



Received: 14 January 2020  
Accepted: 29 March 2020

\*Corresponding author: Dejian Lai,  
Department of Biostatistics and Data  
Science, The University of Texas  
School of Public Health, Houston,  
USA, Texas, 77030  
E-mail: [Dejian.lai@uth.tmc.edu](mailto:Dejian.lai@uth.tmc.edu)

Reviewing editor:  
Yueqing Hu, Fudan University School  
of Life Sciences, CHINA

Additional information is available at  
the end of the article

## STATISTICS | RESEARCH ARTICLE

# Estimations of treatment effects based on covariate adjusted nonparametric methods

Jiabu Ye<sup>1</sup> and Dejian Lai<sup>2\*</sup>

**Abstract:** Nonparametric tests are commonly used tests for two sample comparison in clinical studies. However, the estimation of treatment effects associated with the tests may not be obvious, especially under the covariate adjustment. In this article, we evaluated the effect of covariate adjustment on estimating treatment effects based on the Wilcoxon Rank Sum test, the van Elteren test, aligned rank test, and Jaeckel, Hettmansperger-McKean test through Monte Carlo simulations via mean square error and coverage probability. Based on the simulation, commonly used ANCOVA-based approach do not have good estimation of treatment effect when the covariate imbalance is severe. Aligned rank test seems perform well across most scenarios.

**Subjects:** Science; Mathematics & Statistics; Statistics & Probability; Statistics; Statistical Theory & Methods

**Keywords:** clinical trials; covariate adjustment; coverage probability; mean square error; treatment effects

### 1. Introduction

It is well known fully randomized trials or experiments have properties that assumptions for statistical comparison are met and observed treatment difference is unbiased estimate of true treatment difference for many statistical procedures. However, these properties are not guaranteed for small, poorly designed, or poorly conducted trials. When there is covariates imbalance, the treatment groups are not comparable even before the start of the trials. And the estimated treatment effect is biased. In this situation, it is necessary to adjust for covariates effect. Adjusting covariates will not only reduce the impact of covariate imbalance, but also improve precision for the estimation of treatment effect. When the covariates and the primary response measure are associated, adjusting for such covariate would generally improve the precision of the estimation.

Matching, stratification, and regression are commonly used approaches to adjust for covariates effects. Matching is commonly used observational studies. The treated subjects are matched with untreated subjects which share similar covariates values. However, it may exclude unmatched

### ABOUT THE AUTHORS

Jiabu Ye is senior statistician at AstraZeneca Pharmaceuticals. He is study statistician for multiple ongoing immuno-oncology phase III studies. His research interest is in clinical trial related topics.

Dejian Lai is professor of biostatistics and data science at The University of Texas Health Science Center at Houston School of Public Health. His research areas include nonparametric statistical methods, spatio-temporal models and their applications in public health.

### PUBLIC INTEREST STATEMENT

Many factors may inference the interpretations of results from clinical trials. Statistical methods for designing and analyzing clinical trials are under investigation all the time. In this article, we performed simulation studies on many nonparametric statistical methods with covariate adjustments.

subjects. Stratification is subclassification of subjects into mutual exclusive categories based on the value of covariates. The overall treatment effect is estimated as weighted average of within strata treatment effect. It will be problematic if there are many covariates need be adjusted. Regression or ANCOVA approaches adjust covariate effect by treating covariate as regression parameters. After adjusting the covariates, within treatment variance is reduced and confounds are eliminated. Regression or ANCOVA approaches rely on their hypotheses which include linearity relationship between covariates and response, independent identical normally distributed residuals and same covariate coefficients. Under certain scenarios, these assumptions may be violated. Moreover, selection of covariates in the regression model is subjective and overfit will cause the estimation be biased.

Although covariate adjustment has been widely applied to clinical studies. Few studies focus on the evaluation combining covariate adjusting approaches with both parametric and nonparametric approaches of estimating treatment effect when covariate are imbalanced. This study tried to fill the gap in between.

## 2. Treatment effect of two sample comparison adjusting for covariates

Before evaluation of estimation of treatment effect by different approaches, it is necessary to test if the magnitude of treatment effect is statistically significant (Ye & Lai, 2019). There are many tests that could evaluate the significance of treatment effect. Some of these tests may be valid after adjusting for covariates when the covariates are not fully balanced. Some tests may be invalid after adjusting for non-balanced covariates. The evaluation of several nonparametric tests in terms of their power and type I error were investigated (Ye & Lai, 2019). In current article, we evaluate the treatment effect after adjusting for unbalanced covariates.

Here, response variable  $Y$  can be measurement of change from baseline to endpoint. For simplicity, we denote the measure of change as  $Y$ .

### 2.1. Mean difference (TT)

The first statistic to estimate the treatment effect is mean difference between groups, it is also known as average treatment effect. This mean difference statistic is associated with  $T$  test.

$$\hat{\beta}_{Tt} = \bar{Y}_{Tt} - \bar{Y}_{ctrl}$$

### 2.2. Hodges Lehmann estimate (WRS)

Hodges Lehmann estimate of treatment effect for two sample is the median of all paired differences between responses in two groups (Hodges & Lehmann, 1963).

$$\hat{\beta}_{Tt} = \text{median}\{ (Y_{Tt,i} - Y_{ctrl,j}), i = 1, \dots, m; j = 1, \dots, n \}$$

where  $Y_{Tt,i}$  is the response in treatment group and  $Y_{ctrl,j}$  is the response in control group.  $m$  and  $n$  denote the number of subjects in treatment group and control group correspondingly. The above Hodges-Lehmann estimator  $\hat{\beta}_{Tt}$  is associated with Wilcoxon rank sum test (Wilcoxon, 1945).

### 2.3. Jaeckel's rank estimation of regression parameter, regress treatment variable only (JHM(n))

Jaeckel proposed a rank regression based approach to estimate regression parameters (Jaeckel, 1972). The basic idea is to find a set of estimate  $\hat{\beta}$  for regression parameters by minimizing a dispersion function  $D_J(Y_i - Z_i'\beta) = \sum a(R_i(\beta))(Y_i - Z_i'\beta)$ , where  $R_i(\beta)$  is the rank of  $Y_i - X_i'\beta$  and  $a(\cdot)$  is non-decreasing rank score function. Generally, there is no closed for the dispersion function and numeric approach is required to get the estimate. In this approach, we only consider treatment variable as the regression parameter and ignore the other covariate. And the estimate of treatment effect  $\hat{\beta}_{Tt}$  is the solution to minimize the following dispersion function.

$$D_J(Y_i - Trt_i \hat{\beta}_{Tt}) = \sum a(R_i(\beta))(Y_i - Trt_i \hat{\beta}_{Tt}) = \min$$

This point estimation is associated with Jaeckel, Hettmansperger-McKean Test (Hettmansperger & McKean, 1978), in which the null hypothesis is  $\beta_{Trt}$  equal to 0 and there is no remaining covariate in the linear rank regression model.

**2.4. Jaeckel’s rank estimation of regression parameter, treatment variable, and the covariate (JHM(x))**

In this Jaeckel’s rank estimation, we treat treatment variable and the covariate as the regression parameter. The estimate of treatment effect  $\hat{\beta}_{Trt}'$  is solution to minimize the corresponding dispersion function.

$$D_J(Y_i - Trt_i \hat{\beta}_{Trt}' - X_i \hat{\beta}_X) = \sum a(R_i(\beta))(Y_i - Trt_i \beta_{Trt} - X_i \beta_X) = \min!$$

In the associated JHM test, the null hypothesis is  $\beta_{Trt}$  equals to 0 and there is one remaining covariate X in the linear rank regression model.

**2.5 Adjusting covariate effect based on Jaeckel’s rank estimation and Hodges Lehmann estimate of treatment effect (JHM(x)-WRS)**

In this approach, we adjust for covariate effect first for each individual subjects and  $Y_i^{adj} = Y_i - X_i \hat{\beta}_X$ .  $\hat{\beta}_X$  is the solution to minimize the corresponding dispersion function  $D_J(Y_i - Trt_i \hat{\beta}_{Trt}' - X_i \hat{\beta}_X)$ . Then the Hodges Lehmann estimator is applied to estimate treatment effect

$$\hat{\beta}_{Trt} = \text{median}\left\{ \left( Y_{Trt,i}^{adj} - Y_{Ctrl,j}^{adj} \right), i = 1, \dots, m; j = 1, \dots, n \right\}$$

$Y_{Trt,i}^{adj}$  is the response in treatment group and  $Y_{Ctrl,j}^{adj}$  is the response in control group.

**2.6. Adjusting covariate effect based on ANCOVA and Hodges Lehmann estimate of treatment effect (ANCOVA-WRS)**

Similar to JHM(x)-WRS, we adjust for covariate effect first and  $Y_i^{adj} = Y_i - X_i \hat{\beta}_X$ .  $\hat{\beta}_X$  is estimate of covariate coefficient based on ANCOVA model:  $Y_i = \beta_0 + Trt_i \beta_{Trt} + X_i \hat{\beta}_X$ . The Hodges Lehmann estimator is similar to the step in JHM(x)-WRS.

**2.7. Adjusting covariate effect based on quintile stratification of the covariate and align for strata effect; Hodges Lehmann estimate of treatment effect (AR(x5))**

The continuous covariate is stratified based on its quintile. The covariate effect is adjusted by alignment within each strata and the  $k^{th}$  strata effect is estimated by the Walsh average  $\hat{\delta}_i$ ,  $\hat{\delta}_i = \text{median}\left\{ \frac{Y_i + Y_i'}{2}, i \leq j = 1, \dots, n \right\}$ , here n is number of subjects in  $k^{th}$  strata and  $Y_i^{align} = Y_i - \hat{\delta}_i$ , where  $Y_i$  is subjects in the  $k^{th}$  strata. And the treatment effect is defined as

$$\hat{\beta}_{Trt} = \text{median}\left\{ \left( Y_{Trt,i}^{align} - Y_{Ctrl,j}^{align} \right), i = 1, \dots, m; j = 1, \dots, n \right\}$$

$Y_{Trt,i}^{align}$  is the aligned response in treatment group and  $Y_{Ctrl,j}^{align}$  is the aligned response in the control group.

**2.8. Adjusting covariate effect based on treatment matched quintile stratification of the covariate and align for strata effect; Hodges Lehmann estimate of treatment effect (AR(MS,x5))**

The continuous covariate is stratified based on its conditional quintile condition on the treatment assignment. Therefore, the number of subjects of both treatment groups within each stratum are equal, whereas in AR(x5) the number of each treatment group within each stratum are not equal. The following procedures are similar to AR(x5). The  $k^{th}$  strata effect is estimated by one sample Hodges Lehmann estimator  $\hat{\delta}_i' = \text{median}\left\{ \frac{Y_i' + Y_i}{2}, i \leq j = 1, \dots, n \right\}$ . And each response aligned for the strata effect and  $Y_i^{align'} = Y_i - \hat{\delta}_i'$ , where  $Y_i$  is subjects in the  $k^{th}$  strata. And the treatment effect is defined as

$$\hat{\beta}_{Trt} = \text{median} \left\{ \left( Y_{Trt,i}^{align'} - Y_{Ctrl,j}^{align'} \right), i = 1, \dots, m; j = 1, \dots, n \right\}$$

$Y_{Trt,i}^{align'}$  is the aligned response in treatment group and  $Y_{Ctrl,j}^{align'}$  is the aligned response in the control group (Hodges & Lehmann, 1962, 1963).

