Appendix C. PCATS Aim 1b: 2-Stage Treatment Assignment Results

PCATS Aim 1b: Two-stage treatment assignment results

Two-Stage Simulation Study Results

After making sure the performances of the GPMatch in the one-stage setting, we extended the GPMatch and the BART for the two-staged ATS. Five different simulation studies are performed. Since the primary goal of the study is to evaluate the performance of 2-stage BART and GPMatch for estimating the averaged treatment effort for ATS, four simulation designs were considered a sequential, multiple assignment, randomized trial (SMART), where the treatment is assigned at random at the 1st stage; the 2nd stage treatment is assigned adaptive to the patient's responses to the previous treatment. The design also considered heterogeneous treatment effect setting. Specifically, the five sets of simulation studies are:

1. A SMART trial with nonlinear outcome and treatment model.
2. A SMART trial with including an unmeasured binary confounder (U_0) at the baseline, and linear outcome and treatment model.
3. A SMART trial modified based on Kang and Schafer's dual misspecification setting.
4. A SMART trial with the 1st stage disease outcome modifies the 2nd stage treatment effects, a linear model setting.
5. Observational study subgroup treatment effect at both stages, a nonlinear setting.

In all setting, we consider the observed outcomes are measured without error. That is at the end of the 1st stage, the observed outcomes at the end of the 1st stage of treatment (L_1) is determined by

$$L_1 = A_0\, L_1(1) + \ (1 - A_0)\, L_1(0)$$

Then, at the end of 2nd stage, the observed outcome (Y) is determined by

$$Y = A_0\, A_1\, Y(1,1) + \ (1 - A_0)\, A_1\, Y(0,1) + A_0\, (1 - A_1)\, Y(1,0) + \ (1 - A_0)\, (1 - A_1)\, Y(0,0).$$

Without loss of generality, the simulation study considered a binary treatment decisions. The treatment assignment at the initial decision point (A_0) is determined by the baseline covariates (X). The treatment at the end of 1st stage (A_1) is assignment based on the observed patient's responses L_1, previous treatment assignment (A_0) and the baseline covariates (X), i.e. $[A_1|A_0, X, L_1]$. The potential outcomes at the end of 1st stage is determined by the baseline covariates (X). The potential outcomes at the end of 2nd stage is determined by the baseline covariates and the potential outcomes at end of the 1st stage, i.e. $[Y(00), Y(01), Y(10), Y(11)|X, L_1(0), L_1(1)]$. Under the ATS, there are two types of confounders - X denotes the baseline confounder, which may include age, gender, race, insurance status, baseline disease severity and health related quality of life; the L_1 is a time-varying confounder, which may include disease severity and health at the end of 1st stage. Of note, the L_1 may include the intermediate outcome measure, e.g. disease severity.

The recent paper by Newsome, Keogh and Daniel (2018) presented the most comprehensive review of six existing methods currently used for the evaluating causal treatment effects of ATS.

1. IPW: Invers probability weighting of marginal structural model, where the weight is the inverse of joint probability of treatment assignment $[A_0, A_1|X]$;
2. IPW(truncated): same as the IPW, with the PS truncated at 0.1 and 0.99;
3. HA-MSM: history-adjusted marginal structural models (MSM) $[Y(a_0, a_1)|X]$, where the weight is the inverse of joint time-varying PS $[A_0, A_1|X, L_1] = [A_0|X] \times [A_1|A_0, X, L_1]$.
4. HA-MSM(truncated): same as the HA-MSM, with the stabilized weight truncated at the 1st and 99th percentile;
5. G-computation: G-computatoin approaches impute the missing potential outcome similarly as the missing imputation chained equation (MICE), imputing $[Y(a_0, a_1)|X, L_1(a_0)]$ and $[L_1|X]$ using the observed outcomes and covariates.
6. G-estimation: g-estimation includes the PS estimates at each decision point into the structural nested model $[Y(a_0, a_1)|X, L_1(a_0)]$, and solve the corresponding estimating equation.

The authors reported persistent weaker performances of IPW method compared to the HA-MSM methods. Thus they are not considered in our study. The simulation studies compared the rest four methods against both 2-stage BART and the 2-stage GPMatch in our simulation studies.

For the 2-stage ATS (A_0, A_1), there are 4 causal treatment effect of ATS, $\{\tau_1, \psi_1, \psi_2, \psi_3\}$ which are defined by the six potential outcomes $(L_1(0), L_1(1), Y(00), Y(01), Y(10), Y(11))$:

$$L_1(a_0) = L_1(0) + \tau_1 a_0;$$
$$Y(a_0 a_1) = Y(00) + \psi_1 a_0(1 - a_1) + \psi_2 a_1(1 - a_0) + \psi_3 a_0 a_1.$$

Thus, the τ_1 is the 1st line treatment effect at the end of the 1st stage:

$$\tau_1 = L_1(1) - L_1(0);$$

The $\{\psi_1, \psi_2, \psi_3\}$ is the 2nd line treatment effect of ATS (a_0, a_1) at the end of 2nd stage. Thus the all pairwise comparisons of the four 2nd stage potential outcomes can be written as:

$$Y(10) - Y(00) = \psi_1;\ Y(01) - Y(00) = \psi_2;\ Y(11) - Y(00) = \psi_3;$$

$$Y(10) - Y(01) = \psi_2 - \psi_1; Y(11) - Y(01) = \psi_3 - \psi_1; Y(11) - Y(10) = \psi_3 - \psi_2.$$

The simulation results are summarized for each of the 7 causal contrast of ATS at both stages. All simulation results are summarized over 200 replicates. The root mean square error (RMSE) and median absolute error (MAE) are summarized over all replicates and plotted for all method of comparisons. For GPMath and BART, histogram of the posterior estimates are plotted.

1. A SMART Trial Nonlinear Model Setting

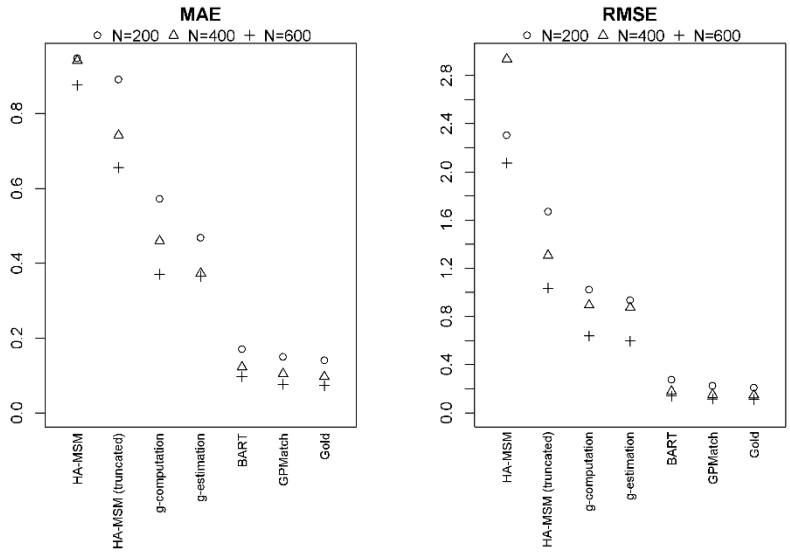
This simulation resembles a SMART trial setting, where the initial treatment is assigned at random. The stage 1 treatment has no effect on the disease progress at the end of 1st stage treatment. Both models of treatment assignment at the 2nd stage and the potential outcomes are nonlinear functions of the end of 1st stage responses L_1,

$$X \sim Bernoulli(0.4), \qquad A_0 \sim Bernoulli(0.5), \qquad L_1 \sim N(0,1)$$

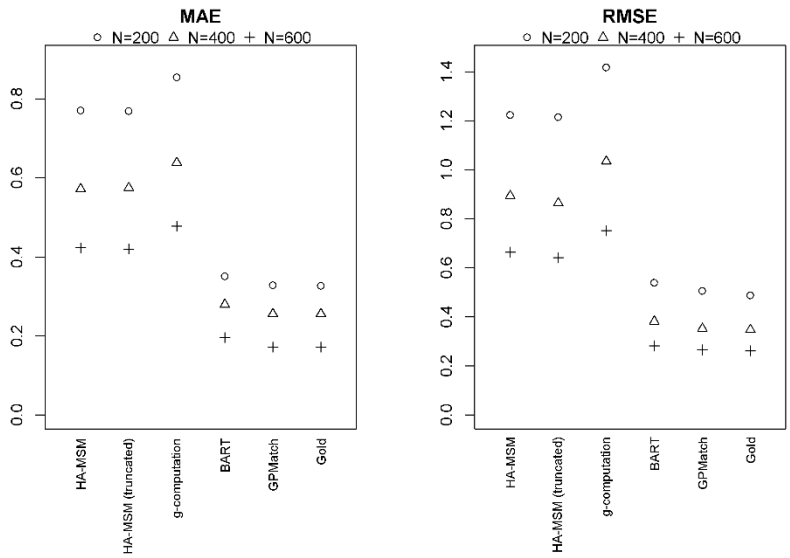
$$A_1 | L_1, A_0, X \sim Bernoulli(expit\left(0.2 - 0.2A_0 + L_1^{1/3}\right))$$

