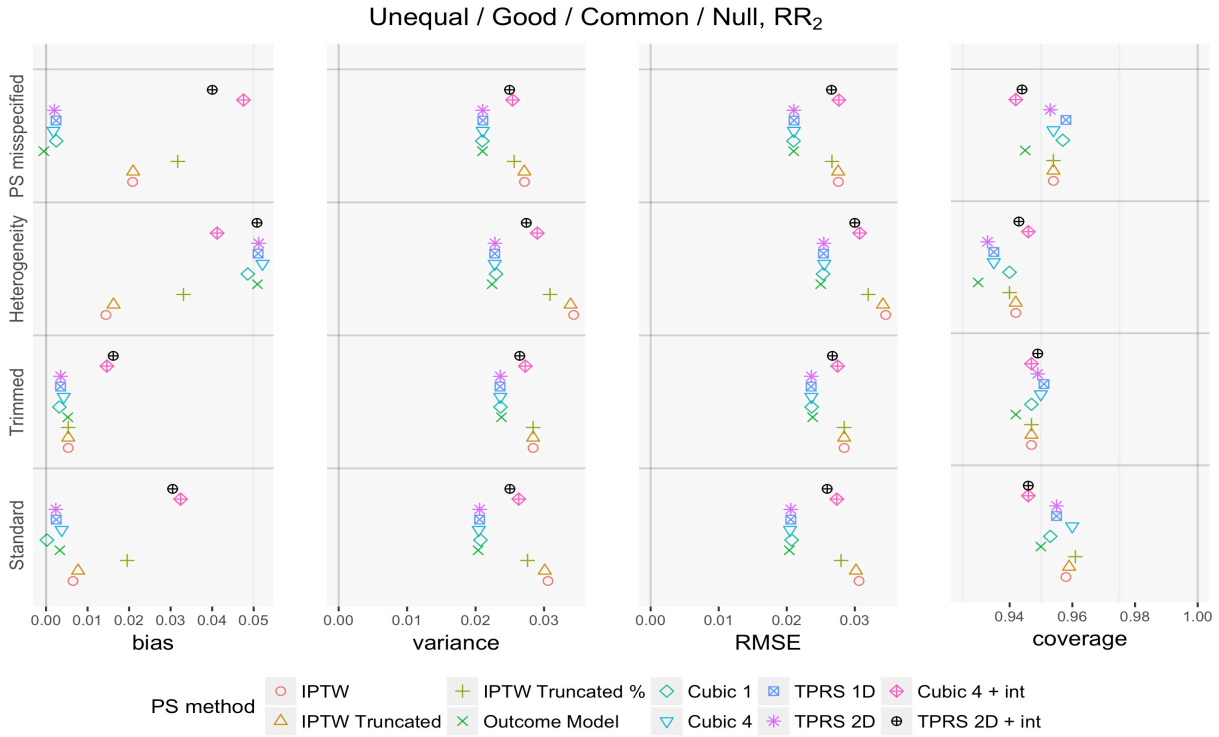
6.14 Unequal / Good / Rare / Non-null



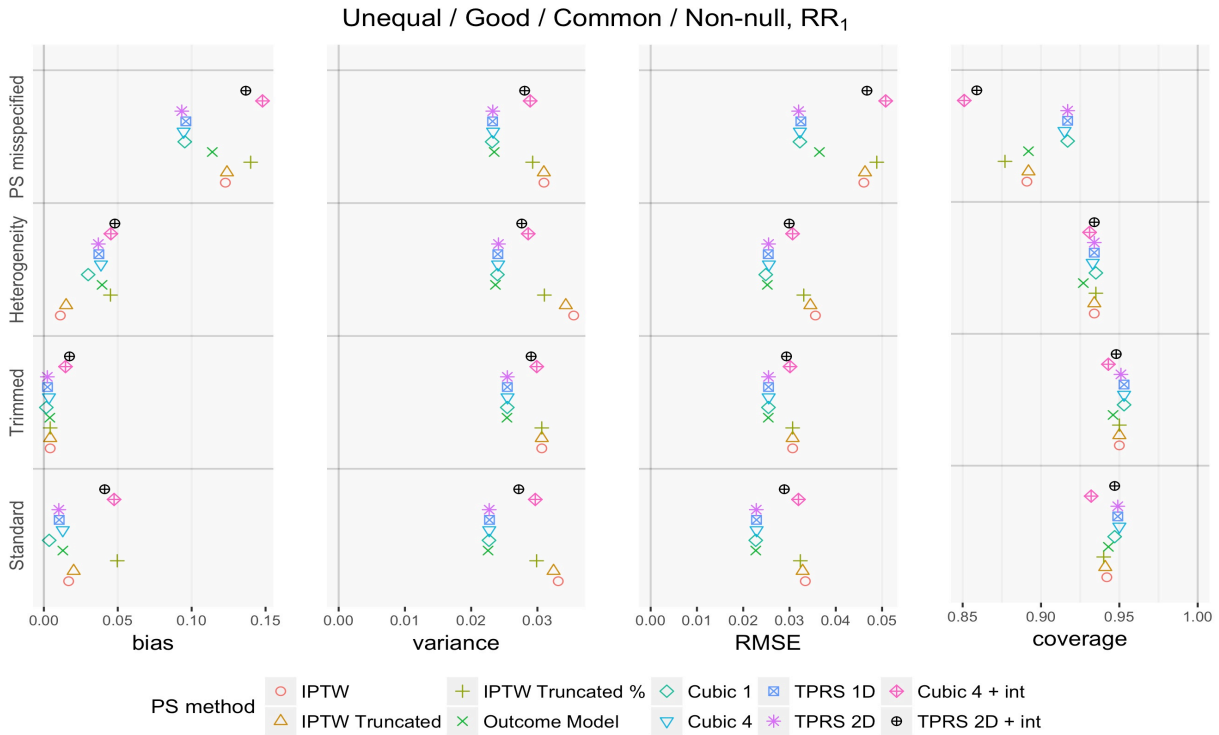


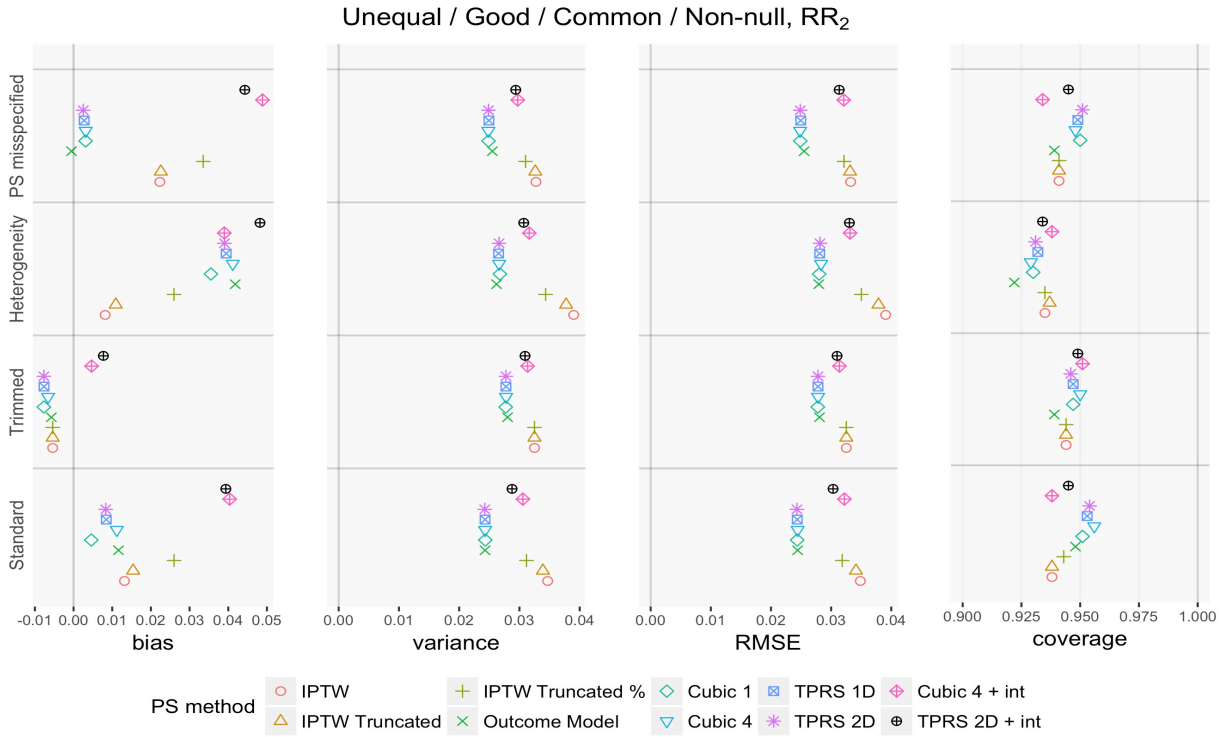
6.15 Unequal / Good / Common / Null



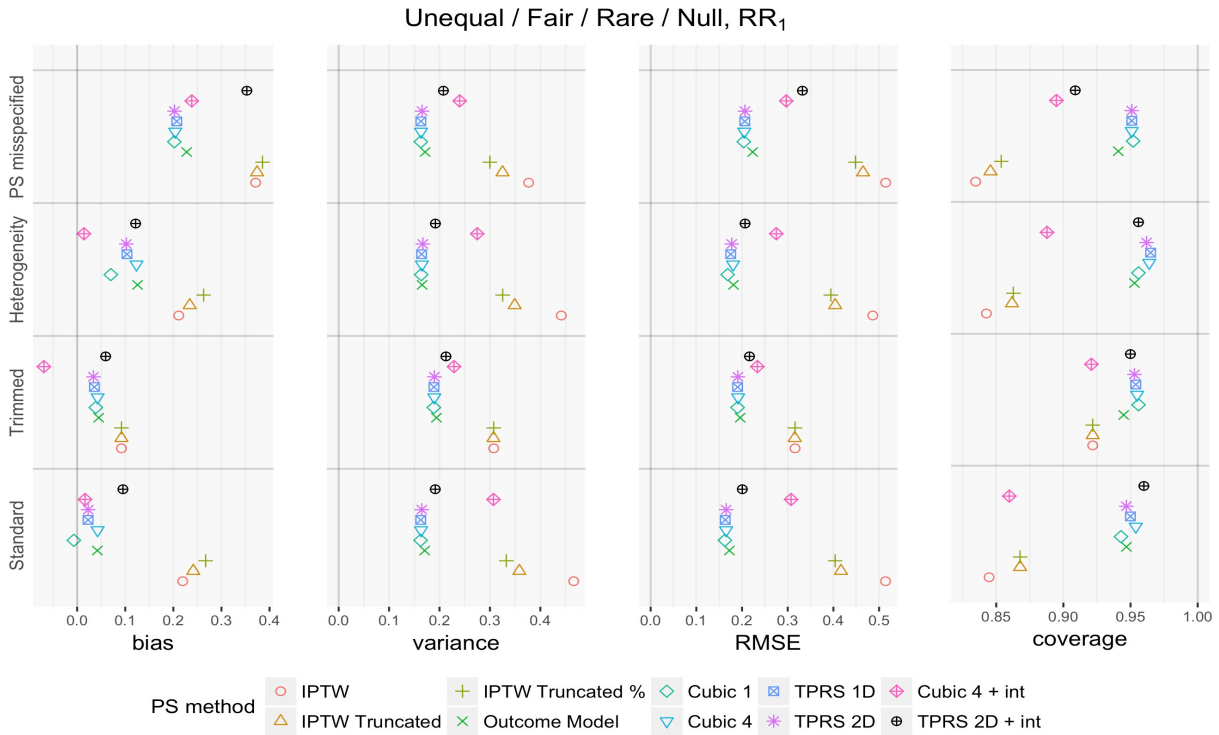


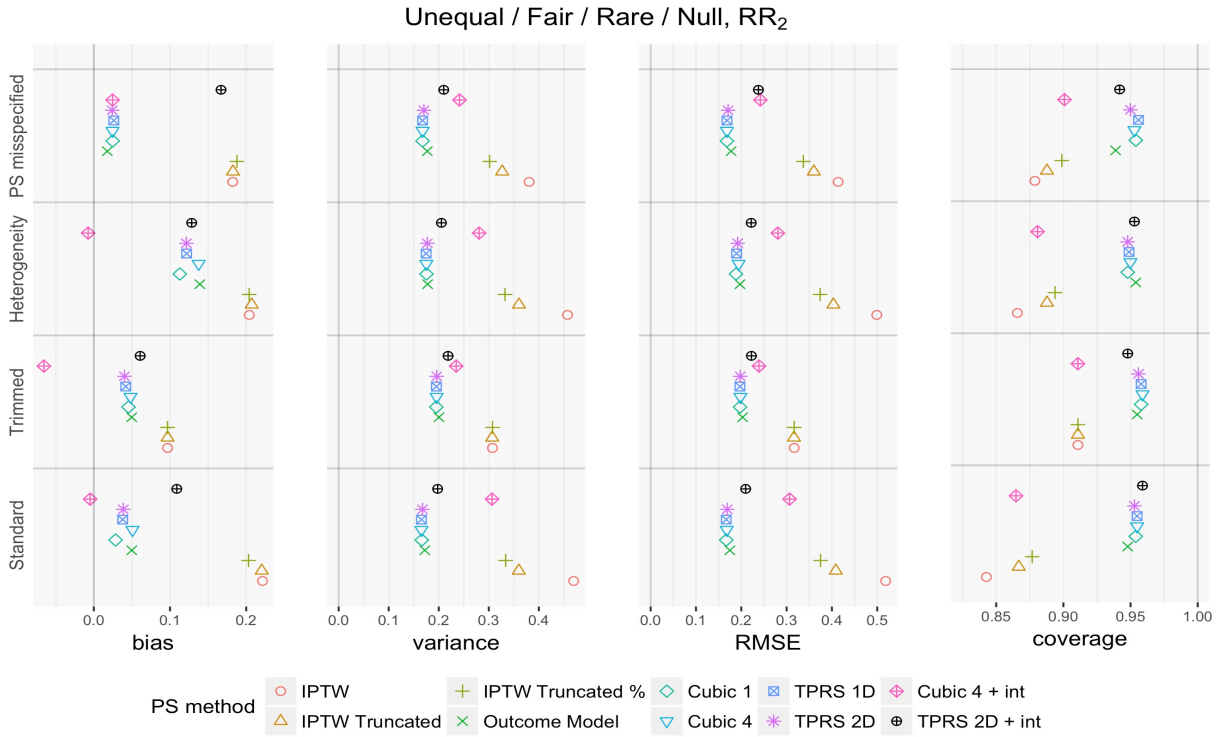