**2.9 Adjusting covariate effect based on quintile stratification of the covariate and align for strata effect; Hodges Lehmann estimate of within strata treatment effect; inverse variance weighted overall treatment effect (Aligned-WS(HL,iv,x5))**

The first step is also to stratify the continuous covariate based on its quintile. Then the response variable is adjusted by alignment. In the  $k^{th}$  strata effect, the strata effect is estimated by the Walsh average  $\hat{\delta}_k, \hat{\delta}_k = \text{median} \left\{ \frac{Y_i + Y_j}{2}, i \leq j = 1, \dots, n \right\}$ , here  $n$  is number of subjects in  $k^{th}$  strata and  $Y_i^{align} = Y_i - \hat{\delta}_k$ , where  $Y_i$  is subjects in the  $k^{th}$  strata. Unlike previous approaches to adjust covariate effect by estimating strata effect, the treatment effects within each strata are estimated by Hodges Lehmann estimator. And the treatment effect within  $k^{th}$  stratum is defined as

$$\hat{\beta}_{kTrt} = \text{median} \left\{ \left( Y_{kTrt,i}^{align'} - Y_{kCtrl,j}^{align'} \right), i = 1, \dots, m; j = 1, \dots, n \right\}$$

Where  $Y_{kTrt,i}$  is the response in treatment group in the  $k^{th}$  stratum and  $Y_{kCtrl,j}$  is the response in control group in the  $k^{th}$  stratum.  $m$  and  $n$  denote the number of subjects in treatment group and control group in the  $k^{th}$  stratum correspondingly.

The overall effect is estimated by weighted average of strata effect.

$$\hat{\beta}_{Trt} = \frac{\sum_{k=1}^5 W_k \hat{\beta}_{kTrt}}{\sum_{k=1}^5 W_k}$$

The weight  $w_k = (\sigma_k^2)^{-1}$ , where  $\hat{\sigma}_k^2 = \left( \frac{W^{(n_{k1}n_{k2}+1-c)} - W^{(c)}}{2 \cdot Z_{(1-\alpha/2)}} \right)^2$  and  $(W^{(c)}, W^{(n_{k1}n_{k2}+1-c)})$  is the  $(1 - \alpha)$  symmetric two sided confidence interval of ordered pairs of difference in Hodges Lehmann estimator (Byun et al., 2013).

**2.10. Adjusting covariate effect based on quintile stratification of the covariate and align for strata effect; within strata treatment effect by mean difference; inverse variance weighted overall treatment effect (Aligned-WS(mean,iv,x5))**

The first step is to stratify the continuous covariate based on its quintile. And similar to Aligned-WS (HL,iv,x5), the response variable is adjusted by alignment. The average treatment effect is estimated by mean difference for each stratum. For the  $k^{th}$  stratum, the treatment effect is defined as

$$\hat{\beta}_{kTrt} = \bar{Y}_{kTrt}^{align} - \bar{Y}_{kCtrl}^{align}$$

The overall effect is estimated by weighted average of strata effect.

$$\hat{\beta}_{Trt} = \frac{\sum_{k=1}^5 W_k \hat{\beta}_{kTrt}}{\sum_{k=1}^5 W_k}$$

The weight  $w_k = (\sigma_k^2)^{-1}$ , where  $\hat{\sigma}_k^2 = \frac{\widehat{var}_{kTrt}^2}{m} + \frac{\widehat{var}_{kCtrl}^2}{n}$ , where  $\widehat{var}_{kTrt}^2, \widehat{var}_{kCtrl}^2$  is the sample variance with stratum  $k$  for each treatment group and  $m, n$  is the number of subjects for each treatment group correspondingly.

**2.11. Adjusting covariate effect based on treatment matched quintile stratification of the covariate and align for strata effect; Hodges Lehmann estimate of within strata treatment effect; inverse variance weighted overall treatment effect (Aligned-WMS(HL,iv,x5))**

This is similar to Aligned-WS(HL,iv,x5) and the only difference is the continuous covariate is stratified based on its conditional quintile condition on the treatment assignment. Therefore, the number of subjects of both treatment groups within each stratum are equal.

**2.12. Adjusting covariate effect based on treatment matched quintile stratification of the covariate and align for strata effect; within strata treatment effect by mean difference; inverse variance weighted overall treatment effect (Aligned-WMS(mean,iv,x5))**

This approach is similar to Aligned-WS(mean,iv,x5) and the only difference is the continuous covariate is stratified based on its conditional quintile condition on the treatment assignment. There are equal number of subjects from both treatment group within each stratum.

**2.13. Adjusting covariate effect based on quintile stratification of the covariate; Hodges Lehmann estimate of within strata treatment effect; inverse variance weighted overall treatment effect (WS(HL,iv,x5))**

The first step is also to stratify the continuous covariate based on its quintile. The treatment effects within each strata are estimated by Hodges Lehmann estimator. And the treatment effect within  $k^{\text{th}}$  stratum is defined as

$$\hat{\beta}_{kTrt} = \text{median} \left\{ \left( Y_{kTrt,i}^{\text{align}'} - Y_{kCtrl,j}^{\text{align}'} \right), i = 1, \dots, m; j = 1, \dots, n \right\}$$

where  $Y_{kTrt,i}$  is the response in treatment group in the  $k^{\text{th}}$  stratum and  $Y_{kCtrl,j}$  is the response in control group in the  $k^{\text{th}}$  stratum.  $m$  and  $n$  denote the number of subjects in treatment group and control group in the  $k^{\text{th}}$  stratum correspondingly.

The overall effect is estimated by weighted average of strata effect.

$$\hat{\beta}_{Trt} = \frac{\sum_{k=1}^5 w_k \hat{\beta}_{kTrt}}{\sum_{k=1}^5 w_k}$$

The weight  $w_k = (\sigma_k^2)^{-1}$ , where  $\hat{\sigma}_k^2 = \left( \frac{W^{(n_{k1}n_{k2}+1-c)} - W^{(c)}}{2 * Z_{(1-\alpha/2)}} \right)^2$  and  $(W^{(c)}, W^{(n_{k1}n_{k2}+1-c)})$  is the  $(1 - \alpha)$  symmetric two-sided confidence interval of ordered pairs of difference in Hodges Lehmann estimator (Byun et al., 2013).

**2.14. Adjusting covariate effect based on quintile stratification of the covariate; within strata treatment effect by mean difference; inverse variance weighted overall treatment effect (WS(mean,iv,x5))**

The first step is to stratify the continuous covariate based on its quintile. The average treatment effect is estimated by mean difference for each stratum. For the  $k^{\text{th}}$  stratum, the treatment effect is defined as

$$\hat{\beta}_{kTrt} = \bar{Y}_{kTrt} - \bar{Y}_{kCtrl}$$

Similar to Aligned-WS(mean, iv, x5), The overall effect is estimated by weighted average of strata effect.

$$\hat{\beta}_{Trt} = \frac{\sum_{k=1}^5 w_k \hat{\beta}_{kTrt}}{\sum_{k=1}^5 w_k}$$

The weight  $w_k = (\sigma_k^2)^{-1}$ , where  $\hat{\sigma}_k^2 = \frac{\widehat{\text{var}}_{kTrt}^2}{m} + \frac{\widehat{\text{var}}_{kCtrl}^2}{n}$ , where  $\widehat{\text{var}}_{kTrt}^2, \widehat{\text{var}}_{kCtrl}^2$  is the sample variance with stratum  $k$  for each treatment group and  $m, n$  is the number of subjects for each treatment group correspondingly.

**2.15. Adjusting covariate effect based on treatment matched quintile stratification of the covariate; Hodges Lehmann estimate of within strata treatment effect; inverse variance weighted overall treatment effect (WMS(HL,iv,x5))**

This is similar to WS(HL,iv,x5) and the only difference is the continuous covariate is stratified based on its conditional quintile condition on the treatment assignment. Therefore, the number of subjects of both treatment groups within each stratum are equal.

**2.16. Adjusting covariate effect based on treatment matched quintile stratification of the covariate; within strata treatment effect by mean difference; inverse variance weighted overall treatment effect (WMS(mean,iv,x5))**

This approach is similar to WS(mean,iv,x5) and the only difference is the continuous covariate is stratified based on its conditional quintile condition on the treatment assignment. There are equal number of subjects from both treatment group within each stratum.

**2.17. Adjusting covariate effect based on alignment of strata effect of binary covariate; Hodges Lehmann estimate of treatment effect (AR)**

This approach is for binary covariate. The binary covariate is treated as stratum variable. Unlike the quintile stratified strata, the sample size within each stratum could be uneven. The covariate effect is adjusted by alignment within each strata and the strata effect is estimated by the Walsh average in AR(x5). And the aligned response is adjusted similarly,  $Y_i^{align} = Y_i - \hat{\delta}_i$ , where  $Y_i$  is subjects in the  $k^{th}$  strata. And the treatment effect is defined as

$$\hat{\beta}_{Trt} = median\left\{\left(Y_{Trt,i}^{align} - Y_{Ctrl,j}^{align}\right), i = 1, \dots, m; j = 1, \dots, n\right\}$$

$Y_{Trt,i}^{align}$  is the aligned response in treatment group and  $Y_{Ctrl,j}^{align}$  is the aligned response in the control group.

**2.18. Adjusting covariate effect based on alignment of strata effect of binary covariate; Hodges Lehmann estimate of within strata treatment effect; inverse variance weighted overall treatment effect (Aligned-WS(HL,iv))**

This is for binary covariate besides treatment variable. This is similar to Aligned-WS(HL,iv,x5). The only difference is the binary variable is treated as strata variable instead of quintile stratification of continuous variable.

**2.19. Adjusting covariate effect based on alignment of strata effect of binary covariate; within strata treatment effect by mean difference; inverse variance weighted overall treatment effect (Aligned-WS(mean,iv))**

This is for binary covariate besides treatment variable. This is similar to Aligned-WS(mean,iv,x5). The only difference is the binary variable is treated as strata variable instead of quintile stratification of continuous variable.

**2.20. Hodges Lehmann estimate of within strata treatment effect; inverse variance weighted overall treatment effect (WS(HL,iv))**

This is for binary covariate besides treatment variable. This is similar to WS(HL,iv,x5). The only difference is the binary variable is treated as strata variable instead of quintile stratification of continuous variable.