$$Y(a_0, a_1) \sim N(-2 + 2.5a_0 + 3.5a_1 + 0.5a_0a_1 - 3\exp(-L_1), sd = 1).$$

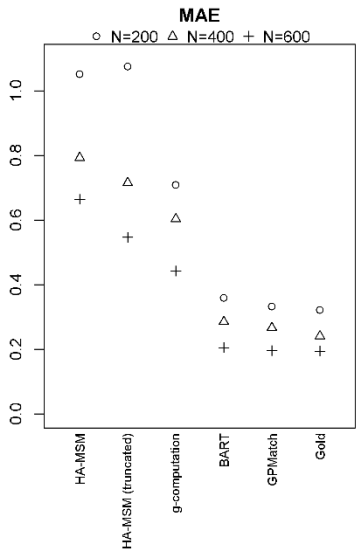
Y01-Y00

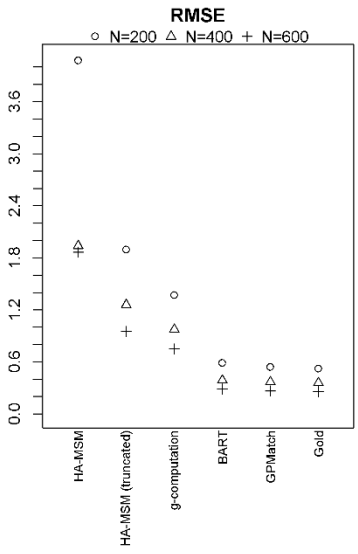


Y10-Y00

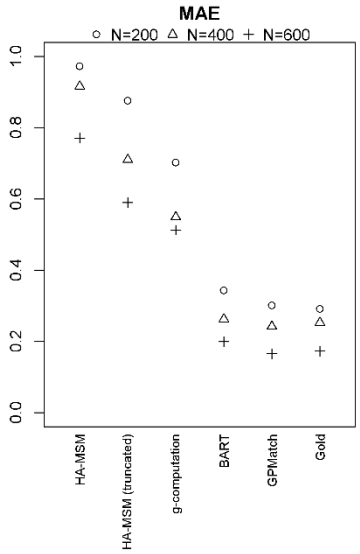


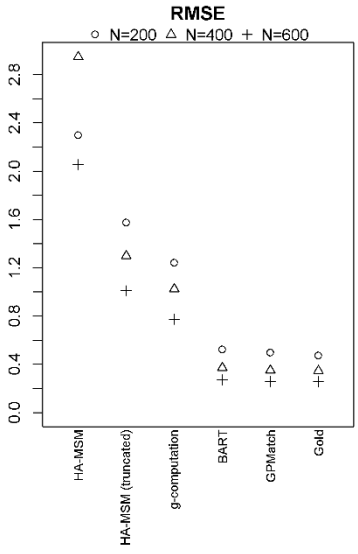
Y11-Y00



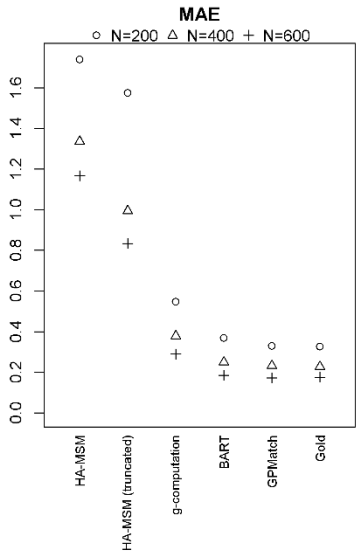


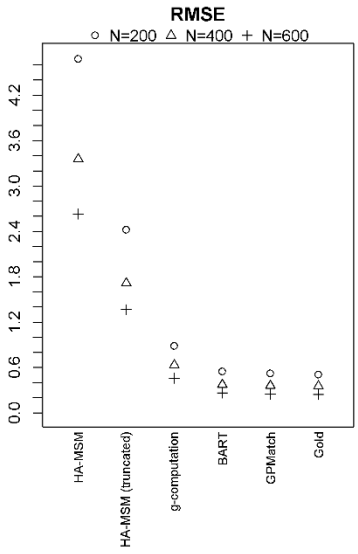
Y10-Y01



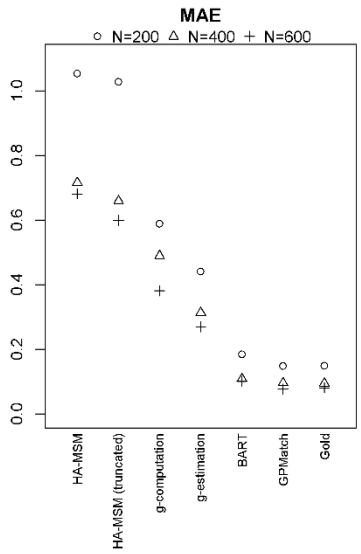


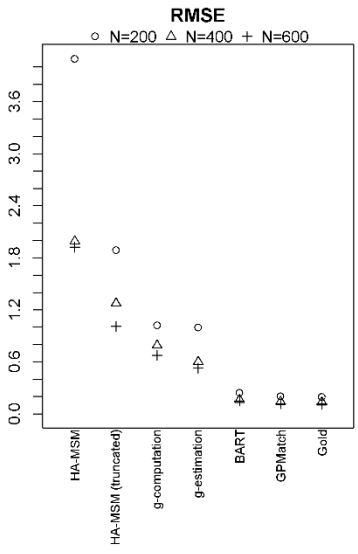
Y11-Y01

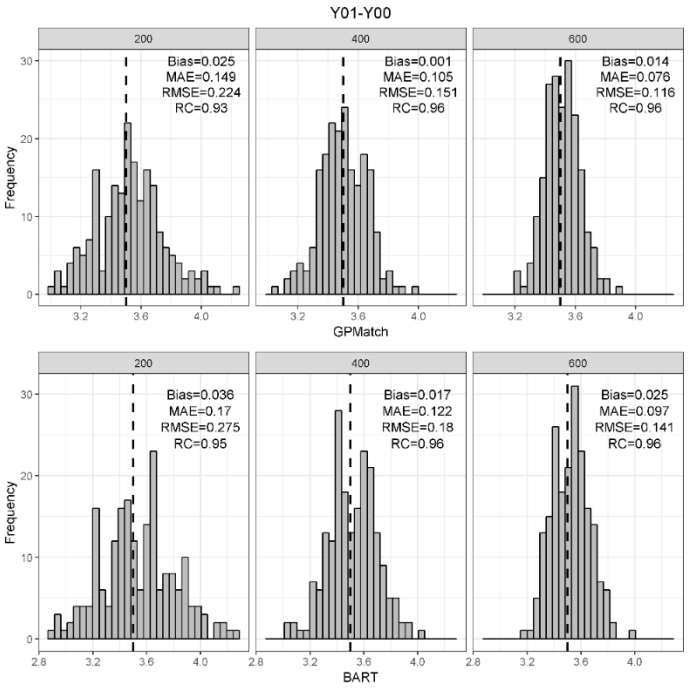


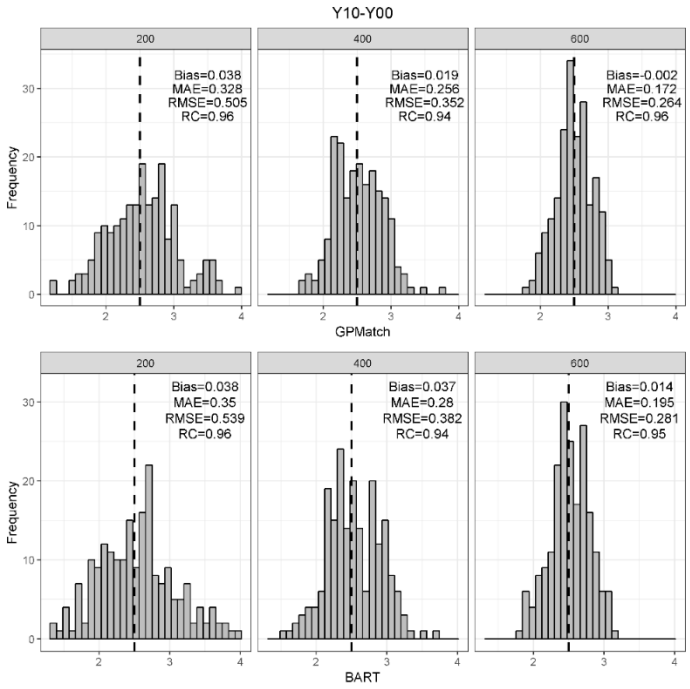


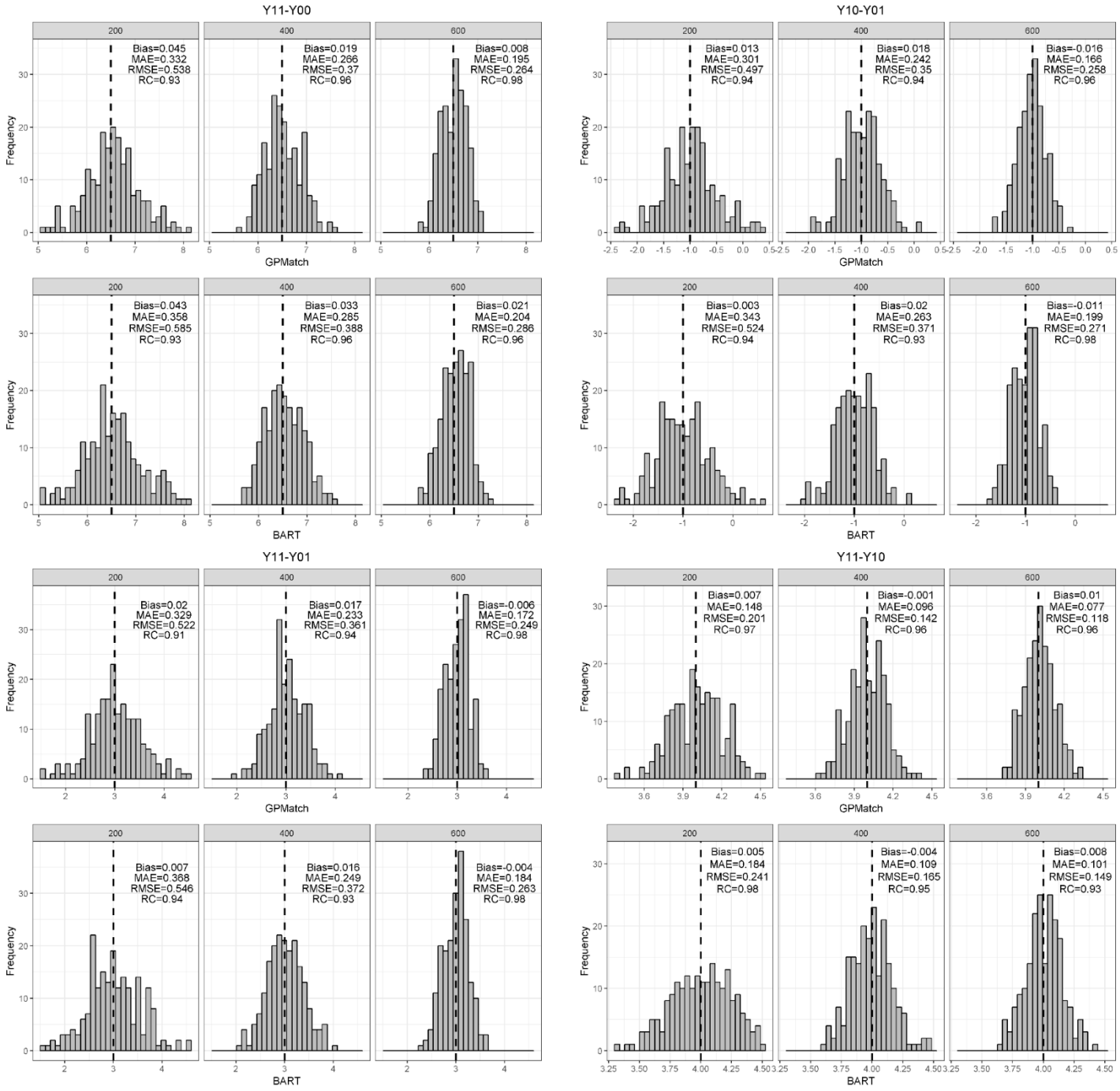
Y11-Y10

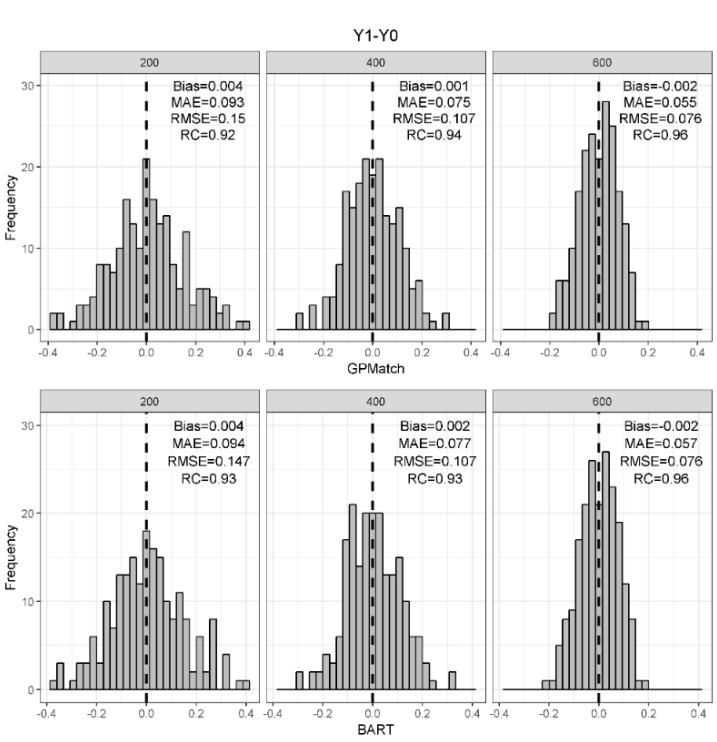

MAE
○ N=200 △ N=400 + N=600
0.0
0.2
0.4
0.6
0.8
1.0
HA-MSM
HA-MSM (truncated)
g-computation
g-estimation
BART
GPMatch
Gold

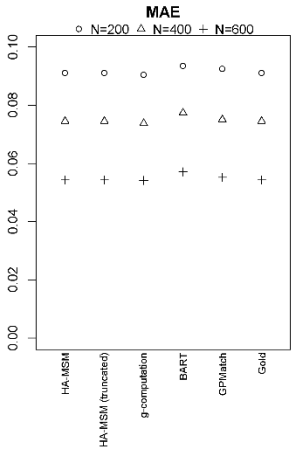

RMSE
○ N=200 △ N=400 + N=600
0.0
0.6
1.2
1.8
2.4
3.0
3.6
HA-MSM
HA-MSM (truncated)
g-computation
g-estimation
BART
GPMatch
Gold

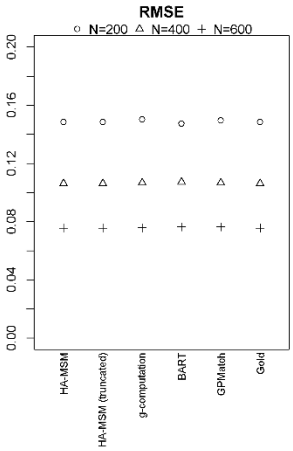

Y01-Y00
200
400
600
Bias=0.025
MAE=0.149
RMSE=0.224
RC=0.93
Bias=0.001
MAE=0.105
RMSE=0.151
RC=0.96
Bias=0.014
MAE=0.076
RMSE=0.116
RC=0.96
Frequency
GPMatch
Bias=0.036
MAE=0.17
RMSE=0.275
RC=0.95
Bias=0.017
MAE=0.122
RMSE=0.18
RC=0.96
Bias=0.025
MAE=0.097
RMSE=0.141
RC=0.96
BART


Y10-Y00
200
400
600
Bias=0.038
MAE=0.328
RMSE=0.505
RC=0.96
Bias=0.019
MAE=0.256
RMSE=0.352
RC=0.94
Bias=-0.002
MAE=0.172
RMSE=0.264
RC=0.96
Frequency
GPMatch
Bias=0.038
MAE=0.35
RMSE=0.539
RC=0.96
Bias=0.037
MAE=0.28
RMSE=0.382
RC=0.94
Bias=0.014
MAE=0.195
RMSE=0.281
RC=0.95
BART


Y11-Y00
200
400
600
Bias=0.045
MAE=0.332
RMSE=0.538
RC=0.93