6.16 Unequal / Good / Common / Non-null



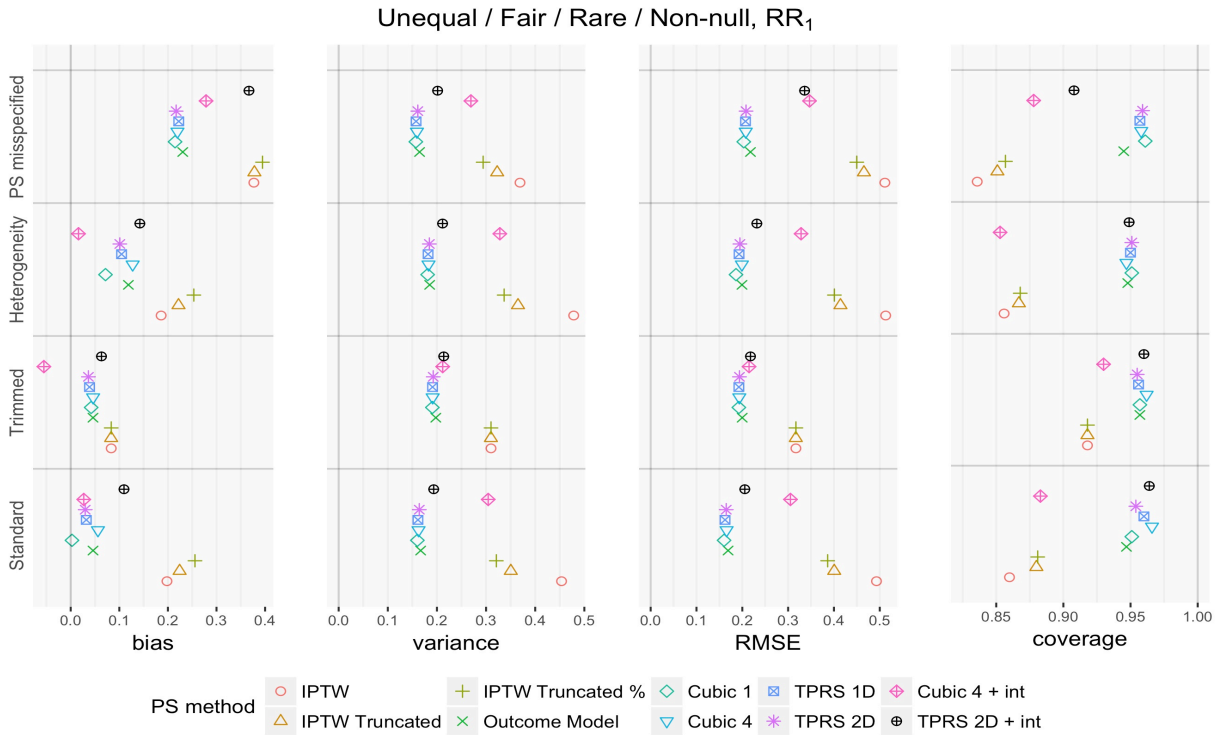


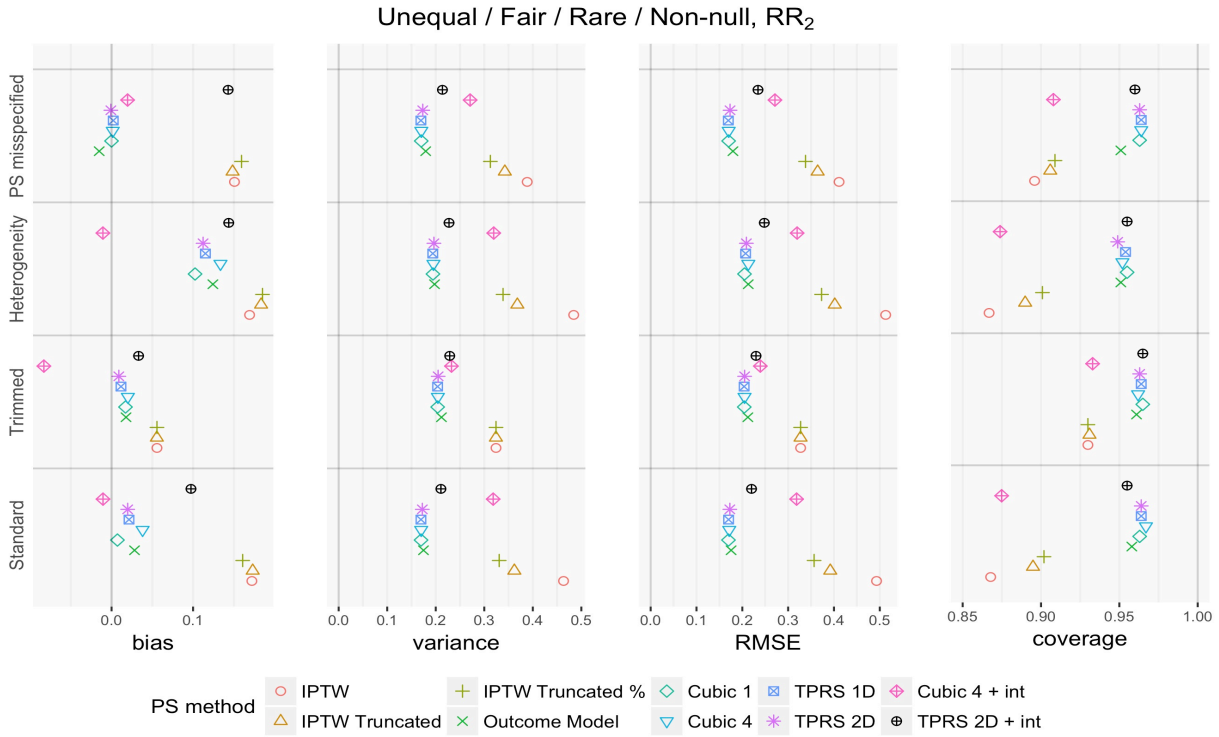
6.17 Unequal / Fair / Rare / Null



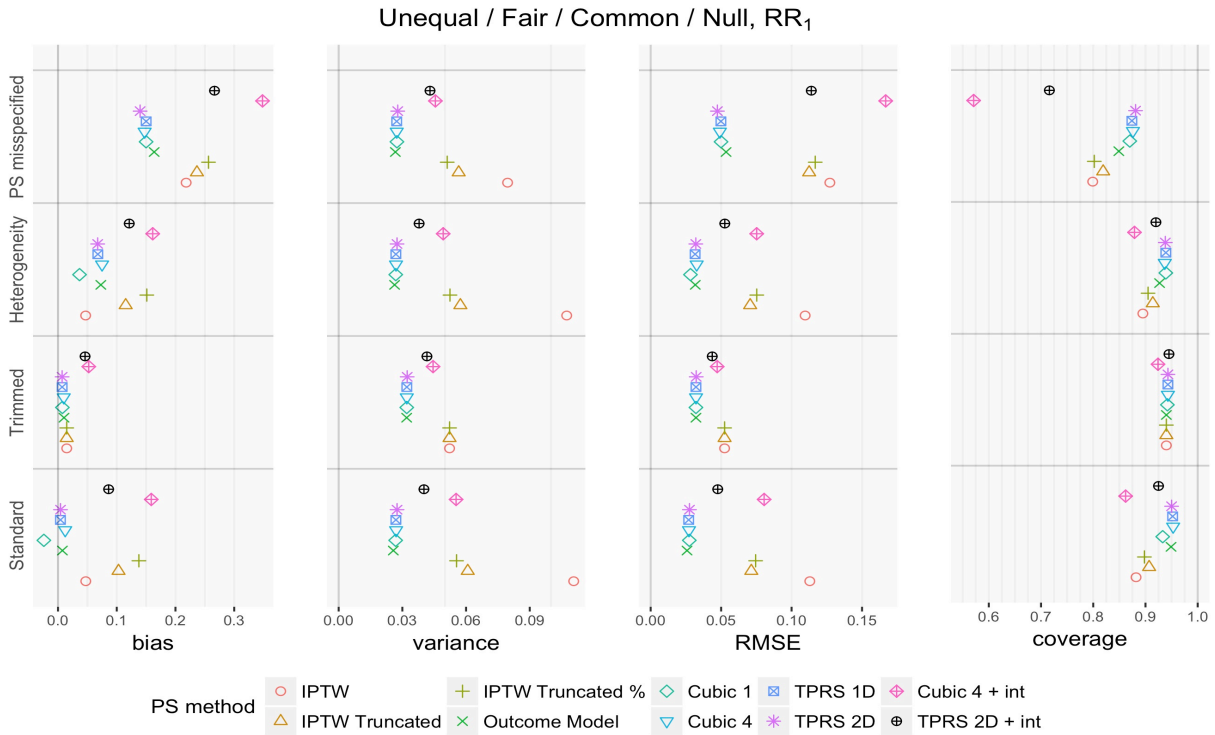


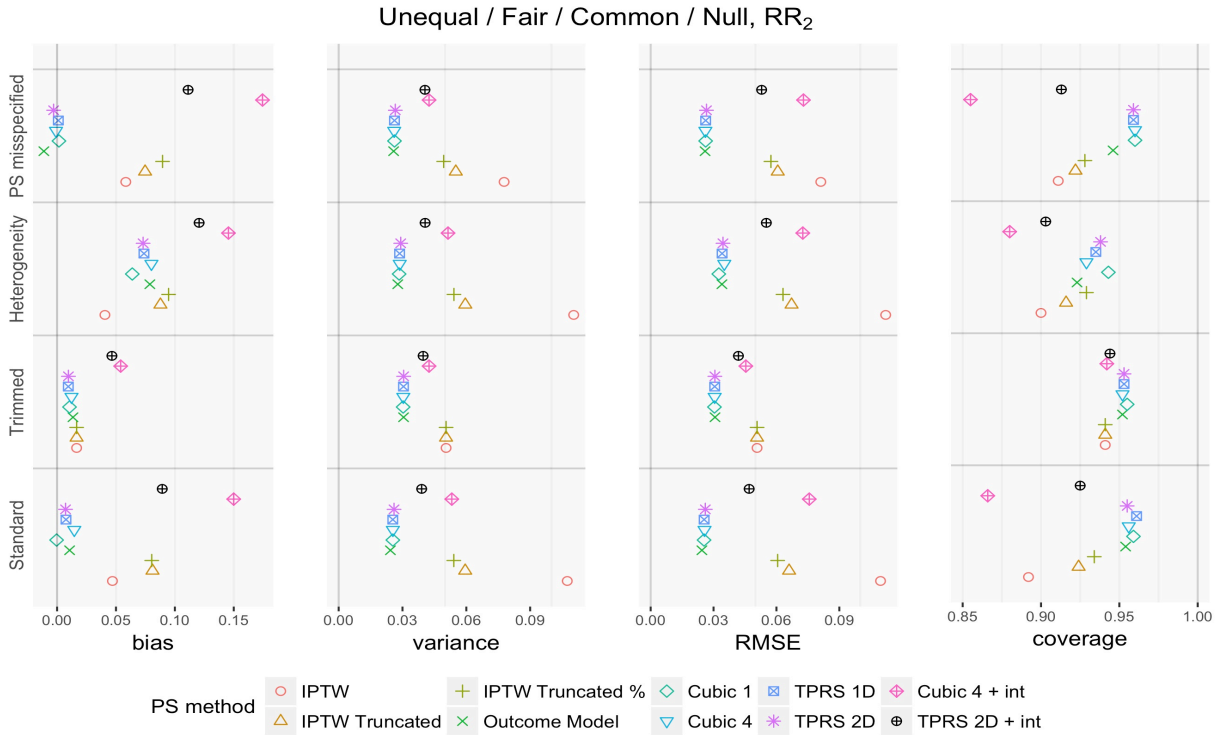
6.18 Unequal / Fair / Rare / Non-null