**2.21. Within strata treatment effect by mean difference; inverse variance weighted overall treatment effect (WS(mean,iv))**

This is for binary covariate besides treatment variable. This is similar to WS(mean,iv,x5). The only difference is the binary variable is treated as strata variable instead of quintile stratification of continuous variable.

### 3. Simulation settings

The motivation for this simulation study is to evaluate if covariate adjusting approaches based on their empirical mean square error (MSE) and coverage probability. **MSE** is defined as  $\widehat{\text{Var}}(\widehat{\beta}_{\text{Trt}}) + (\widehat{\beta}_{\text{Trt}} - \beta_{\text{Trt}})^2$ , where  $\beta_{\text{Trt}}$  is the true parameter of treatment effect and  $\widehat{\beta}_{\text{Trt}}$  is estimate of treatment effect by different approaches. And coverage probability is defined as rate that  $\beta_{\text{Trt}}$  is within calculated 95% confidence interval for the parameter. In this simulation, only single covariate scenarios will be evaluated in this study. Setting 1: one normal covariate and linear regression.

In the first simulation, one normal covariate is included for adjusting. The simulation is through following steps:

- (1) Simulate population of 10,000 subjects with outcome Y, treatment variable Trt, and covariate X. Trt follow Bernoulli(0.5) distribution and the covariate X follow standard normal distribution.
- (2) Both scenario with outlier and without outliers are simulated. In the scenarios of no outlier, error follow Normal(0,  $\frac{1}{16}$ ). In the scenarios of outliers, error follow Normal(0,  $\frac{1}{16}$ ) with 80% chance and follow Normal(0,  $\frac{10}{16}$ ) with 20% chance.
- (3) Compute outcome Y through true model  $Y_i = b_0 + b_1 \text{Trt}_i + b_2 X_i + e_i$ .  $b_0$  is set to 0.1.  $b_1$  is determined so that under alternative hypothesis the power is close to 0.8 for most tests when the covariate X are fully balanced.  $b_2$  is determined so that correlation coefficient between Y and X is 0.3 for scenario with moderate covariate effect.
- (4) Create new indicator variable indX so that if  $X_i \geq \text{median}(X)$ ,  $\text{indX}_i = 1$ ; otherwise,  $\text{indX}_i = 0$ .
- (5) Sample 200 subjects from the population. Each arm has 100 subjects. In the control arm,  $\Pr(\text{indX}_i = 1 | \text{Trt} = 0)$  is set to 0.5. Thus, in the control arm, there is 50% chance that covariate X is greater than true population median. In the treatment arm,  $\Pr(\text{indX}_i = 1 | \text{Trt} = 1)$  range from 0.5 to 0.95 with increment 0.05. Thus in the first scenario, treatment arm has 50% chance that covariate X is greater than true population median. In the second scenario, the treatment arm has 55% chance that covariate X is greater than true population median. In the next scenario, the probability increment is 5% more. And in the 10<sup>th</sup> scenario, the treatment arm has 95% chance that covariate X is greater than true population median. Thus, in the first scenario, the covariate X is fully balanced between the two arms. In the second scenario, the covariate X is slightly imbalanced and in treatment arm has more large values of X comparing to control arm. In the 10<sup>th</sup> scenario, the covariate X is extremely imbalanced.
- (6) Under each scenario, the baseline covariate t-test are computed for each iteration. For each covariate adjusting approaches for estimation of treatment effect, the empirical MSE and coverage probability is computed.

#### 3.1. Setting 2: One binary covariate and linear regression

In the second simulation, one binary covariate is included for adjusting. The simulation is similar to Simulation 1 with following modifications:

- (1) In Step 1, the covariate X follow Bernoulli(0.5).
- (2) Skip Step 4. No indicator was generated.
- (3) In Step 5, In the control arm,  $\Pr(X_i = 1 | \text{Trt} = 0)$  is set to 0.5. In the treatment arm,  $\Pr(X_i = 1 | \text{Trt} = 1)$  range from 0.5 to 0.95 with increment 0.05.
- (4) In Step 6, the baseline covariate Chi-square test are computed for each iteration.

#### 3.2. Setting 3: One normal covariate and exponential regression

In the third simulation, one normal covariate is included for adjusting. The simulation is similar to Simulation 1 with difference in the true model. The true model is  $Y_i = b_0 + \exp(b_1 \text{Trt}_i + b_2 X_i) + e_i$ .

#### 4. Simulation results

The empirical MSE of above approaches are plotted in Figure 1 and coverage probability is plotted in Figure 2. In the subfigures, the first digit of the figure indicating the MSE (1) or coverage probability (2) and the second digit indicates the three simulation settings. The third digit indicates various cases of covariate effects: (1) no covariate effect with no outlier; (2) moderate covariate effect with no outlier; (3) no covariate effect with outliers; and (4) moderate covariate effect with outliers. In each figures, the labels in x-axis represent scenarios of covariate imbalance. From the original of x-axis, '0.5 : 0.5' represent the scenario that covariate X has 50% chance greater than true population median in both control arm and treatment arm. This represents the covariate fully balanced scenario. '0.55 : 0.5' represent the scenario that covariate X has 55% chance greater than true population median in the treatment arm and 50% chance greater than true population median in control arm. The other labels represent the scenarios in the similar ways. '0.95 : 0.5' represent the scenario that covariate X has 95% chance greater than true population median in the treatment arm and 50% chance greater than true population median in the control arm. This represent the most extreme scenario of covariate imbalance. From left to right of x-axis in each figures, the covariate imbalance become more extreme. In the two sub-plots of each figure, the first plot (a) includes all the tests and second plot (b) show the selected tests in smaller scale.

#### 4.1. The mean square error

##### 4.1.1. Results from setting 1 (normal covariate)

When there is no covariate effect and the covariate are fully balanced, the approaches estimating overall treatment effect based on weighted treatment effect within each stratum by nonparametric Hodges Lehmann estimator have slightly larger empirical MSE comparing to other approaches (Figure 1.1.1a). These approaches include Aligned-WS(HL,iv,x5), Aligned-WMS(HL,iv, x5), WS(HL,iv,x5), and WMS(HL,iv,x5)

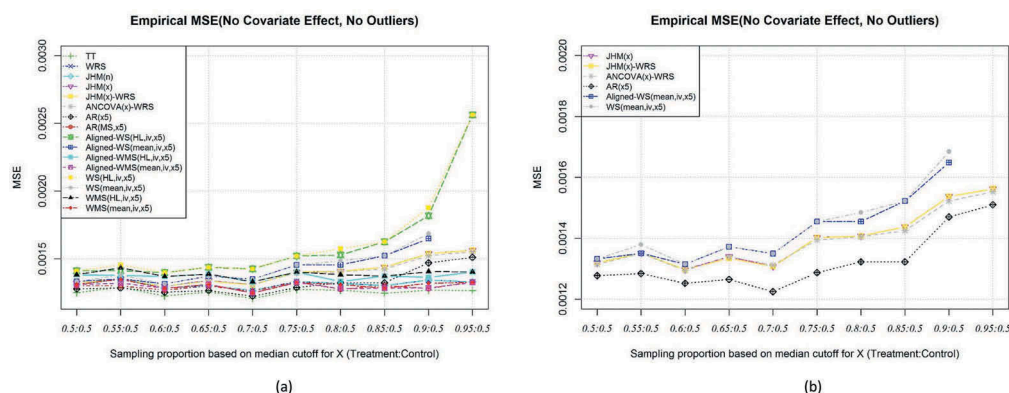
As covariate imbalance got more severe, the empirical MSE of approaches not involving covariate adjustment (TT, WRS, and JHM(n)) do not increase. The empirical MSE of approaches estimating overall treatment effect based on weighted treatment effect with matching will maintain at the covariate fully balanced level. These approaches includes Aligned-WMS(HL,iv,x5), Aligned-WMS (mean,iv,x5), WMS(HL,iv,x5), and WMS(mean,iv,x5). The empirical MSE of AR(MS,x5) does not increase as covariate imbalance gets severe since it adjusts covariate effect based on matched subjects for each stratum.

The empirical MSE of approaches estimating overall treatment effect based on weighted treatment effect without matching (Including Aligned-WS(HL,iv,x5), Aligned-WS(mean,iv,x5), WS(HL,iv, x5), and WS(mean,iv,x5)), increases slightly as the covariate imbalance get more severe. The other

**Figure 1.1.1.** Empirical MSE, true model:

$$Y = 0.1 + 0.1 * Trt + 0 * X + e,$$

where  $X \sim N(0, 1)$  and  $e \sim N(0, 0.063)$ .



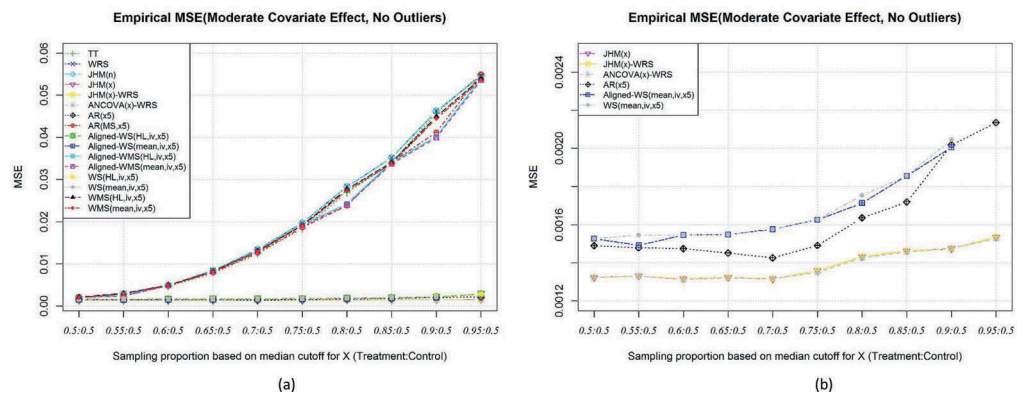
covariate adjusting approaches (including JHM(x), JHM(x)-WRS, ANCOVA(x)-WRS, and AR(x5)) will slightly increase empirical MSE when covariate imbalance get more severe (Figure 1.1.1b).

At presence of Moderate effect, the empirical MSE of approaches not involving covariate adjustment (TT, WRS, and JHM(n)) and the empirical MSE of approaches involve treatment group matched stratification (AR(MS,x5), Aligned-WMS(HL,iv,x5), Aligned-WMS(mean,iv,x5), WMS(HL,iv,x5), and WMS(mean,iv,x5)) increase dramatically as covariate imbalance get severe. (Figure 1.1.2a). The empirical MSE of approaches estimating overall treatment effect based on weighted treatment effect without matching (Including Aligned-WS(HL,iv,x5), Aligned-WS(mean,iv,x5), WS(HL,iv,x5), and WS(mean,iv,x5)) and JHM(x), JHM(x)-WRS, ANCOVA(x)-WRS, and AR(x5) is controlled as covariate imbalanced get severe. And JHM(x), JHM(x)-WRS, ANCOVA(x)-WRS has smallest empirical MSE across the scenarios of covariate imbalance (Figure 1.1.2b).