Bias=0.019
MAE=0.266
RMSE=0.37
RC=0.96
Bias=0.008
MAE=0.195
RMSE=0.264
RC=0.98
Frequency
GPMatch
Y10-Y01
Bias=0.013
MAE=0.301
RMSE=0.497
RC=0.94
Bias=0.018
MAE=0.242
RMSE=0.35
RC=0.94
Bias=-0.016
MAE=0.166
RMSE=0.258
RC=0.96
Bias=0.043
MAE=0.358
RMSE=0.585
RC=0.93
Bias=0.033
MAE=0.285
RMSE=0.388
RC=0.96
Bias=0.021
MAE=0.204
RMSE=0.286
RC=0.96
BART
Bias=0.003
MAE=0.343
RMSE=0.524
RC=0.94
Bias=0.02
MAE=0.263
RMSE=0.371
RC=0.93
Bias=-0.011
MAE=0.199
RMSE=0.271
RC=0.98
Y11-Y01
Bias=0.02
MAE=0.329
RMSE=0.522
RC=0.91
Bias=0.017
MAE=0.233
RMSE=0.361
RC=0.94
Bias=-0.006
MAE=0.172
RMSE=0.249
RC=0.98
Y11-Y10
Bias=0.007
MAE=0.148
RMSE=0.201
RC=0.97
Bias=-0.001
MAE=0.096
RMSE=0.142
RC=0.96
Bias=0.01
MAE=0.077
RMSE=0.118
RC=0.96
Bias=0.007
MAE=0.368
RMSE=0.546
RC=0.94
Bias=0.016
MAE=0.249
RMSE=0.372
RC=0.93
Bias=-0.004
MAE=0.184
RMSE=0.263
RC=0.98
Bias=0.005
MAE=0.184
RMSE=0.241
RC=0.98
Bias=-0.004
MAE=0.109
RMSE=0.165
RC=0.95
Bias=0.008
MAE=0.101
RMSE=0.149
RC=0.93







2. A SMART Trial Linear Model, Unmeasured Covariate Setting

Following a setting used in Daniels (2013), this simulation considered an unmeasured confounder U_0. Specifically, the data are simulated according to the following setup:

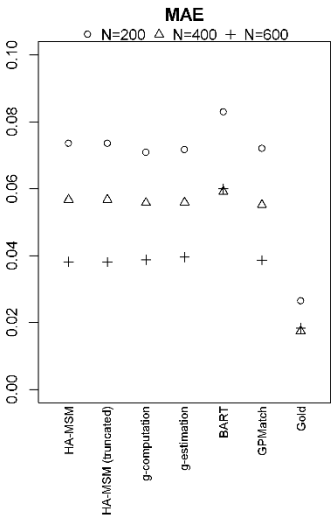
$$U_0 \sim Bernoulli(0.4)\,,$$
$$A_0 \sim Bernoulli(0.5),$$

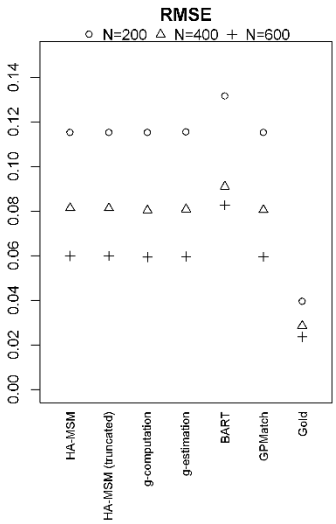
$$L_1(a_0) \sim Bernoulli(expit(0.25 + 0.3a_0 - 0.2U_0 - 0.05a_0U_0))$$

$$A_1 \sim Bernoulli(expit(0.4 + 0.5A_0 - 0.3L_1 - 0.4A_0L_1))$$

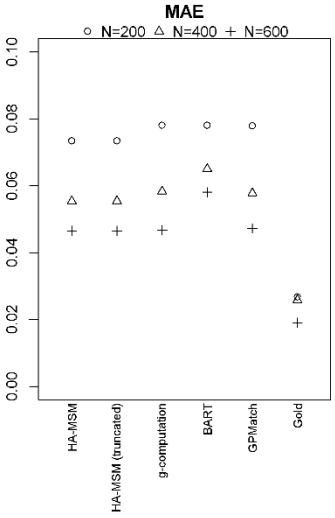
$$Y(a_0, a_1) \sim N(0.25 - 0.5a_0 - 0.75a_1 + 0.2a_0a_1 - U_0, 0.2)$$