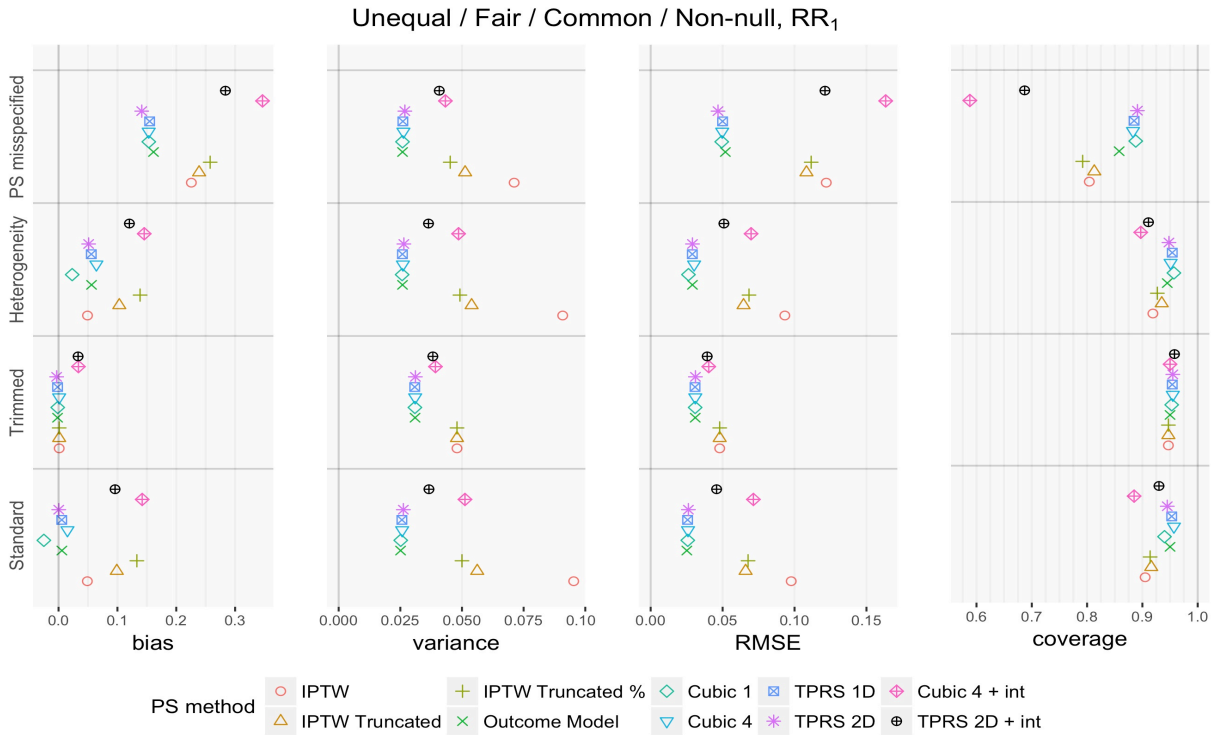


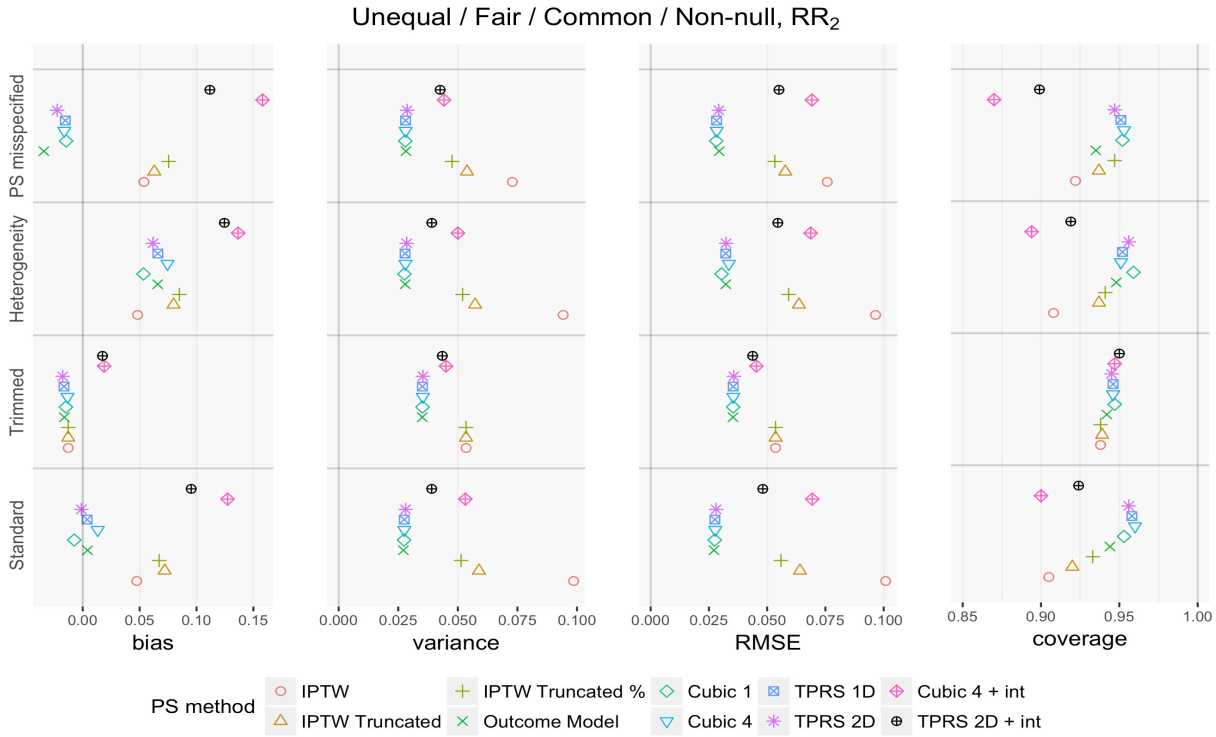
6.19 Unequal / Fair / Common / Null



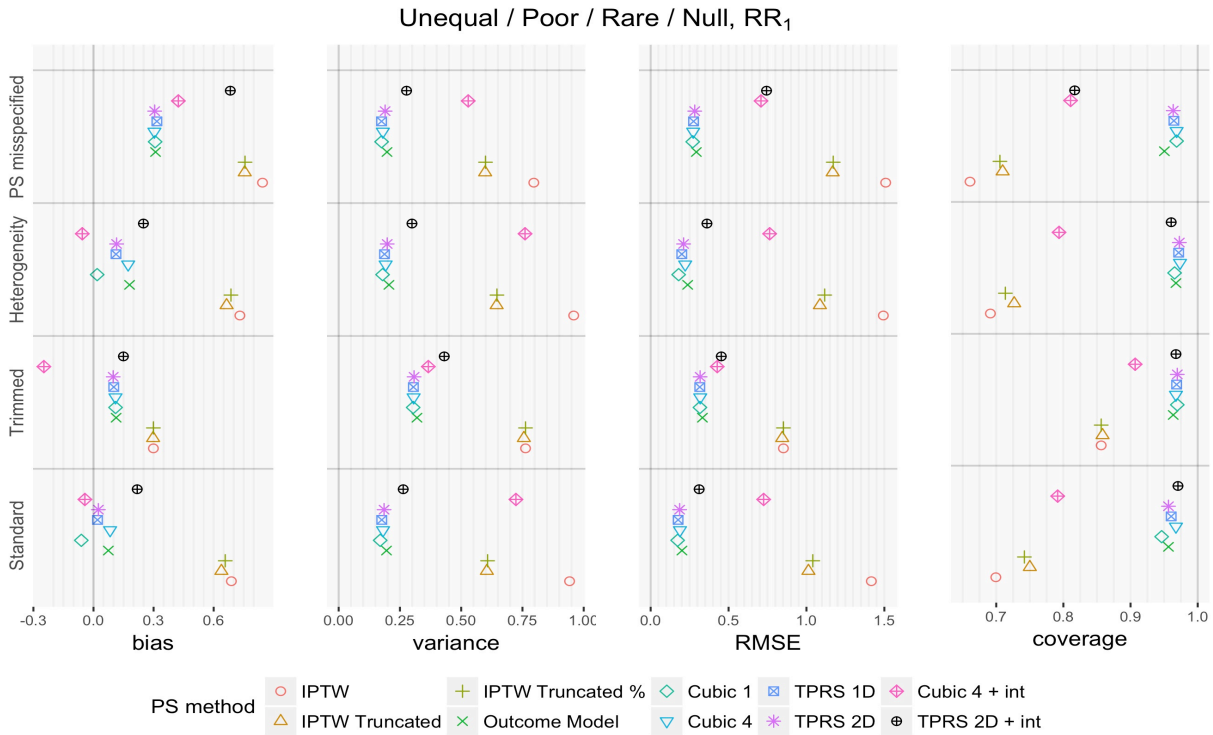


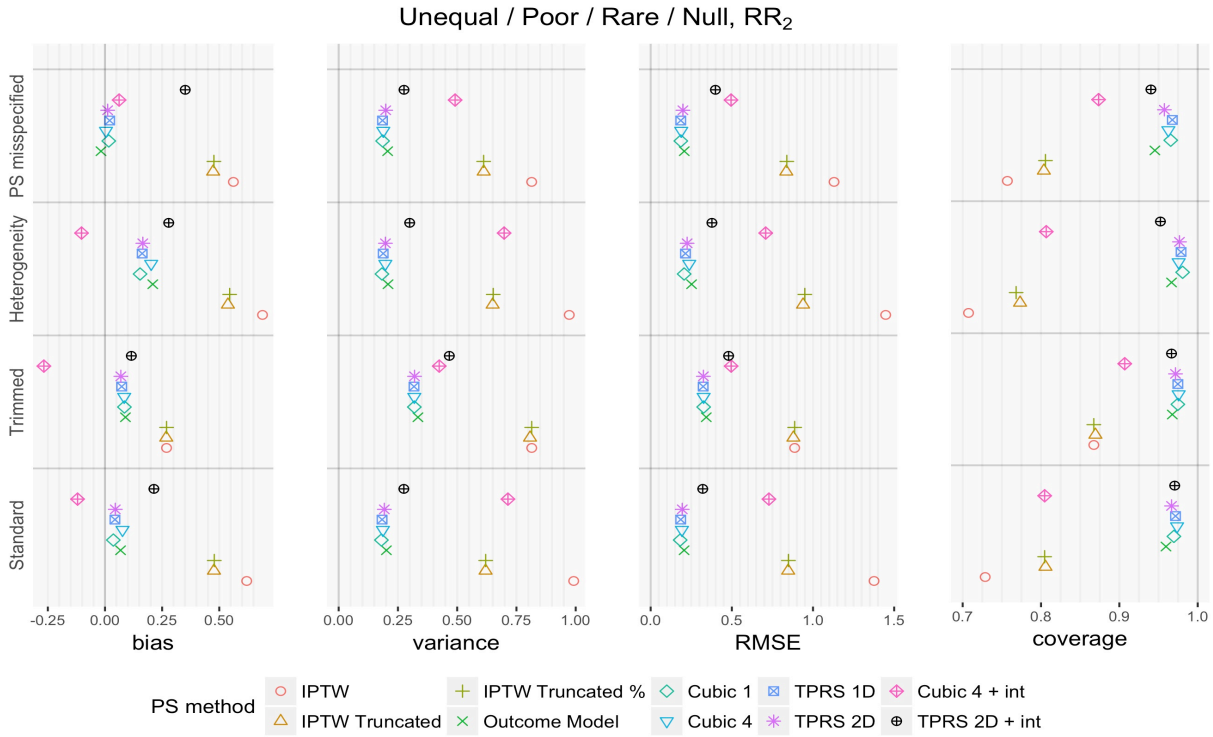
6.20 Unequal / Fair / Common / Non-null



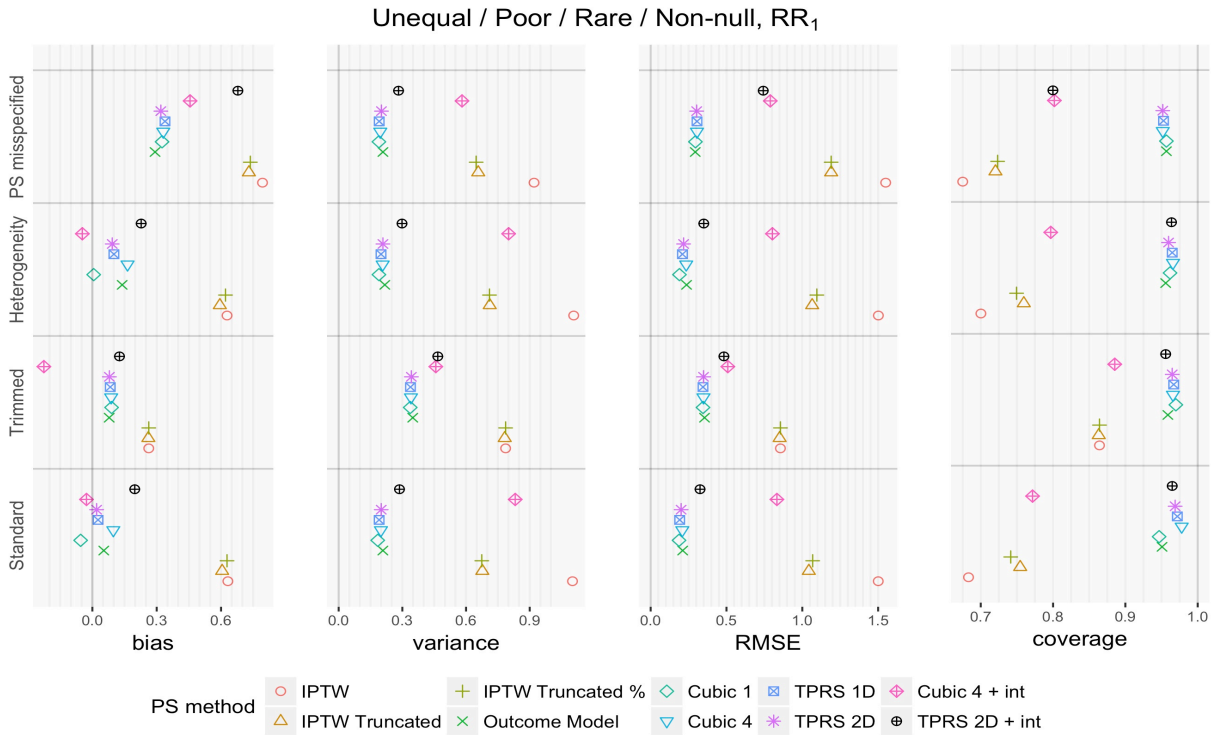