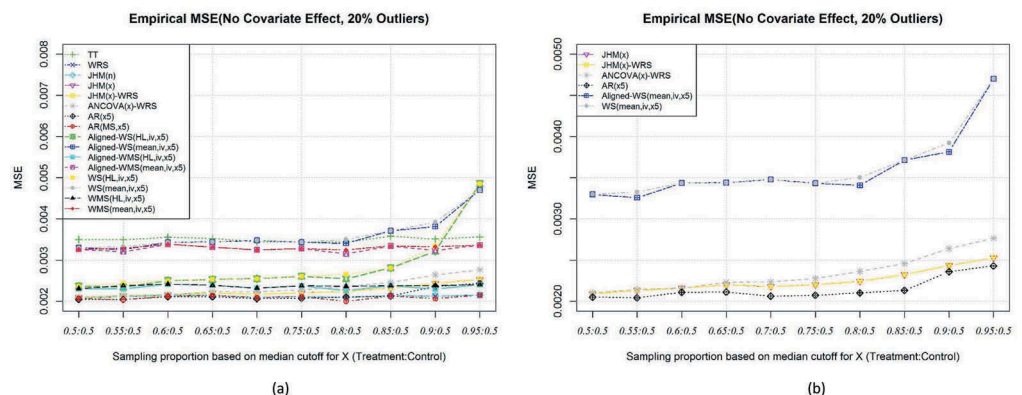
At presence of 20% outliers without covariate effect, the approaches involving mean differences (including TT, Aligned-WS(mean,iv,x5), Aligned-WMS(mean,iv,x5), WS(mean,iv,x5), and WMS(mean,iv,x5)) have largest empirical MSE comparing to other approaches when the covariate are fully balanced. Among the others, the empirical MSE of the approaches estimating overall treatment effect based on weighted treatment effect within each stratum by nonparametric Hodges Lehmann estimator (Aligned-WS(HL,iv,x5), Aligned-WMS(HL,iv,x5), WS(HL,iv,x5), and WMS(HL,iv,x5)) have slightly larger MSE (Figure 1.1.3a, Figure 1.1.3b).

As the covariate imbalance got more severe, the empirical MSE of approaches estimating overall treatment effect based on weighted treatment effect without matching (Including Aligned-WS(HL,iv,

**Figure 1.1.2.** Empirical MSE, true model:  
 $Y = 0.1 + 0.1 * Trt + 0.32 * X + e$ ,  
 where  $X \sim N(0, 1)$  and  $e \sim N(0, 0.063)$ .



**Figure 1.1.3.** Empirical MSE, true model:  
 $Y = 0.1 + 0.1 * Trt + 0 * X + e$ ,  
 where  $X \sim N(0, 1)$  and  $e \sim N(0, 0.063)$  with 80% or  $e \sim N(0, 0.63)$  with 20%.



x5), Aligned-WS(mean,iv,x5), WS(HL,iv,x5), and WS(mean,iv,x5)) will increase. The empirical MSE of JHM(x), JHM(x)-WRS, ANCOVA(x)-WRS, and AR(x5) will increase slightly as covariate imbalance get severe. While the empirical MSE of approaches not involving covariate adjustment (TT, WRS, and JHM(n)) and approaches estimating overall treatment effect based on weighted treatment effect with matching (Aligned-WMS(HL,iv,x5), Aligned-WMS(mean,iv,x5), WMS(HL,iv,x5), and WMS(mean,iv,x5)) will maintain at the same level as covariate imbalance increase (Figure 1.1.3a, Figure 1.1.3b).

At presence of outliers with moderate covariate effect, empirical MSE of approaches not involving covariate adjustment (TT, WRS, and JHM(n)) have higher MSE. Aligned-WS(HL,iv,x5), Aligned-WS(mean,iv,x5), WS(HL,iv,x5), WS(mean,iv,x5), JHM(x), JHM(x)-WRS, ANCOVA(x)-WRS, and AR(x5) controls the empirical MSE as covariate imbalance get severe and the empirical MSE of JHM(x), JHM(x)-WRS, ANCOVA(x)-WRS is the smallest (Figure 1.1.4a, Figure 1.1.4b).

4.1.2. Results from setting 2 (Binary covariate)

When there is no covariate effect and the covariate are fully balanced, the approaches involving mean differences (including TT, Aligned-WS(mean,iv), and WS(mean,iv)) has smallest empirical MSE. As covariate imbalance get severe, the empirical MSE of approaches not involving covariate adjustment (TT, WRS, and JHM(n)) do not increase (Figure 1.2.1a, Figure 1.2.1b). The empirical MSE of approaches estimating overall treatment effect based on weighted treatment effect within each stratum by nonparametric Hodges Lehmann estimator (Including Aligned-WS(HL,iv), WS(HL,iv)) increases as covariate imbalance get severe. The empirical MSE of other approaches have slightly increase as covariate get imbalanced (Figure 1.2.1a and Figure 1.2.1b).

When there is true covariate effect, the empirical MSE of approaches do no adjusting for the covariate (TT, WRS, JHM(n)) will increase dramatically as to covariate imbalance get severe (Figure

Figure 1.1.4. Empirical MSE, true model:

$$Y = 0.1 + 0.1 * Trt + 0.32 * X + e,$$

where  $X \sim N(0, 1)$  and  $e \sim N(0, 0.063)$  with 80% or  $e \sim N(0, 0.63)$  with 20%.

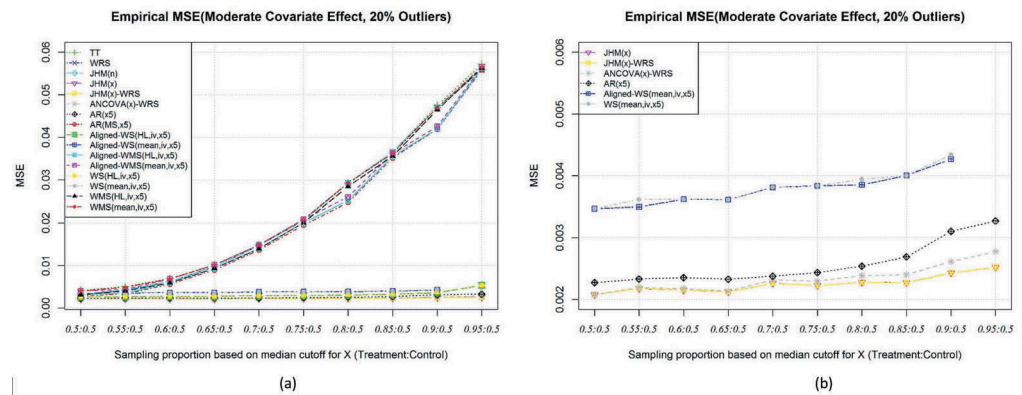
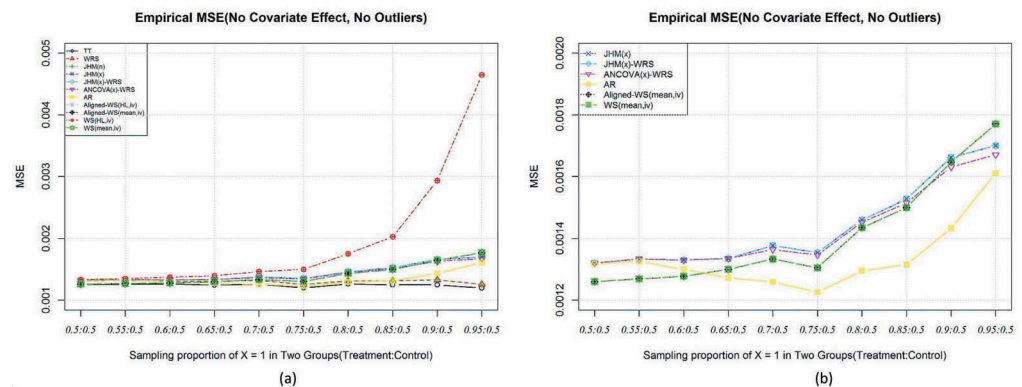


Figure 1.2.1. Empirical MSE, true model:

$$Y = 0.1 + 0.1 * Trt + 0 * X + e,$$

where  $X \sim Bern(0.5)$  and  $e \sim N(0, 0.063)$ .



1.2.2a). Among the approaches adjusting for the covariates effect, The empirical MSE of approaches estimating overall treatment effect based on weighted treatment effect within each stratum by nonparametric Hodges Lehmann estimator (Including Aligned-WS(HL,iv), WS(HL,iv)) increases as covariate imbalance get severe. The other approaches have similar Empirical MSE. As covariate get severely imbalanced, AR has lowest MSE (Figure 1.2.2b).

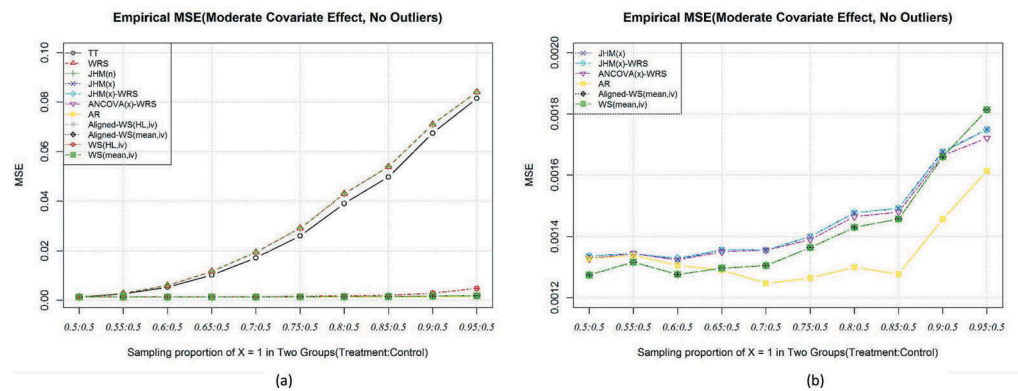
At presence of outliers with no covariate effect, the empirical MSE of approaches involving mean differences (including TT, Aligned-WS(mean,iv), and WS(mean,iv)) have smallest empirical MSE when covariate are balanced. Similarly to no outlier scenario, as the covariate imbalance get severe, the empirical MSE of Aligned-WS(HL,iv) and WS(HL,iv) increase (Figure 1.2.3a). And the empirical MSE of approaches not involving covariate adjustment (TT, WRS, and JHM(n)) do not increase. The empirical MSE of other approaches have slightly increase (Figure 1.2.3b).

At presence of 20% outliers with moderate covariate effect, the situation for MSE is similar as no outliers. The empirical MSE of approaches do no adjusting for the covariate (TT, WRS, JHM(n)) will increase dramatically as to covariate imbalance get severe (Figure 1.2.4a). The empirical MSE of Aligned-WS(HL,iv) and WS(HL,iv) increase most as the covariate get imbalanced among the rest. AR has the lowest MSE across the scenarios of covariate imbalance (Figure 1.2.4b).