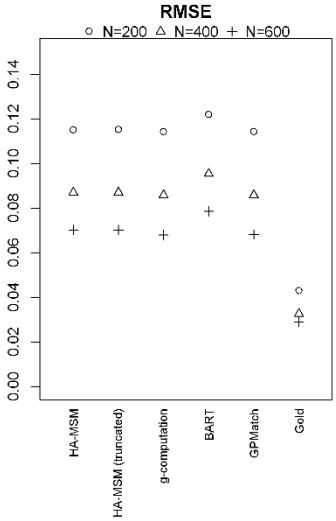
Y01-Y00



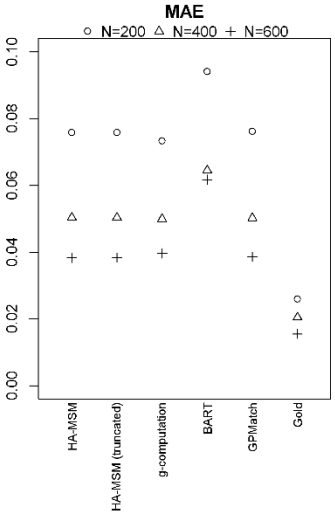


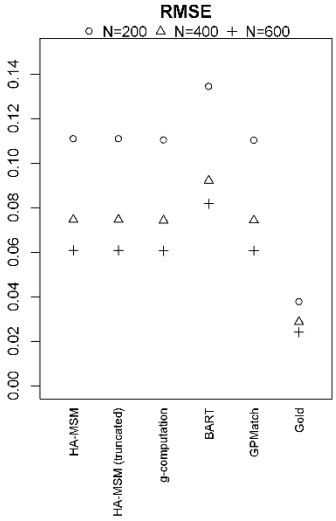
Y10-Y00



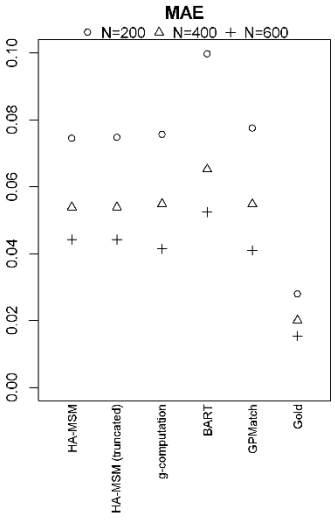


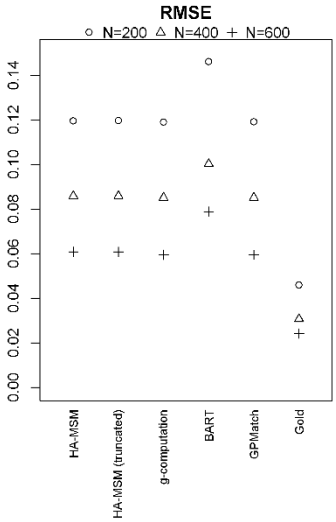
Y11-Y00



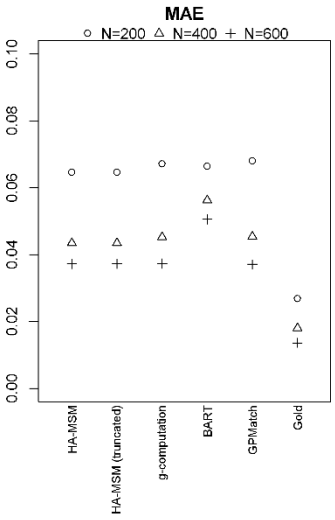


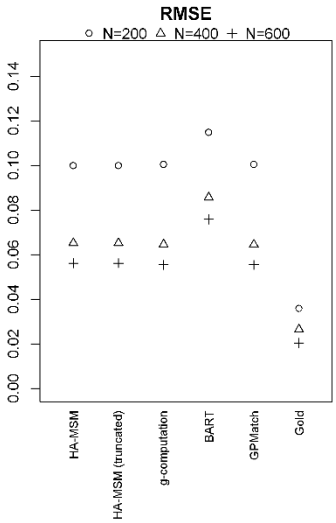
Y10-Y01



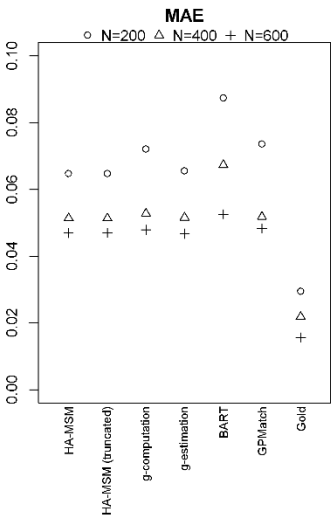


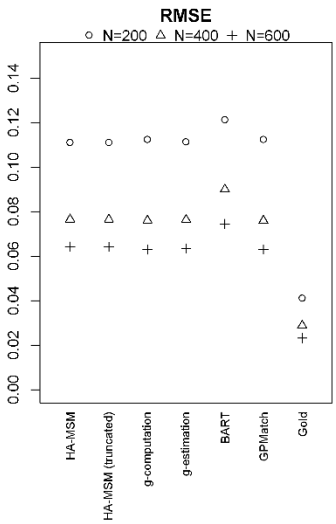
Y11-Y01

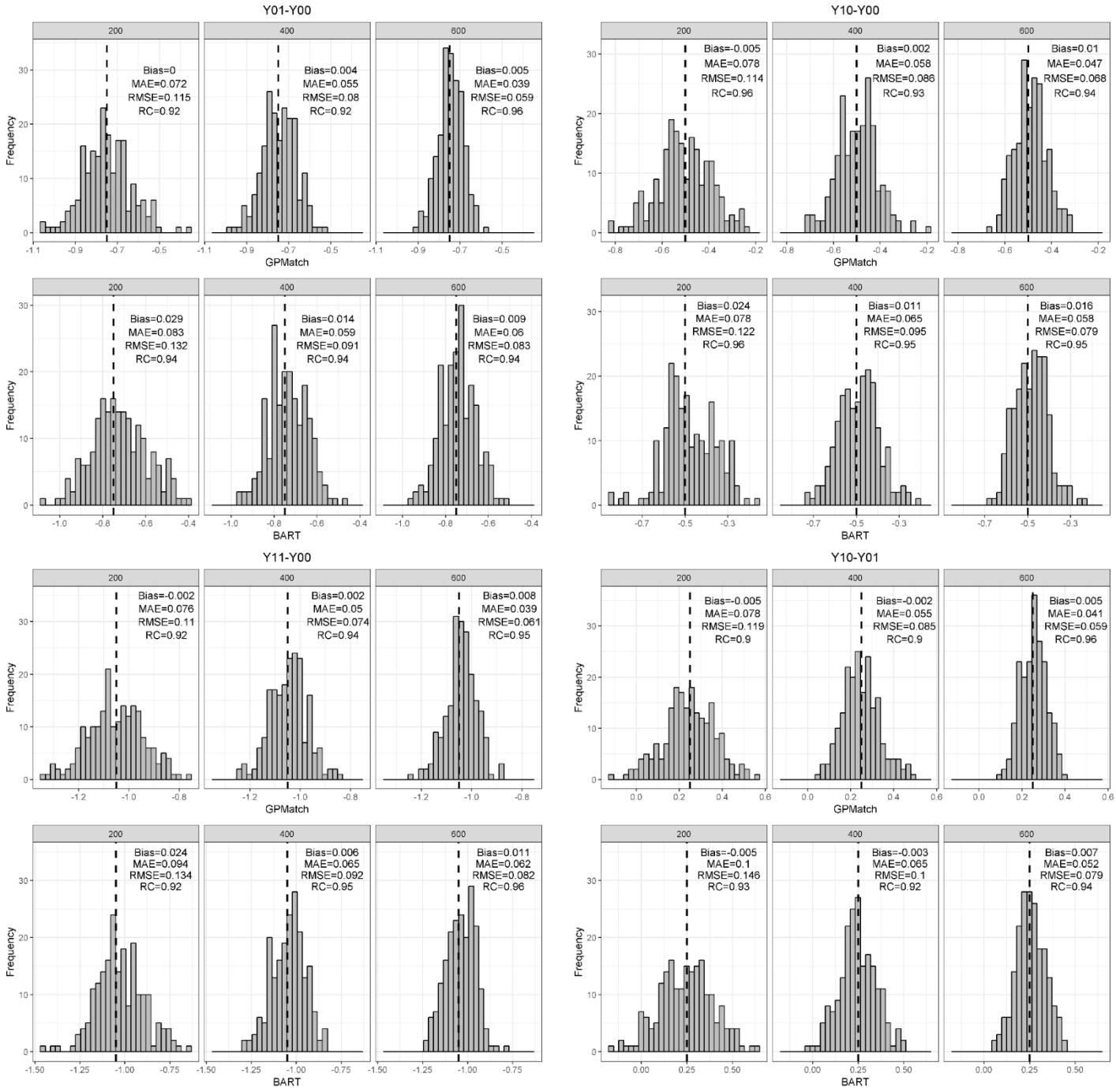


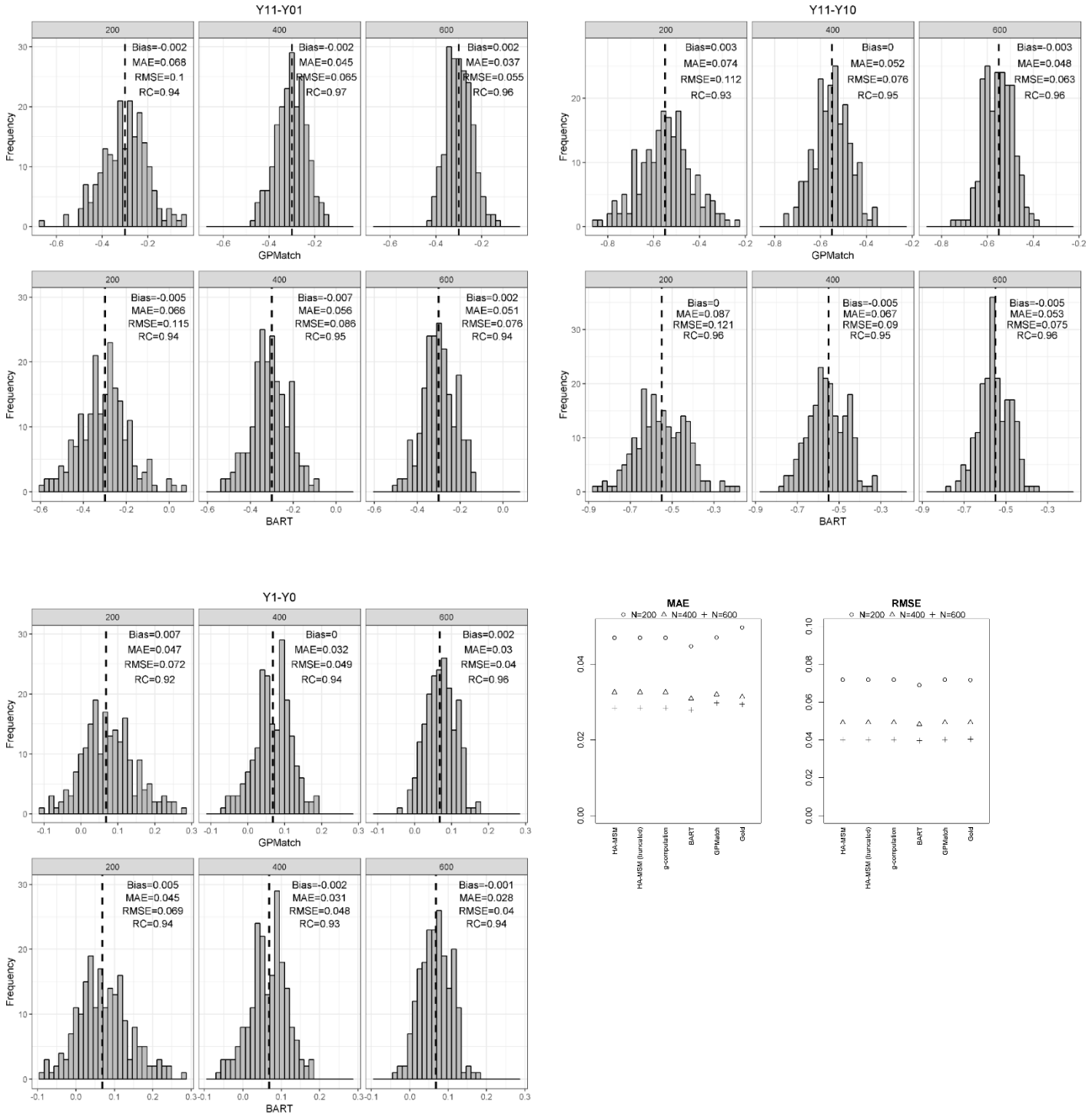


Y11-Y10






Y01-Y00
200
400
600
Bias=0
MAE=0.072
RMSE=0.115
RC=0.92
Bias=0.004
MAE=0.055
RMSE=0.08
RC=0.92
Bias=0.005
MAE=0.039
RMSE=0.059
RC=0.96
Frequency
GPMatch
Y10-Y00
Bias=-0.005
MAE=0.078
RMSE=0.114
RC=0.96
Bias=0.002
MAE=0.058
RMSE=0.086
RC=0.93
Bias=0.01
MAE=0.047
RMSE=0.068
RC=0.94
Bias=0.029
MAE=0.083
RMSE=0.132
RC=0.94
Bias=0.014
MAE=0.059
RMSE=0.091
RC=0.94
Bias=0.009
MAE=0.06
RMSE=0.083
RC=0.94
BART
Bias=0.024
MAE=0.078
RMSE=0.122
RC=0.96
Bias=0.011
MAE=0.065
RMSE=0.095
RC=0.95
Bias=0.016
MAE=0.058
RMSE=0.079
RC=0.95
Y11-Y00
Bias=-0.002
MAE=0.076
RMSE=0.11
RC=0.92
Bias=0.002
MAE=0.05
RMSE=0.074
RC=0.94
Bias=0.008
MAE=0.039
RMSE=0.061
RC=0.95
Y10-Y01
Bias=-0.005
MAE=0.078
RMSE=0.119
RC=0.9
Bias=-0.002
MAE=0.055