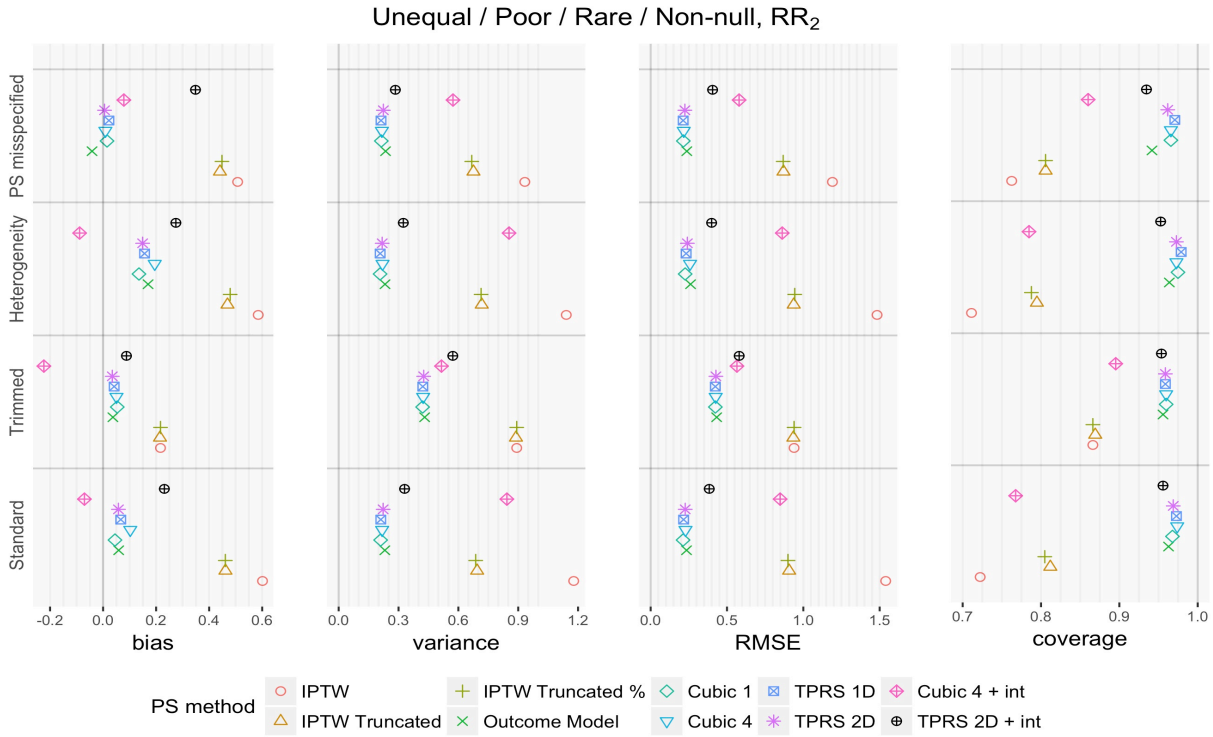
6.21 Unequal / Poor / Rare / Null



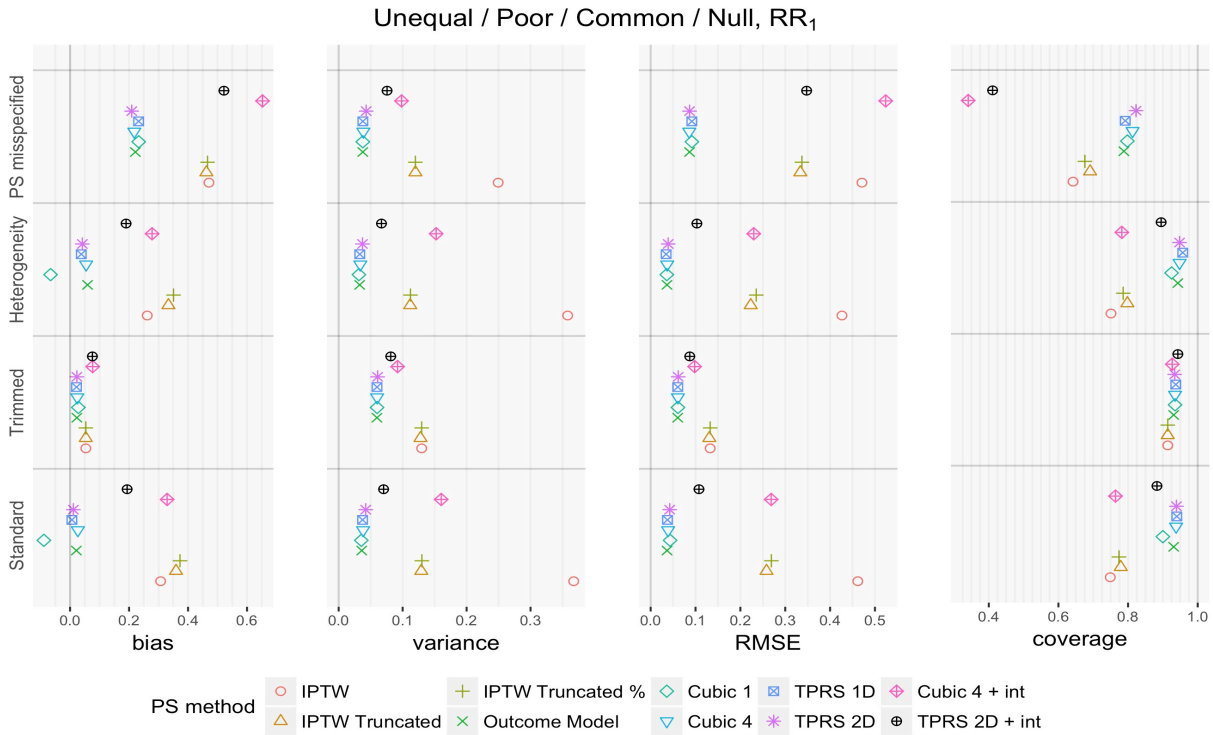


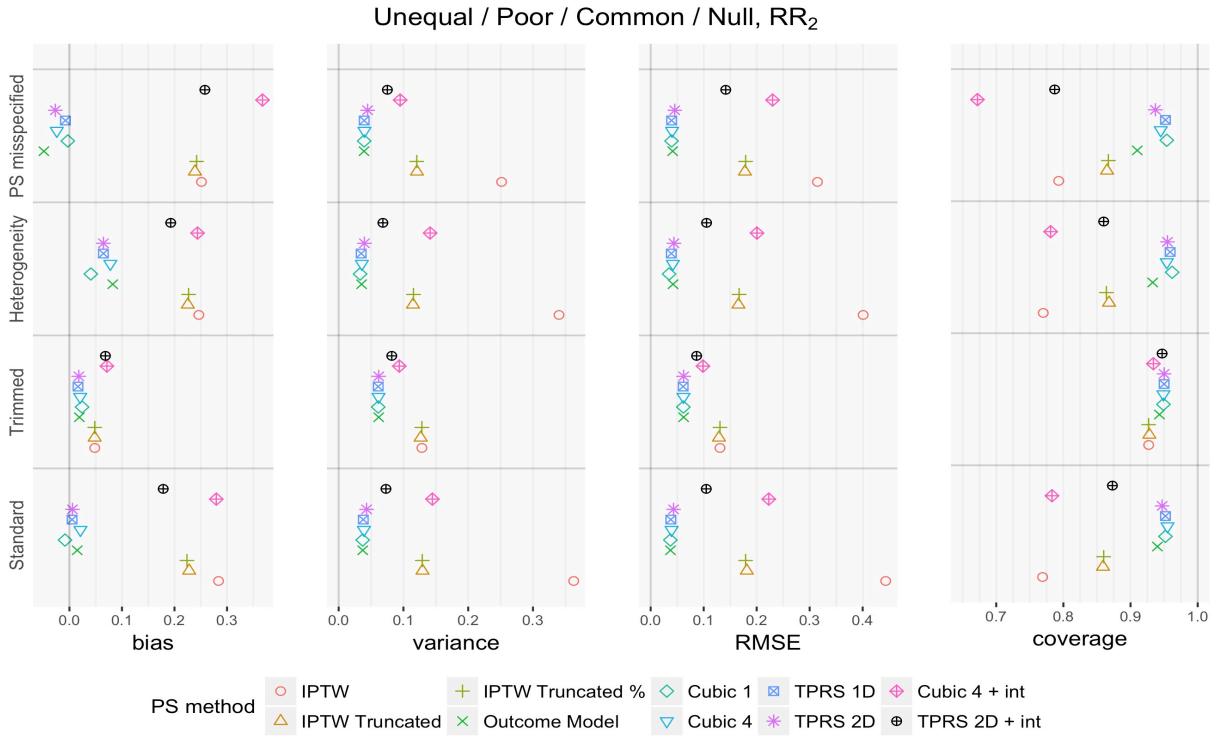
6.22 Unequal / Poor / Rare / Non-null



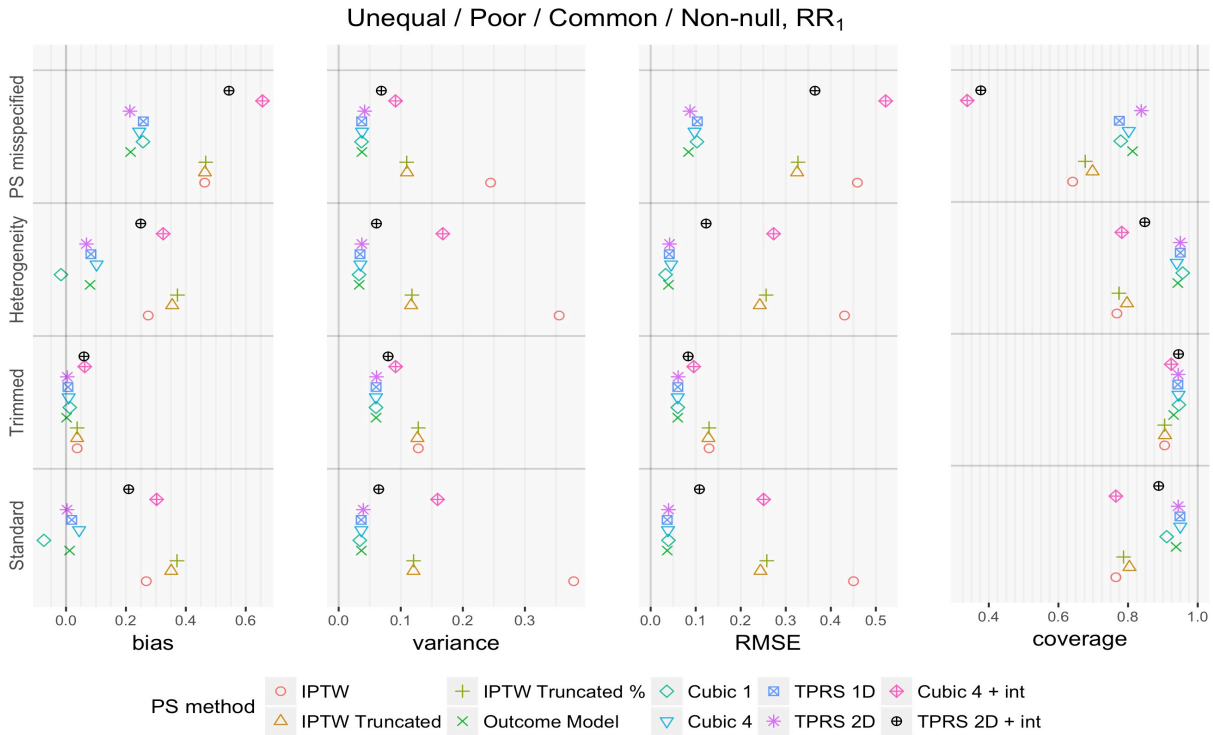


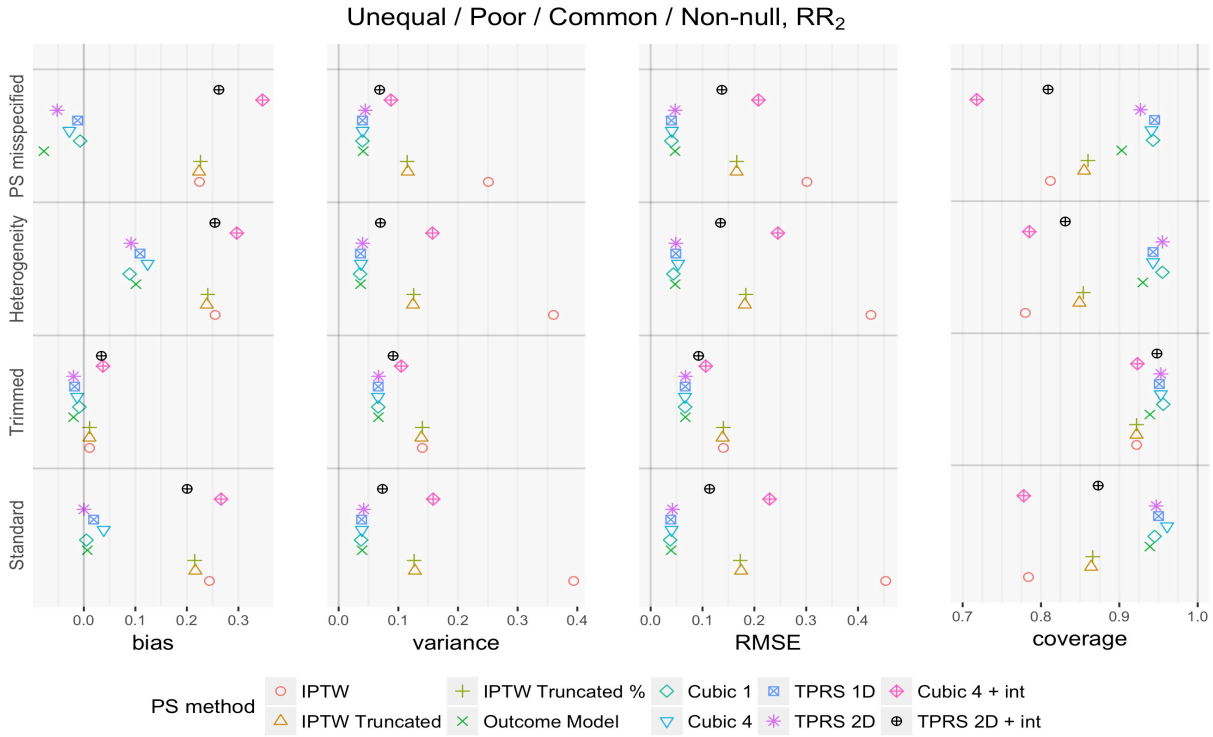
6.23 Unequal / Poor / Common / Null