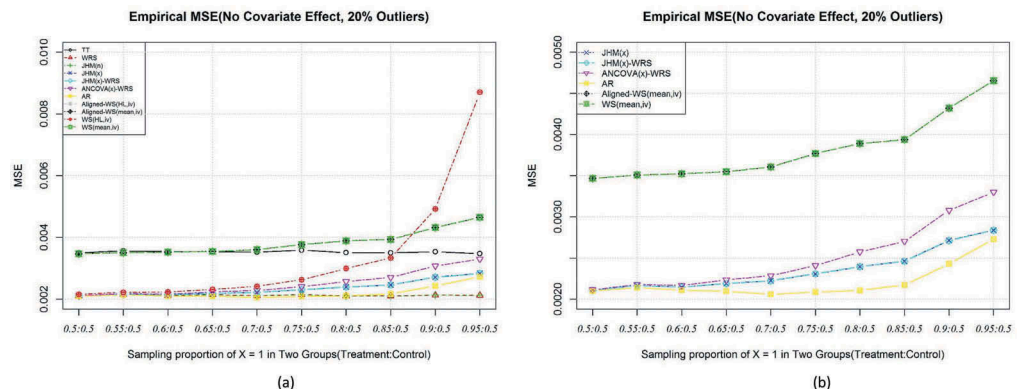
4.1.3. Results from setting 3 (exponential covariate)

When there is no covariates effect and the covariate is fully balanced, the empirical MSE of approaches estimating overall treatment effect based on weighted treatment effect within each

**Figure 1.2.2.** Empirical MSE, true model:  
 $Y = 0.1 + 0.1 * Trt + 0.63 * X + e$ ,  
 where  $X \sim \text{Bern}(0.5)$  and  $e \sim N(0, 0.063)$ .



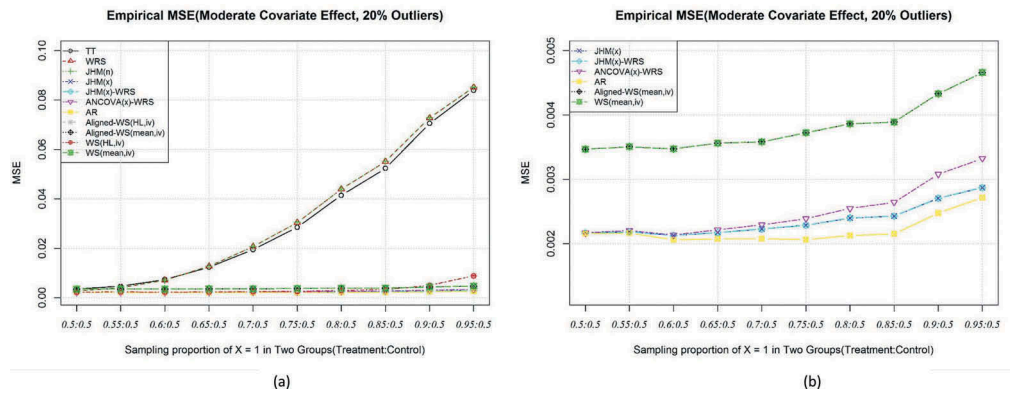
**Figure 1.2.3.** Empirical MSE, true model:  
 $Y = 0.1 + 0.1 * Trt + 0 * X + e$ ,  
 where  $X \sim \text{Bern}(0.5)$  and  $e \sim N(0, 0.063)$  with 80% or  $e \sim N(0, 0.63)$  with 20%.



**Figure 1.2.4.** Empirical MSE, true model:

$$Y = 0.1 + 0.1 * Trt + 0.63 * X + e,$$

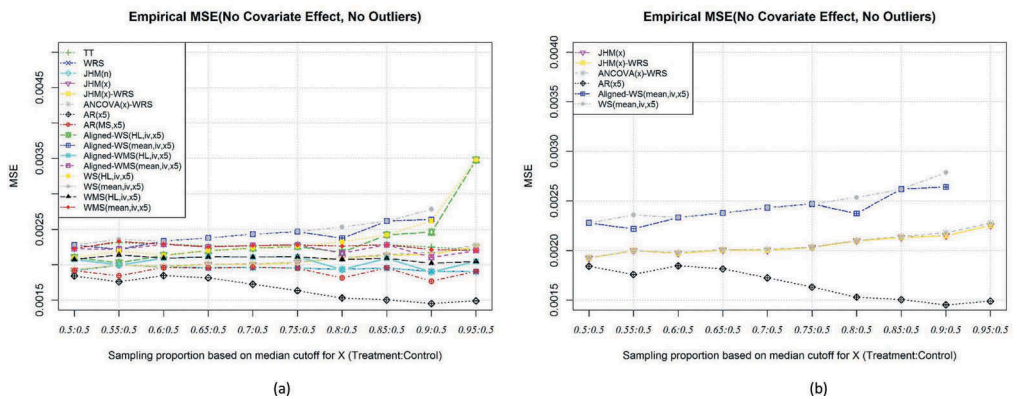
where  $X \sim \text{Bern}(0.5)$  and  $e \sim N(0, 0.063)$  with 80% or  $e \sim N(0, 0.63)$  with 20%.



**Figure 1.3.1.** Empirical MSE, true model:

$$Y = 0.1 + \exp(0.1 * Trt + 0 * X) + e,$$

where  $X \sim N(0, 1)$  and  $e \sim N(0, 0.063)$ .



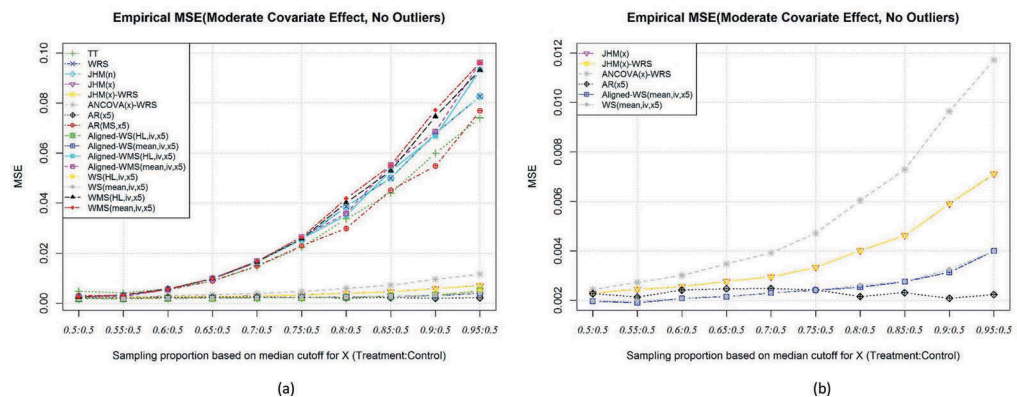
stratum by mean differences (Aligned-WS(mean,iv,x5), Aligned-WMS(mean,iv,x5), WS(mean,iv,x5), and WMS(mean,iv,x5)) are slightly larger than others (Figure 1.3.1a, Figure 1.3.1b). As covariate imbalance get severe, the empirical MSE of approaches estimating overall treatment effect based on weighted treatment effect without matching (Including Aligned-WS(HL,iv,x5), Aligned-WS(mean,iv,x5), WS(HL,iv,x5), and WS(mean,iv,x5)) increase (Figure 1.3.1a).

As covariate imbalance get severe, the empirical MSE of approaches involve matching within stratification (AR(MS,x5), Aligned-WMS(HL,iv,x5), Aligned-WMS(mean,iv,x5), WMS(HL,iv,x5), and

**Figure 1.3.2.** Empirical MSE, true model:

$$Y = 0.1 + \exp(0.1 * Trt + 0.34 * X) + e,$$

where  $X \sim N(0, 1)$  and  $e \sim N(0, 0.063)$ .



WMS(mean,iv,x5) and the approaches do not adjust for covariates (TT, WRS, and JHM(n)) increase dramatically (Figure 1.3.2a, Figure 1.3.2b). Among the others, the empirical MSE of rank regression based approaches (JHM(x) and JHM(x)-WRS) and ANCOVA-WRS are slightly larger as covariate imbalance get severe.

At presence of 20% outliers with no covariate effect, the situation of empirical MSE are similar as the empirical MSE of Aligned-WS(mean,iv,x5), Aligned-WMS(mean,iv,x5), WS(mean,iv,x5), and WMS(mean,iv,x5) are slightly larger (Figure 1.3.3a, Figure 1.3.3b).

At presence of 20% outlier with moderate covariate effect, the empirical MSE of approaches involve matching within stratification and approaches do not adjust for covariate effect increase dramatically as covariate imbalance get severe (Figure 1.3.4a). Among the remaining approaches, AR(x5) maintain the empirical MSE as covariate imbalance get severe (Figure 1.3.4a, Figure 1.3.4b).

**4.2. The coverage probability**

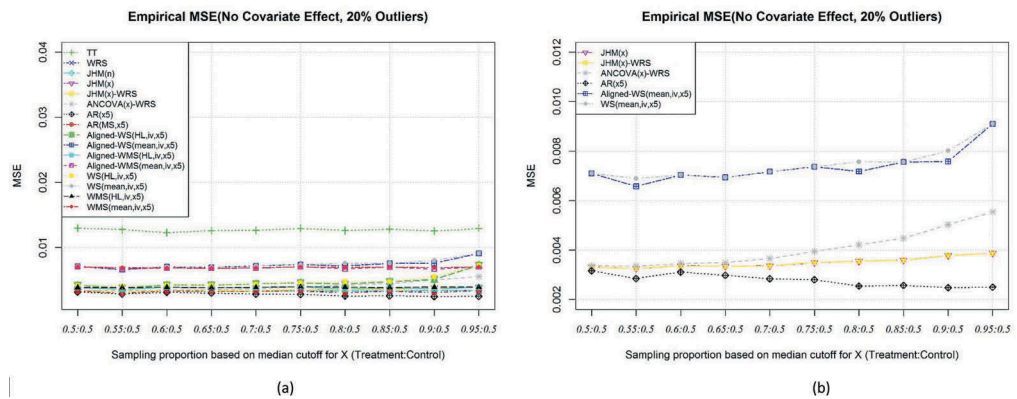
**4.2.1. Results from setting 1 (normal covariate)**

When there is no covariate effect and the covariate are fully balanced, the approaches estimating overall treatment effect based on weighted treatment effect within each stratum by nonparametric Hodges Lehmann estimator. These approaches include Aligned-WS(HL,iv,x5), Aligned-WMS(HL,iv,x5), WS(HL,iv,x5), and WMS(HL,iv,x5). The coverage probabilities for these approaches are also 6% lower than nominal level when covariate are fully balanced (Figure 2.1.1a). The coverage probability of Aligned-WMS(mean,iv,x5) and WMS(mean,iv,x5) are 1% slightly below 95% (Figure 2.1.1a, Figure 2.1.1b).