RMSE=0.085
RC=0.9
Bias=0.005
MAE=0.041
RMSE=0.059
RC=0.96
Bias=0.024
MAE=0.094
RMSE=0.134
RC=0.92
Bias=0.006
MAE=0.065
RMSE=0.092
RC=0.95
Bias=0.011
MAE=0.062
RMSE=0.082
RC=0.96
Bias=-0.005
MAE=0.1
RMSE=0.146
RC=0.93
Bias=-0.003
MAE=0.065
RMSE=0.1
RC=0.92
Bias=0.007
MAE=0.052
RMSE=0.079
RC=0.94



3. A SMART Trial Kang and Schafer Dual Misspecification Setting

We extended the well-known Kang and Schafer (2014) simulation to a 2-stage setting. Like the SMART trial, the 1st treatment is assigned at random, and the outcome at is a linear function of the baseline covariates $Z_1 - Z_4$.

$$Z_1, Z_2, Z_3, Z_4 \sim N(0,1),\ A_0 \sim Bernoulli(0.5)$$

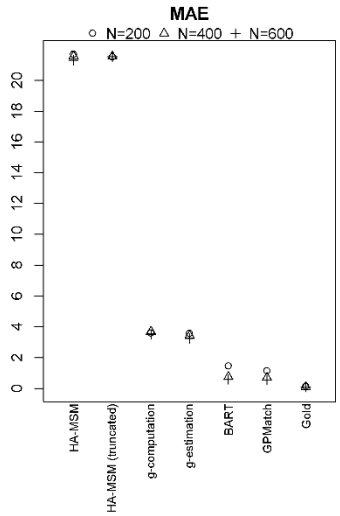
$$L_1 \sim N(-1.5 + 27.4Z_1 + 13.7Z_2 + 0 * Z_3 + 0 * Z_4 - 3A_0, 1)$$

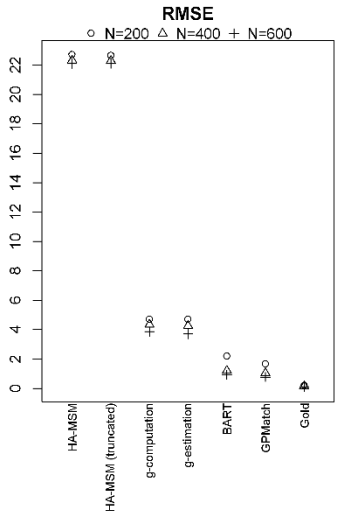
$$A_1 \sim Bernoulli(expit\left(0.25A_0 + (0.1) * L_1^{1/3} + 0.75 * (Z_1 - 0.5Z_2 + 0.25Z_3 + 0.1Z_4)\right))$$

$$Y(A_0, A_1) \sim N(210 + L_1 + 13.7Z_3 + 13.7Z_4 - 5A_0 - 3A_1 - 2A_0A_1, 1)$$

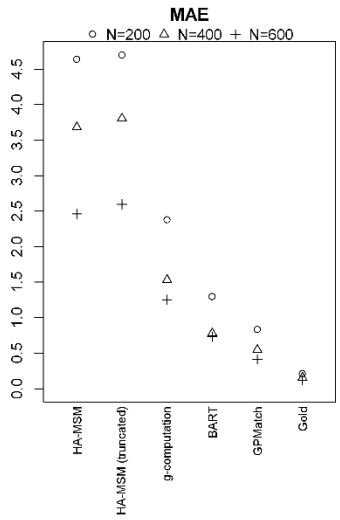
Like in Kang and Schafer, only transformed covariates are observed $x_1 = exp(z_1/2), x_2 = z_2/(1 + exp(z_1)) + 10, x_3 = \left(\frac{z_1 z_3}{25} + 0.6\right)^3$, and $x_4 = (z_2 + z_4 + 20)^2$.