6.24 Unequal / Poor / Common / Non-null





References

Franklin, J. M., Rassen, J. A., Ackermann, D., Bartels, D. B., Schneeweiss, S. (2014). Metrics for covariate balance in cohort studies of causal effects.*Statistics in medicine, 33*(10), 1685-1699.

Marra, G., Wood, S. N. (2012). Coverage properties of confidence intervals for generalized additive model components. *Scandinavian Journal of Statistics, 39*(1), 53-74.

Stürmer, T., Wyss, R., Glynn, R. J., Brookhart, M. A. (2014). Propensity scores for confounder adjustment when assessing the effects of medical interventions using nonexperimental study designs.*Journal of internal medicine,275*(6), 570-580.

Xu, S., Ross, C., Raebel, M. A., Shetterly, S., Blanchette, C., Smith, D. (2010). Use of stabilized inverse propensity scores as weights to directly estimate relative risk and its confidence intervals.*Value in Health,13*(2), 273-277.

Yoshida, K., Hernández-Díaz, S., Solomon, D. H., Jackson, J. W., Gagne, J. J., Glynn, R. J., Franklin, J. M. (2017). Matching weights to simultaneously compare three treatment groups: comparison to three-way matching. *Epidemiology (Cambridge, Mass.), 28*(3), 387.