**Figure 1.3.3.** Empirical MSE, true model:

$$Y = 0.1 + \exp(0.1 * Trt + 0 * X) + e,$$

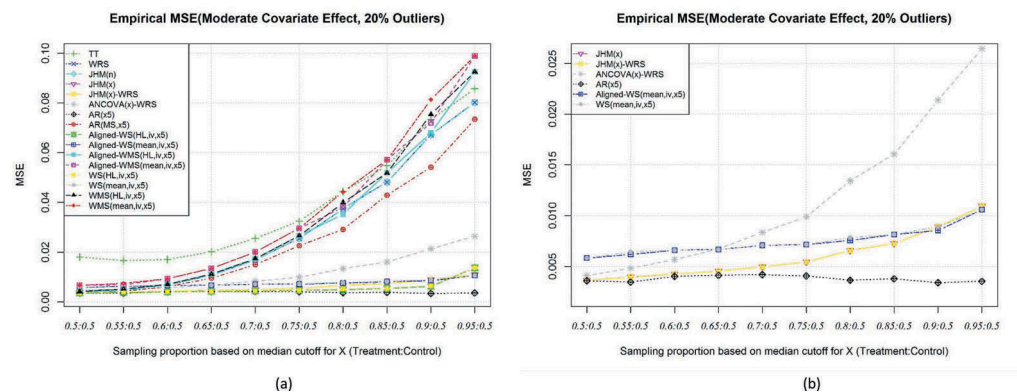
where  $X \sim N(0, 1)$  and  $e \sim N(0, 0.063)$  with 80% or  $e \sim N(0, 0.63)$  with 20%.



**Figure 1.3.4.** Empirical MSE, true model:

$$Y = 0.1 + \exp(0.1 * Trt + 0.34 * X) + e,$$

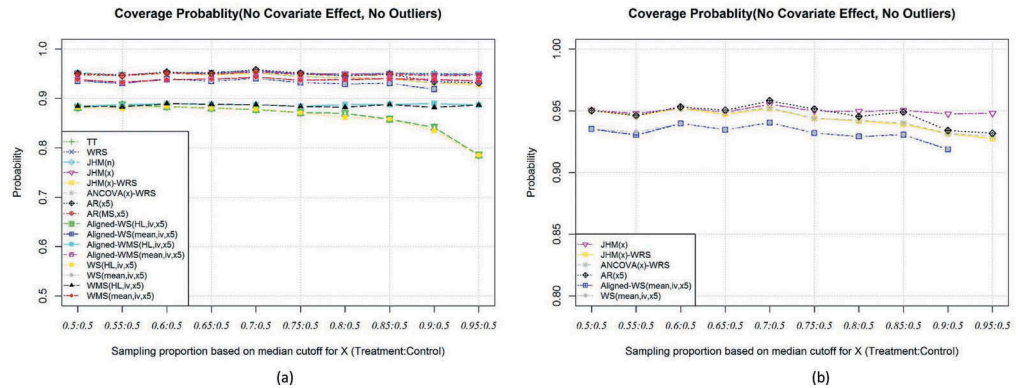
where  $X \sim N(0, 1)$  and  $e \sim N(0, 0.063)$  with 80% or  $e \sim N(0, 0.63)$  with 20%.



**Figure 2.1.1.** Coverage probability, true model:

$$Y = 0.1 + 0.1 * Trt + 0 * X + e,$$

where  $X \sim N(0, 1)$  and  $e \sim N(0, 0.063)$ .



The coverage probabilities for JHM(x), JHM(x)-WRS, ANCOVA(x)-WRS, and AR(x5) also decrease as covariate imbalance get more severe. Among them, Aligned-WS(HL,iv,x5) and WS(HL,iv,x5) decrease to less than 80%, while the coverage probability of other approaches is still above 90% (Figure 2.1.1a).

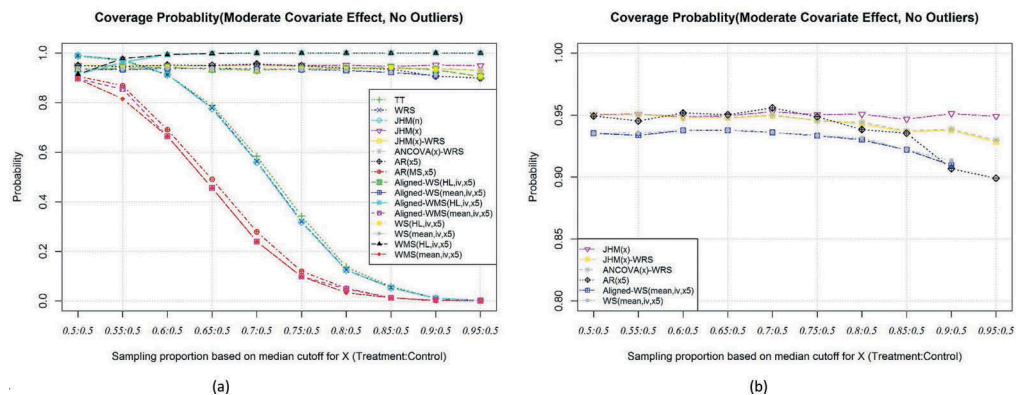
The coverage probability of approaches not involving covariate adjustment (TT, WRS, and JHM(n)) will decrease dramatically as covariate imbalance get severe (Figure 2.1.2a). The coverage probability of Aligned-WMS(mean,iv,x5), WMS(mean,iv,x5) and AR(MS, x5) also decrease as covariate imbalance get severe. The coverage probability of Aligned-WS(HL,iv,x5) and WMS(HL,iv,x5) increase to 1 as covariate get covariate imbalanced. This suggest matching will be problematic when there is covariate imbalance. JHM(x) maintain the coverage probability at 95% as covariate get imbalanced. For other approaches (including JHM(x)-WRS, ANCOVA(x)-WRS, AR(x5), Aligned-WS(HL,iv,x5), Aligned-WS(mean,iv,x5), WS(HL,iv,x5), and WS(mean,iv,x5)), the coverage probability is close to 95% and as covariate imbalance get severe, the coverage decrease slightly decrease but maintain at above 90% (Figure 2.1.2a, Figure 2.1.2b).

At presence of 20% outliers with no covariate effect, the approaches involving mean differences (including TT, Aligned-WS(mean,iv,x5), Aligned-WMS(mean,iv,x5), WS(mean,iv,x5), and WMS(mean,iv,x5)) have largest empirical MSE comparing to other approaches when the covariate are fully balanced. The coverage probability of Aligned-WS(mean,iv,x5), Aligned-WMS(mean,iv,x5), WS(mean,iv,x5), and WMS(mean,iv,x5) is about 94% which is very similar to the scenarios without outliers. The coverage probability of Aligned-WS(HL,iv,x5), Aligned-WMS(HL,iv,x5), WS(HL,iv,x5), and WMS(HL,iv,x5) is 98%–99% indicating these approaches are very conservative at presence of outliers (Figure 2.1.3a, Figure 2.1.3b). As the covariate imbalance get severe, the coverage

**Figure 2.1.2.** Coverage probability, true model:

$$Y = 0.1 + 0.1 * Trt + 0.32 * X + e,$$

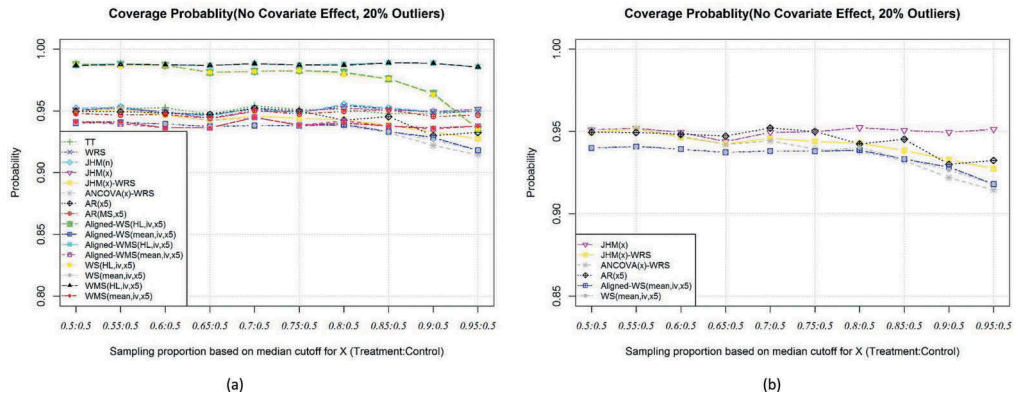
where  $X \sim N(0, 1)$  and  $e \sim N(0, 0.063)$ .



**Figure 2.1.3.** Coverage probability, true model:

$$Y = 0.1 + 0.13*Trt + 0*X + e,$$

where  $X \sim N(0, 1)$  and  $e \sim N(0, 0.063)$  with 80% or  $e \sim N(0, 0.63)$  with 20%.



probability of approaches estimating overall treatment effect based on weighted treatment effect without matching (Including Aligned-WS(HL,iv,x5), Aligned-WS(mean,iv,x5), WS(HL,iv,x5), and WS(mean,iv,x5)) decrease (Figure 2.1.3a, Figure 2.1.3b).

At presence of outliers, the coverage probability of different approaches behave similarly except for Aligned-WS(HL,iv,x5) and WS(HL,iv,x5). These two approaches estimating overall treatment effect based on weighted treatment effect within each stratum by nonparametric Hodges Lehmann estimator behave very conservative and the coverage probability is close to 1. This suggest these using Hodges Lehmann estimator to estimate treatment effect within each stratum may be conservative at presence of outliers if inverse variance is used as the weight (Figure 2.1.4a, Figure 2.1.4b).

4.2.2. Results from setting 2 (binary covariate)

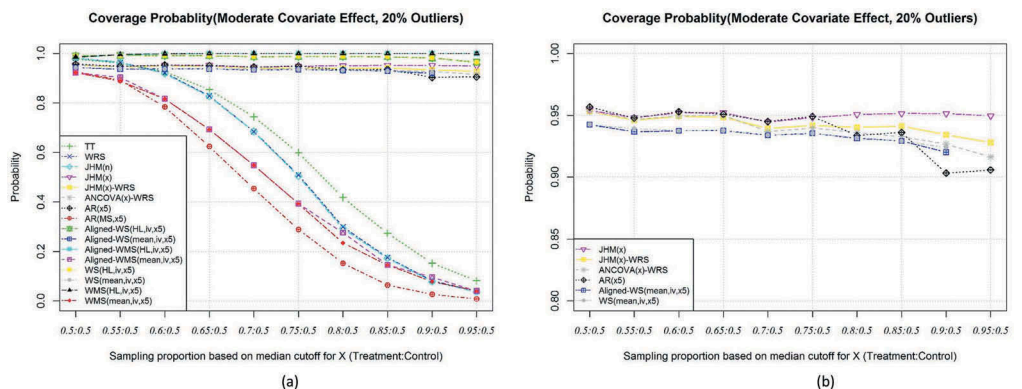
When there is no covariate effect and the covariate are fully balanced, coverage probability of estimates involving mean differences (including TT, Aligned-WS(mean,iv), and WS(mean,iv)) maintain around 95% (Figure 2.2.1a, Figure 2.2.1b). The coverage probability of weighted treatment effect within each stratum by nonparametric Hodges Lehmann estimator (Including Aligned-WS(HL,iv), WS(HL,iv)) are close to 1 (Figure 2.2.1a). The coverage probability for JHM(x) remains at 95% as covariate imbalance get severe. For other remaining approaches, the coverage probability decrease slightly as covariate imbalance get severe.