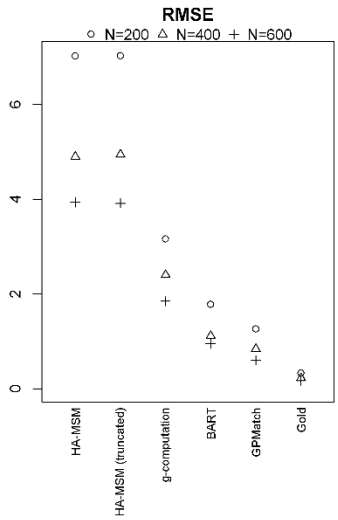
Y01-Y00



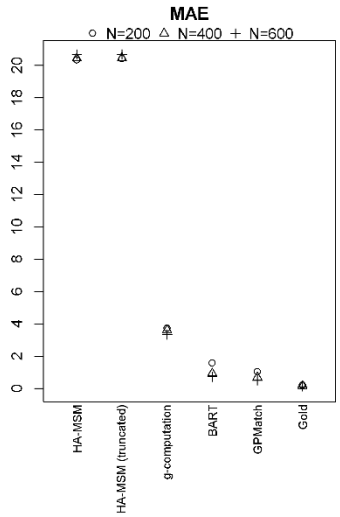


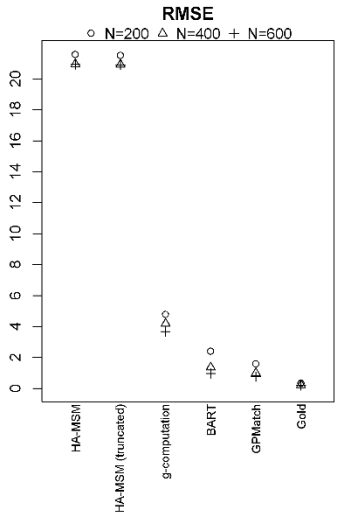
Y10-Y00



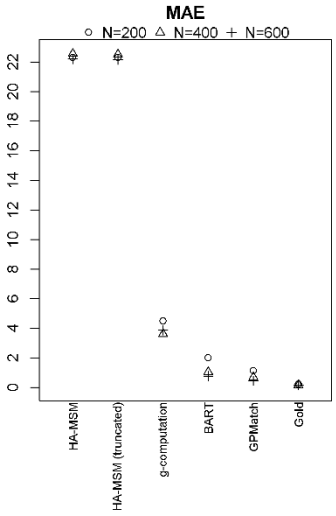


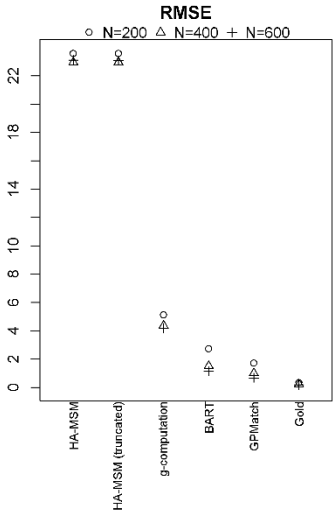
Y11-Y00



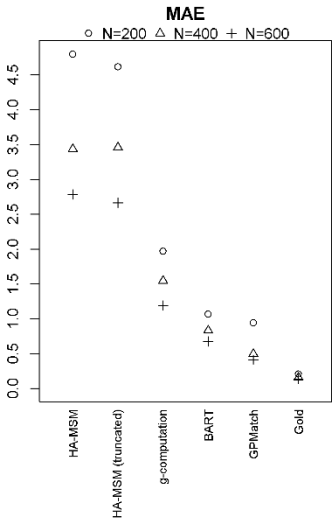


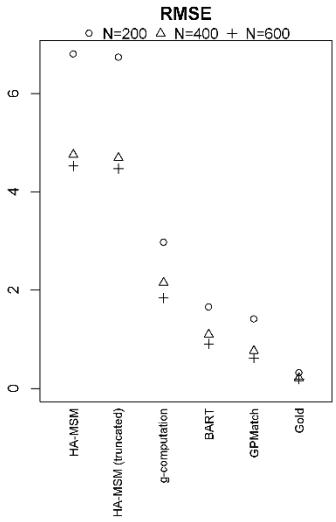
Y10-Y01



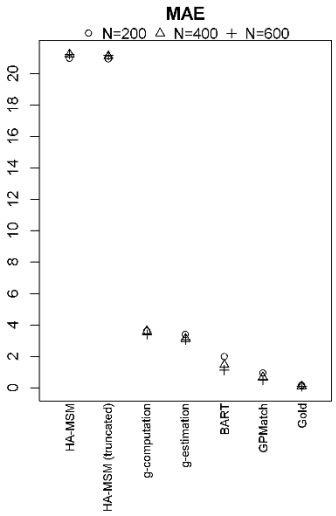


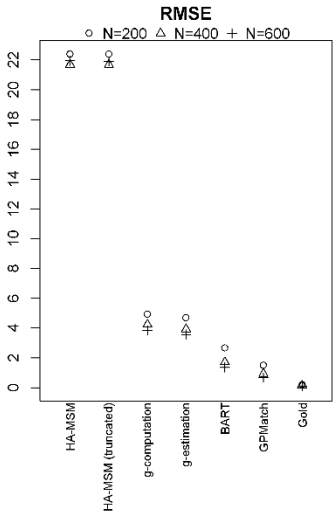
Y11-Y01



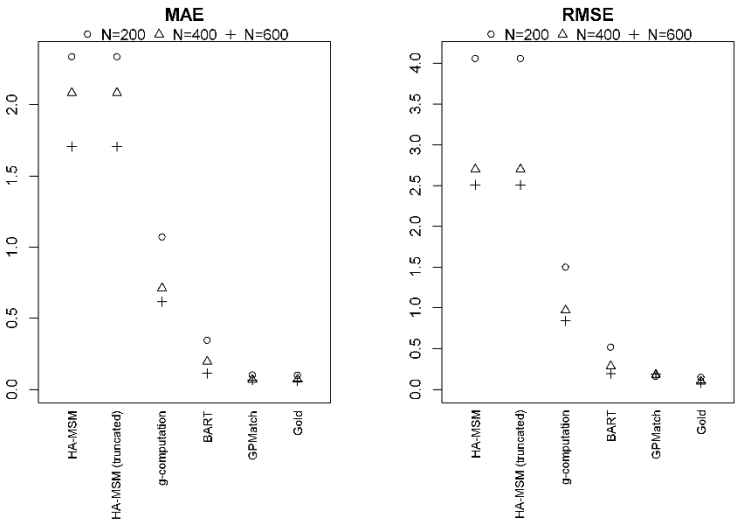


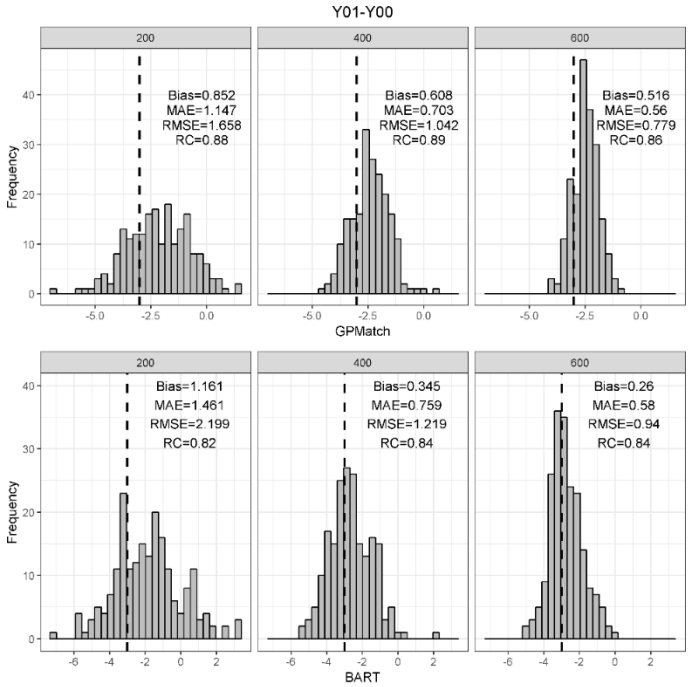
Y11-Y10

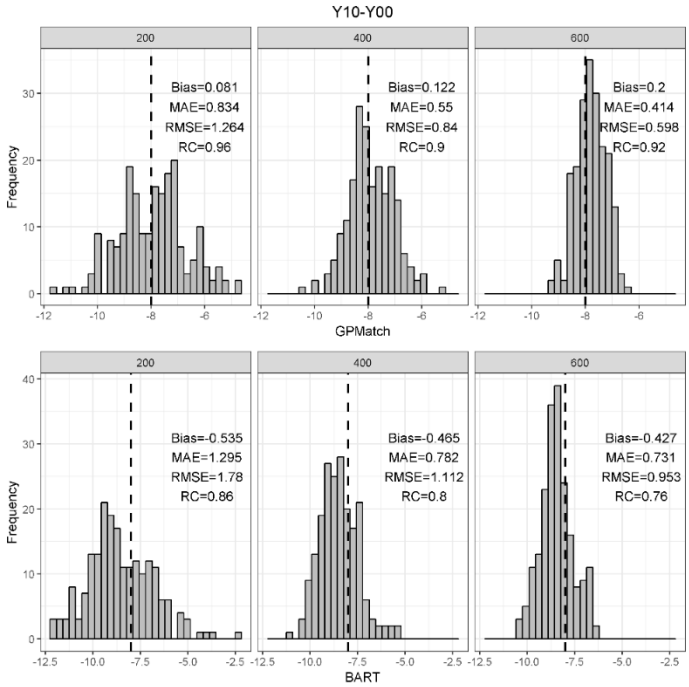


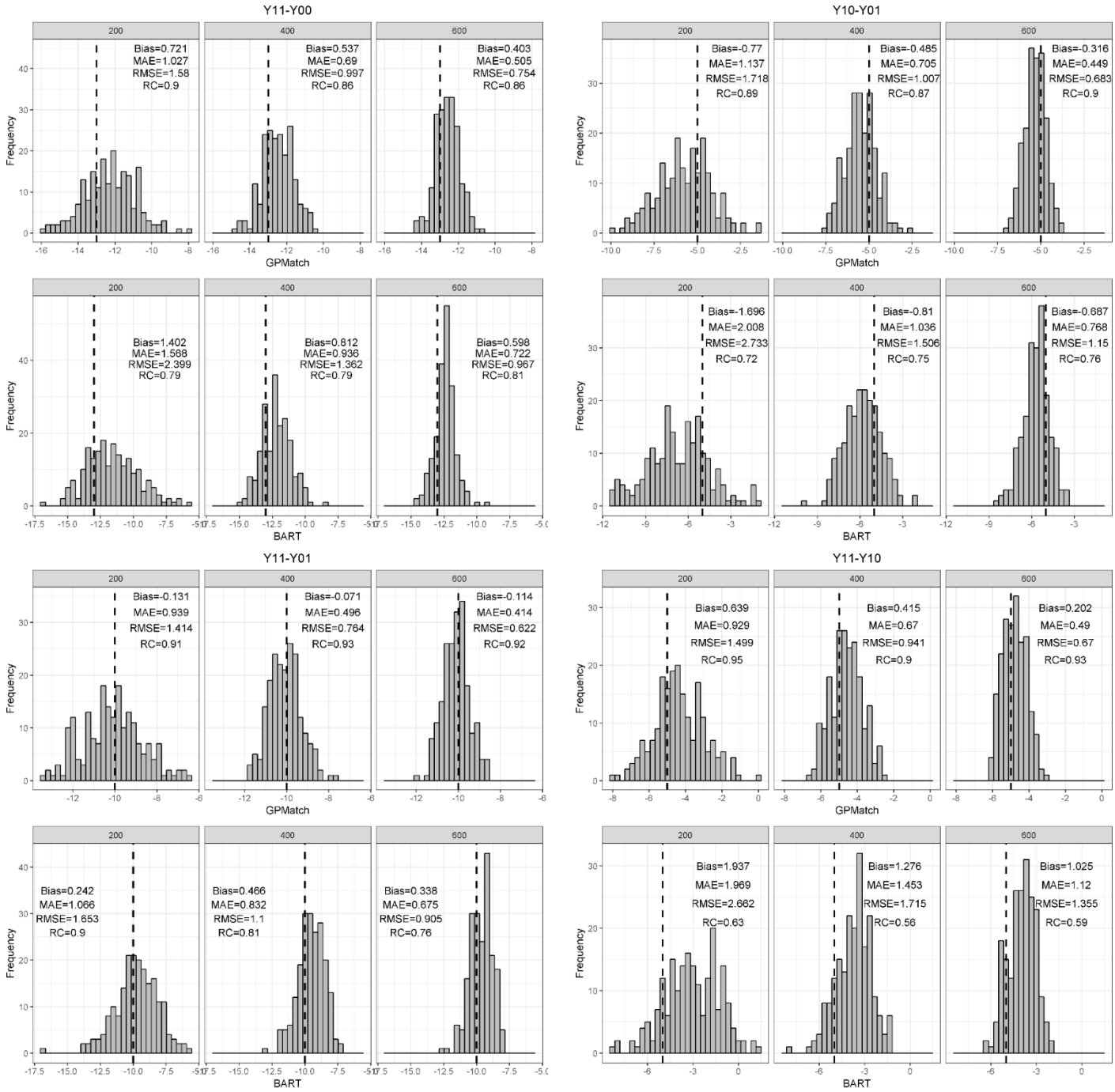


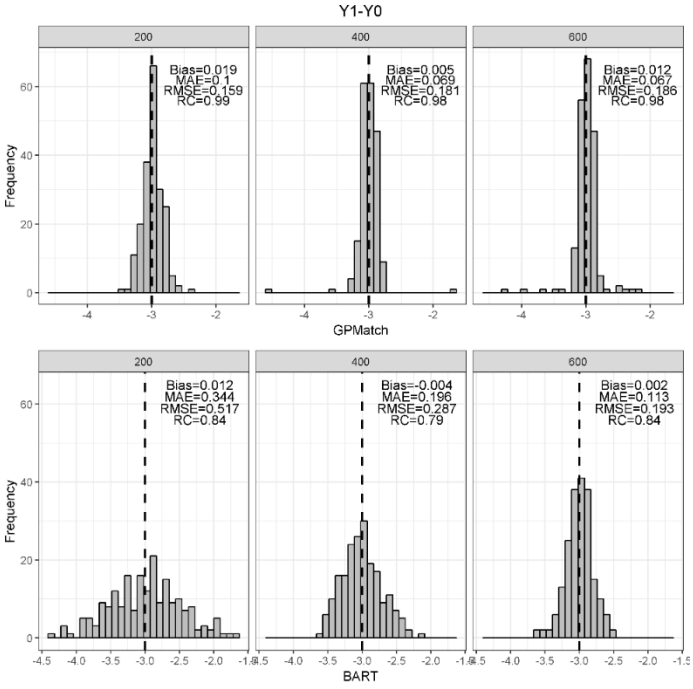
Y1-Y0


MAE
○ N=200 △ N=400 + N=600
0.0
0.5
1.0
1.5
2.0
HA-MSM
HA-MSM (truncated)
g-computation
BART
GPMatch
Gold
RMSE
○ N=200 △ N=400 + N=600
0.0
0.5
1.0
1.5
2.0
2.5
3.0
3.5
4.0
HA-MSM
HA-MSM (truncated)
g-computation
BART
GPMatch
Gold


Y01-Y00
200
400
600
Bias=0.852
MAE=1.147
RMSE=1.658
RC=0.88
Bias=0.608
MAE=0.703
RMSE=1.042
RC=0.89
Bias=0.516
MAE=0.56
RMSE=0.779
RC=0.86
Frequency
GPMatch
Bias=1.161
MAE=1.461
RMSE=2.199
RC=0.82
Bias=0.345
MAE=0.759
RMSE=1.219
RC=0.84
Bias=0.26
MAE=0.58
RMSE=0.94
RC=0.84
Frequency
BART


Y10-Y00
200
400
600
Bias=0.081
MAE=0.834
RMSE=1.264
RC=0.96
Bias=0.122
MAE=0.55
RMSE=0.84
RC=0.9
Bias=0.2
MAE=0.414
RMSE=0.598
RC=0.92
Frequency
GPMatch
Bias=-0.535
MAE=1.295
RMSE=1.78
RC=0.86
Bias=-0.465
MAE=0.782
RMSE=1.112
RC=0.8
Bias=-0.427
MAE=0.731
RMSE=0.953
RC=0.76
Frequency
BART


Y11-Y00
200
400
600
Bias=0.721
MAE=1.027
RMSE=1.58
RC=0.9
Bias=0.537
MAE=0.69
RMSE=0.997
RC=0.86
Bias=0.403
MAE=0.505
RMSE=0.754
RC=0.86
Frequency
GPMatch
Y10-Y01
Bias=-0.77
MAE=1.137
RMSE=1.718
RC=0.89
Bias=-0.485
MAE=0.705
RMSE=1.007
RC=0.87
Bias=-0.316
MAE=0.449
RMSE=0.683
RC=0.9
GPMatch
Bias=1.402
MAE=1.568
RMSE=2.399
RC=0.79
Bias=0.812
MAE=0.936
RMSE=1.362
RC=0.79
Bias=0.598
MAE=0.722
RMSE=0.967
RC=0.81
BART
Bias=-1.696
MAE=2.008
RMSE=2.733
RC=0.72
Bias=-0.81
MAE=1.036
RMSE=1.506
RC=0.75
Bias=-0.687
MAE=0.768
RMSE=1.15
RC=0.76
BART
Y11-Y01
Bias=-0.131
MAE=0.939
RMSE=1.414
RC=0.91
Bias=-0.071
MAE=0.496
RMSE=0.764
RC=0.93
Bias=-0.114
MAE=0.414
RMSE=0.622
RC=0.92
GPMatch
Y11-Y10
Bias=0.639
MAE=0.929
RMSE=1.499
RC=0.95
Bias=0.415
MAE=0.67
RMSE=0.941
RC=0.9
Bias=0.202
MAE=0.49
RMSE=0.67
RC=0.93
GPMatch
Bias=0.242
MAE=1.066
RMSE=1.653
RC=0.9
Bias=0.466
MAE=0.832
RMSE=1.1
RC=0.81
Bias=0.338
MAE=0.675
RMSE=0.905
RC=0.76
BART
Bias=1.937
MAE=1.969
RMSE=2.662
RC=0.63
Bias=1.276
MAE=1.453
RMSE=1.715
RC=0.56
Bias=1.025
MAE=1.12
RMSE=1.355
RC=0.59
BART



4. Heterogeneous Treatment Effect SMART Trial

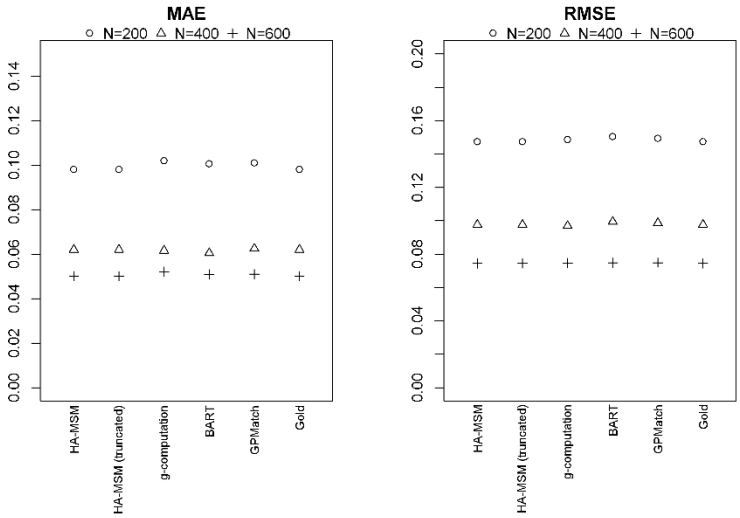
It is expected that the outcome at the 1st stage may modify the effect of the next stage treatment effect. Here, we consider a simple additive interaction effect: s

$$X \sim Bernoulli(0.4),\ A_0 \sim Bernoulli(0.5),\ L_1 \sim N(0,1)$$

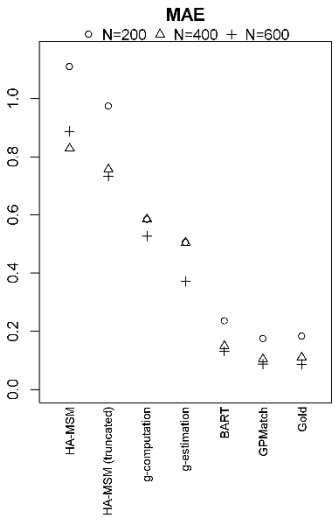
$$A_1|L_1, A_0, X \sim Bernoulli(expit\left(0.2 - 0.2A_0 + L_1^{1/3}\right))$$

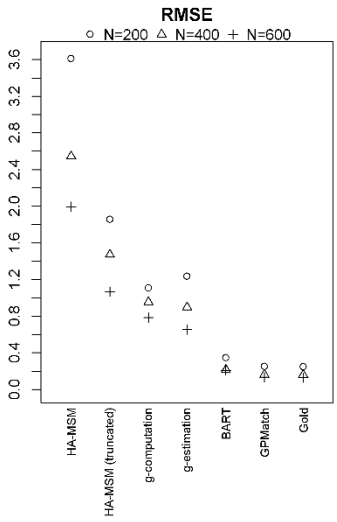
$$Y(A_0, A_1, L_1)|X \sim N(-2 + 2.5A_0 + 3.5A_1 + 0.5A_0A_1 - 3\exp(-L_1) + A_1L_1, sd = 1)$$