When there is true moderate covariate effect, the coverage probability of approaches that do not adjust for the covariate (TT, WRS, JHM(n)) decrease dramatically as covariate get more imbalanced

**Figure 2.1.4.** Coverage probability, true model:

$$Y = 0.1 + 0.13*Trt + 0.32*X + e,$$

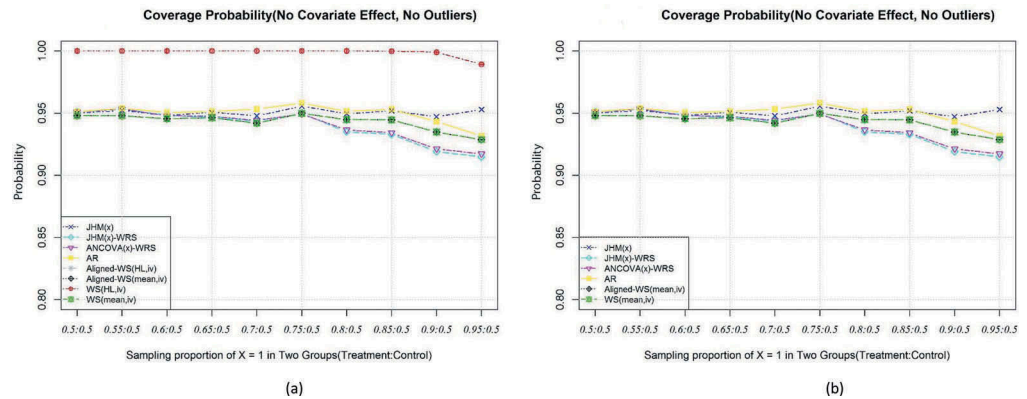
where  $X \sim N(0, 1)$  and  $e \sim N(0, 0.063)$  with 80% or  $e \sim N(0, 0.63)$  with 20%.



**Figure 2.2.1.** Coverage probability, true model:

$$Y = 0.1 + 0.1 * Trt + 0 * X + e,$$

where  $X \sim \text{Bern}(0.5)$  and  $e \sim N(0, 0.063)$ .



(Figure 2.2.a). Among the approaches adjusting for the covariates effect (including Aligned-WS (HL,iv), WS(HL,iv) increases as covariate imbalance get severe), the coverage probability of Aligned-WS(HL,iv), WS(HL,iv) remain at 1 (Figure 2.2.a). The coverage probability of JHM(x) remains at 95% across the covariate imbalance scenarios (Figure 2.2.b). For the other approaches, the coverage probability decreases slightly as covariate imbalance get severe.

At presence of outliers with no covariate effect, the coverage probability of TT, WRS, and JHM(n) keep around 95% as covariate imbalance get severe. The other approaches(AR, ANCOVA(x)-WRS, JHM(x)-WRS, WS(mean, iv), and Aligned-WS(mean,iv)) will have slight decrease as covariate get imbalanced (Figure 2.3.3a, Figure 2.3.3b).

At presence of 20% outliers with moderate covariate, the coverage probability is similar as no outliers. The coverage probability for approaches without adjusting for the covariate (TT, WRS, JHM(n)) decrease dramatically (Figure 2.2.4a). The coverage probability of Aligned-WS(HL,iv) and WS(HL,iv) remains at 1 as covariate imbalance get severe. The coverage probability of JHM(x) maintain at 95% as covariate get imbalance. The coverage probability of AR, Aligned-WS(mean,iv), and WS(mean, iv) slightly decrease as covariate imbalance get severe, while coverage probability of ANCOVA(x)-WRS and JHM(x)-WRS decrease slightly more when the covariate are extremely imbalanced (Figure 2.2.4a, Figure 2.2.4b).

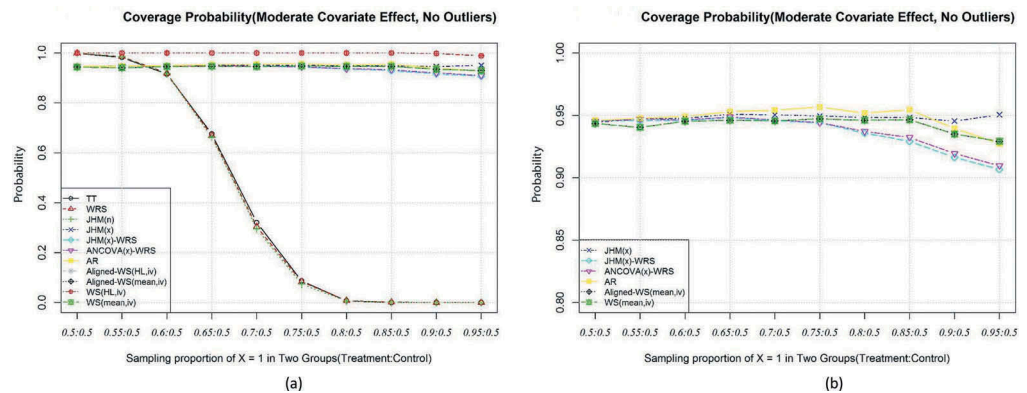
4.2.3. Results from setting 3 (exponential covariate)

When there is no covariates effect and the covariate is fully balanced, The coverage probability for approaches estimating overall treatment effect based on weighted treatment in effect within each

**Figure 2.2.2.** Coverage probability, true model:

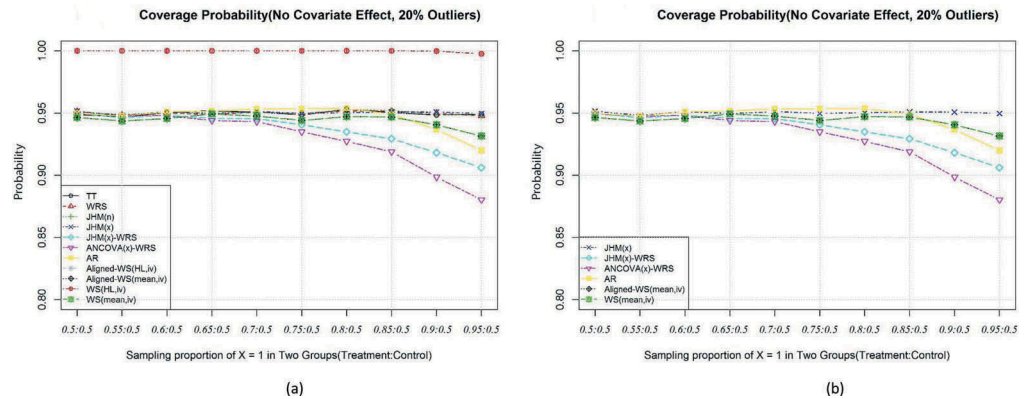
$$Y = 0.1 + 0.1 * Trt + 0.63 * X + e,$$

where  $X \sim \text{Bern}(0.5)$  and  $e \sim N(0, 0.063)$ .



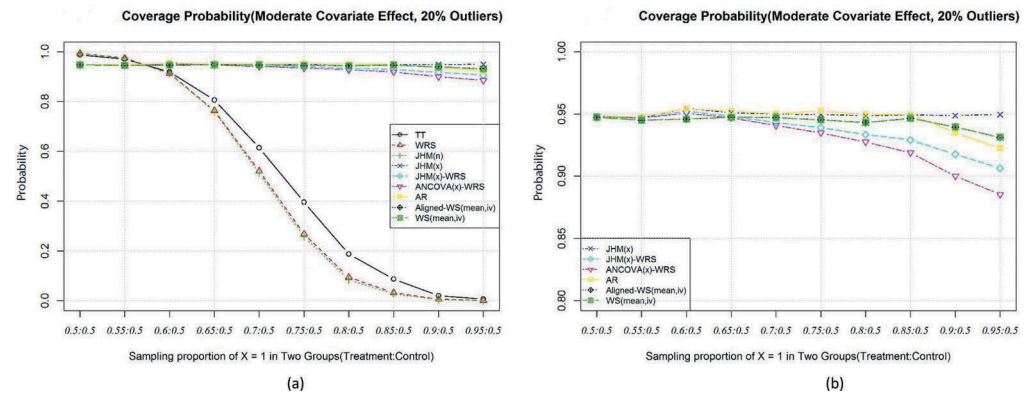
**Figure 2.2.3.** Coverage probability, true model:

$Y = 0.1 + 0.1 * Trt + 0 * X + e$ ,  
 where  $X \sim \text{Bern}(0.5)$  and  $e \sim N(0, 0.063)$  with 80% or  $e \sim N(0, 0.63)$  with 20%.



**Figure 2.2.4.** Coverage probability, true model:

$Y = 0.1 + 0.1 * Trt + 0.63 * X + e$ ,  
 where  $X \sim \text{Bern}(0.5)$  and  $e \sim N(0, 0.063)$  with 80% or  $e \sim N(0, 0.63)$  with 20%.



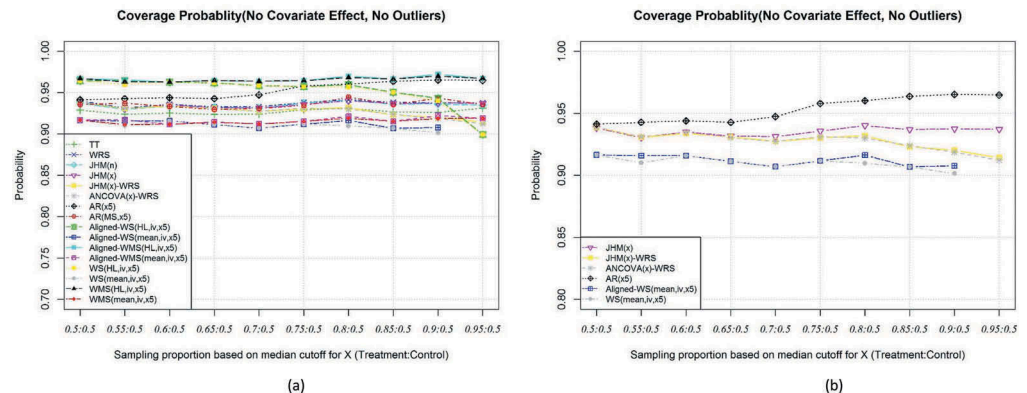
stratum by mean differences (Aligned-WS(mean,iv,x5), Aligned-WMS(mean,iv,x5), WS(mean,iv,x5), and WMS(mean,iv,x5)) about 92% across the scenarios (Figure 2.3.1a, Figure 2.3.1b).

When is moderate covariate effect, the coverage probability for approaches involve matching within stratification (AR(MS,x5), Aligned-WMS(HL,iv,x5), Aligned-WMS(mean,iv,x5), WMS(HL,iv,x5), and WMS(mean,iv,x5)) and the approaches do not adjust for covariates (TT, WRS, and JHM(n)) decrease dramatically (Figure 2.3.2a, Figure 2.3.2b).

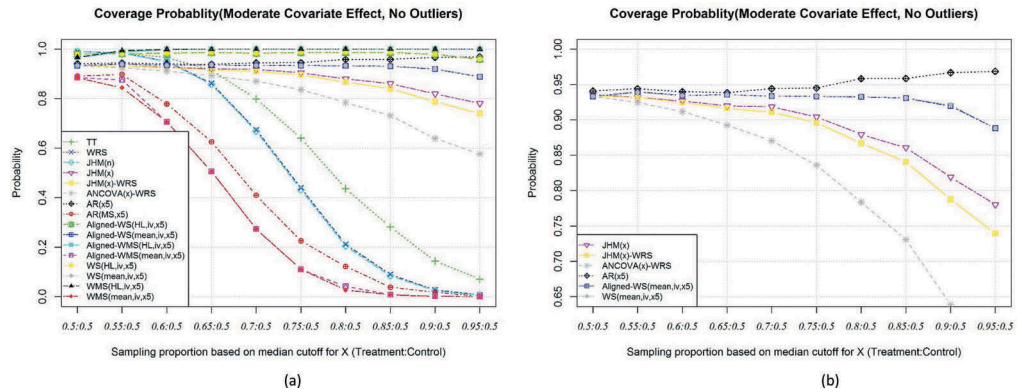
At presence of 20% outliers without covariate effect, the situation of empirical coverage probability are similar to that without outliers as the empirical coverage probability of Aligned-WS

**Figure 2.3.1.** Coverage probability, true model:

$Y = 0.1 + \exp(0.1 * Trt + 0 * X) + e$ ,  
 where  $X \sim N(0, 1)$  and  $e \sim N(0, 0.063)$ .



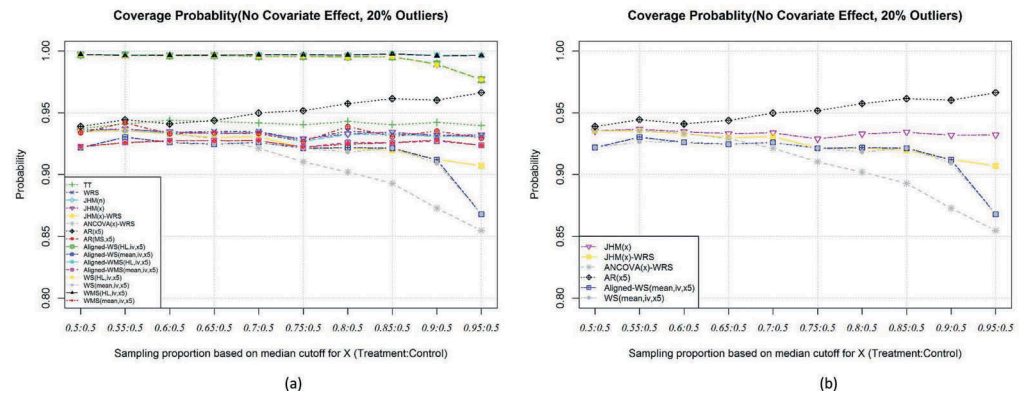
**Figure 2.3.2.** Coverage probability, true model:  
 $Y = 0.1 + \exp(0.1*Trt + 0.34*X) + e$   
 , where  $X \sim N(0, 1)$  and  $e \sim N(0, 0.063)$ .



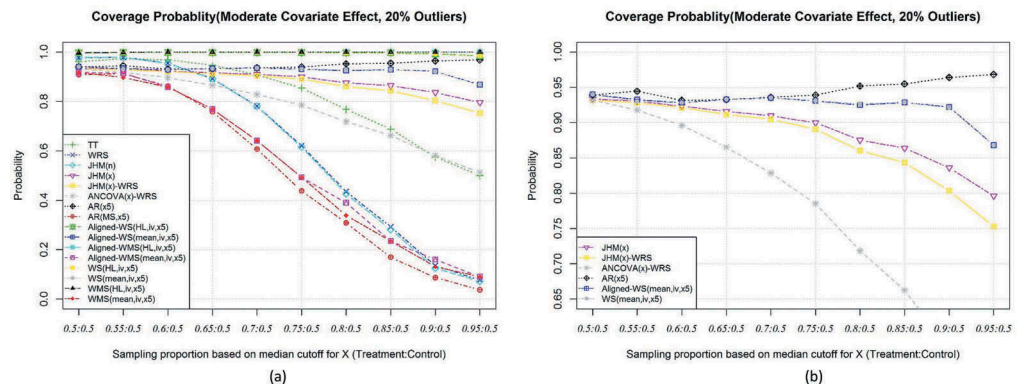
(mean,iv,x5), Aligned-WMS(mean,iv,x5), WS(mean,iv,x5), and WMS(mean,iv,x5) are slightly smaller (Figure 2.3.3a, Figure 2.3.3b).

At presence of 20% outliers with covariate effect, the empirical coverage probability of approaches involve matching within stratification and approaches without adjusting for covariate effect decrease dramatically as covariate imbalance get severe. Among the remaining approaches, AR(x5) maintains the empirical coverage probability as covariate imbalance get severe. Among these approaches, AR(x5) is the only approach have coverage probability close to 95% (Figure 2.3.4a, Figure 2.3.4b).

**Figure 2.3.3.** Coverage probability, true model:  
 $Y = 0.1 + \exp(0.1*Trt + 0*X) + e$   
 , where  $X \sim N(0, 1)$  and  $e \sim N(0, 0.063)$  with 80% or  $e \sim N(0, 0.63)$  with 20%.



**Figure 2.3.4.** Coverage probability, true model:  
 $Y = 0.1 + \exp(0.1*Trt + 0.34*X) + e$   
 , where  $X \sim N(0, 1)$  and  $e \sim N(0, 0.063)$  with 80% or  $e \sim N(0, 0.63)$  with 20%.



## 7. Concluding remarks

Covariate adjustment is necessary to estimate treatment effect when there is covariate imbalance. The approaches without adjustment for covariate effect have large MSE and low coverage probability at presence of covariate imbalance.

Some covariate adjustment approaches improve the estimation of treatment effect based on empirical MSE and coverage probability but not all of them. Approaches based on matching within stratification will fail to adjust the covariate effect when covariate imbalance exists. Matching will introduce extra bias when the covariate is imbalanced.

The approaches estimating overall treatment effect based on weighted treatment effect within each stratum by nonparametric Hodges Lehmann estimator do not work well either. At presence of outliers or covariate imbalance, these approaches are conservative and coverage probability are close to 1 for most scenarios.

ANCOVA-based approach do not have good estimation of treatment effect when the covariate imbalance is severe. It is likely the basic assumption for ANCOVA is violated at these extreme cases. The coverage probability tends to be much lower and empirical MSE is larger comparing to other covariate adjusting approaches.

Rank regression based approaches work well when the linearity assumption holds. But these approaches does not work well under some nonlinear scenario.

Among these approaches, alignment for covariate strata effect and then estimate treatment effect based on Hodges Lehmann approach seems perform well across all the scenarios in the simulation.

### Funding

This work was partially supported by the Cancer Prevention and Research Institute of Texas [RP170668].

### Author details

Jiabu Ye<sup>1</sup>

Dejian Lai<sup>2</sup>

E-mail: [Dejian.lai@uth.tmc.edu](mailto:Dejian.lai@uth.tmc.edu)

<sup>1</sup> AstraZeneca Pharmaceuticals, Gaithersburg, Maryland, 20854, USA.

<sup>2</sup> Department of Biostatistics and Data Science, The University of Texas School of Public Health, Houston, Texas, 77030, USA.

### Citation information

Cite this article as: Estimations of treatment effects based on covariate adjusted nonparametric methods, Jiabu Ye & Dejian Lai, *Cogent Mathematics & Statistics* (2020), 7: 1750878.

### References

Byun, J., Lai, D., Luo, S., Risser, J., Tung, B., & Hardy, R. J. (2013). A hybrid method in combining treatment effects from matched and unmatched studies.

*Statistics in Medicine*, 32(28), 4924–4937. <https://doi.org/10.1002/sim.5887>

Hettmansperger, T. P., & McKean, J. W. (1978). Statistical inference based on ranks. *Psychometrika*, 43, 69–79.

Hodges, J. L., & Lehmann, E. L. (1962). Rank methods for combination of independent experiments in analysis of variance. *The Annals of Mathematical Statistics*, 33(2), 482–497. <https://doi.org/10.1214/aoms/1177704575>

Hodges, J. L., Jr, & Lehmann, E. L. (1963). Estimates of location based on rank tests. *The Annals of Mathematical Statistics*, 34(2), 598–611. <https://doi.org/10.1214/aoms/1177704172>

Jaekel, L. A. (1972). Estimating regression coefficients by minimizing the dispersion of the residuals. *The Annals of Mathematical Statistics*, 43(5), 1449–1458. <https://doi.org/10.1214/aoms/1177692377>

Wilcoxon, F. (1945). Individual comparisons by ranking methods. *Biometrics Bulletin*, 1(6), 80–83. <https://doi.org/10.2307/3001968>

Ye, J., & Lai, D.: “Empirical power and Type I error of covariate adjusted nonparametric methods”. (2019). Submitted



© 2020 The Author(s). This open access article is distributed under a Creative Commons Attribution (CC-BY) 4.0 license.

You are free to:

Share — copy and redistribute the material in any medium or format.

Adapt — remix, transform, and build upon the material for any purpose, even commercially.

The licensor cannot revoke these freedoms as long as you follow the license terms.

Under the following terms:

Attribution — You must give appropriate credit, provide a link to the license, and indicate if changes were made.

You may do so in any reasonable manner, but not in any way that suggests the licensor endorses you or your use.

No additional restrictions

You may not apply legal terms or technological measures that legally restrict others from doing anything the license permits.



***Cogent Mathematics & Statistics* (ISSN: 2574-2558) is published by Cogent OA, part of Taylor & Francis Group.**

**Publishing with Cogent OA ensures:**

- Immediate, universal access to your article on publication
- High visibility and discoverability via the Cogent OA website as well as Taylor & Francis Online
- Download and citation statistics for your article
- Rapid online publication
- Input from, and dialog with, expert editors and editorial boards
- Retention of full copyright of your article
- Guaranteed legacy preservation of your article
- Discounts and waivers for authors in developing regions

**Submit your manuscript to a Cogent OA journal at [www.CogentOA.com](http://www.CogentOA.com)**