Y1-Y0

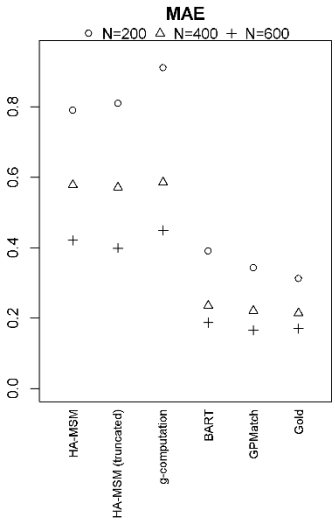


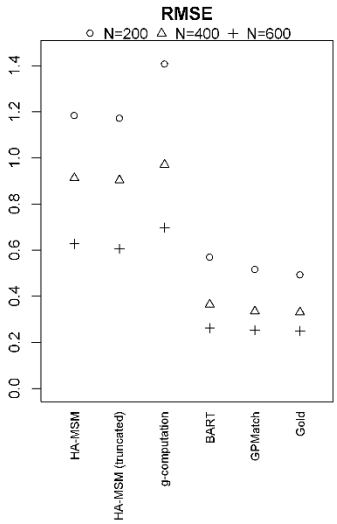
Y01-Y00



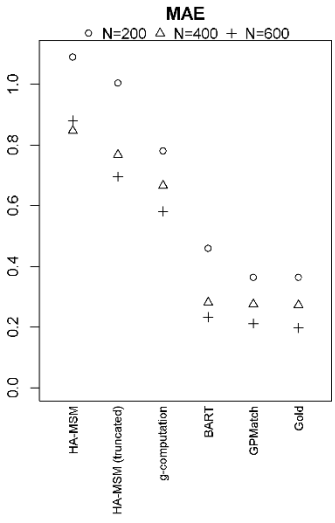


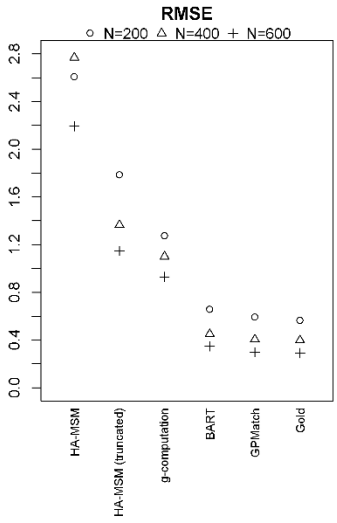
Y10-Y00



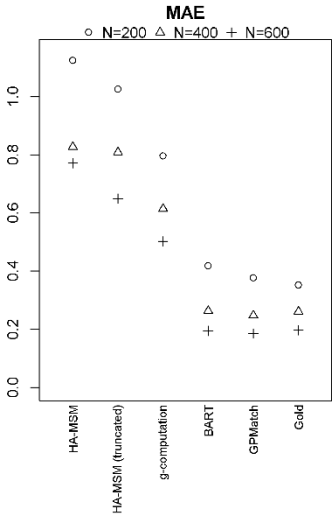


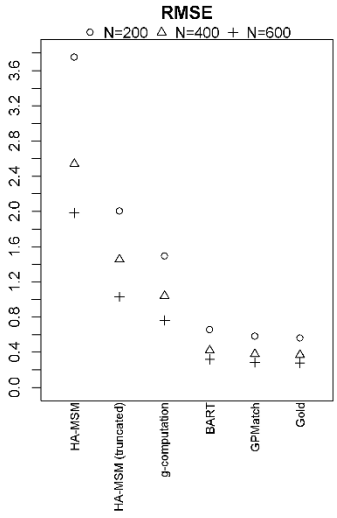
Y11-Y00



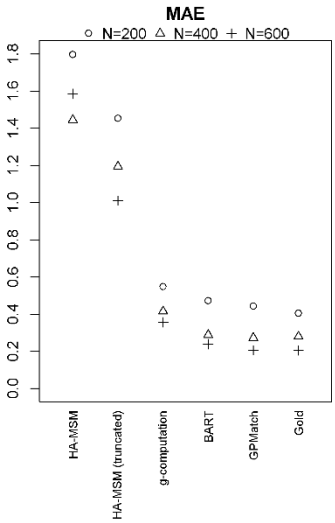


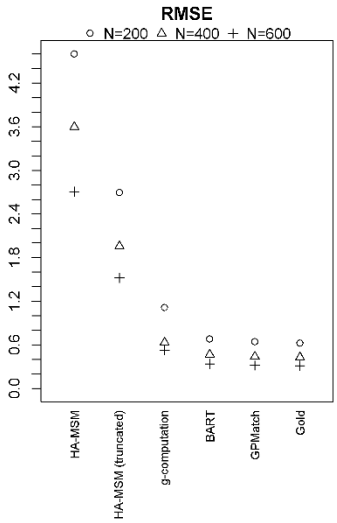
Y10-Y01



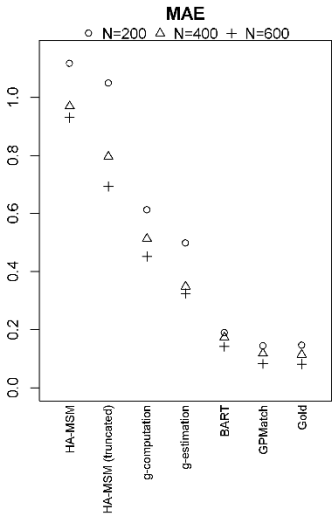


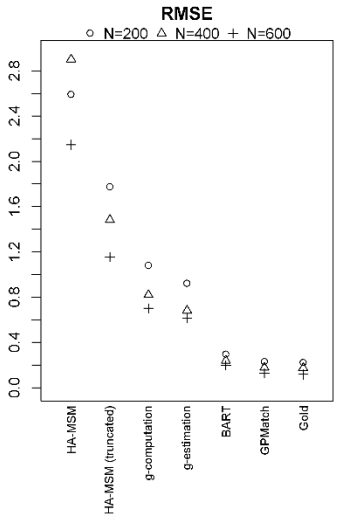
Y11-Y01

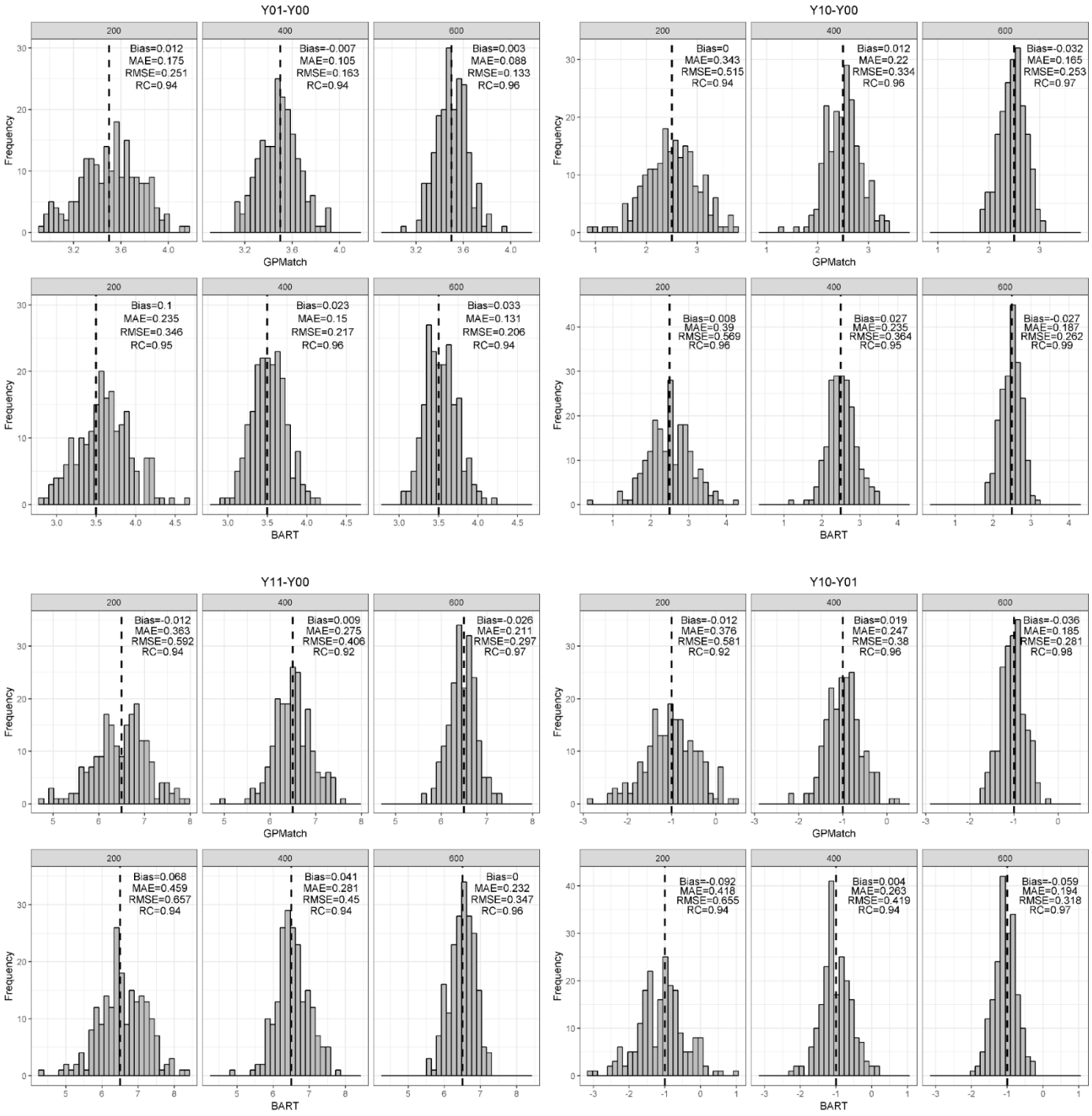


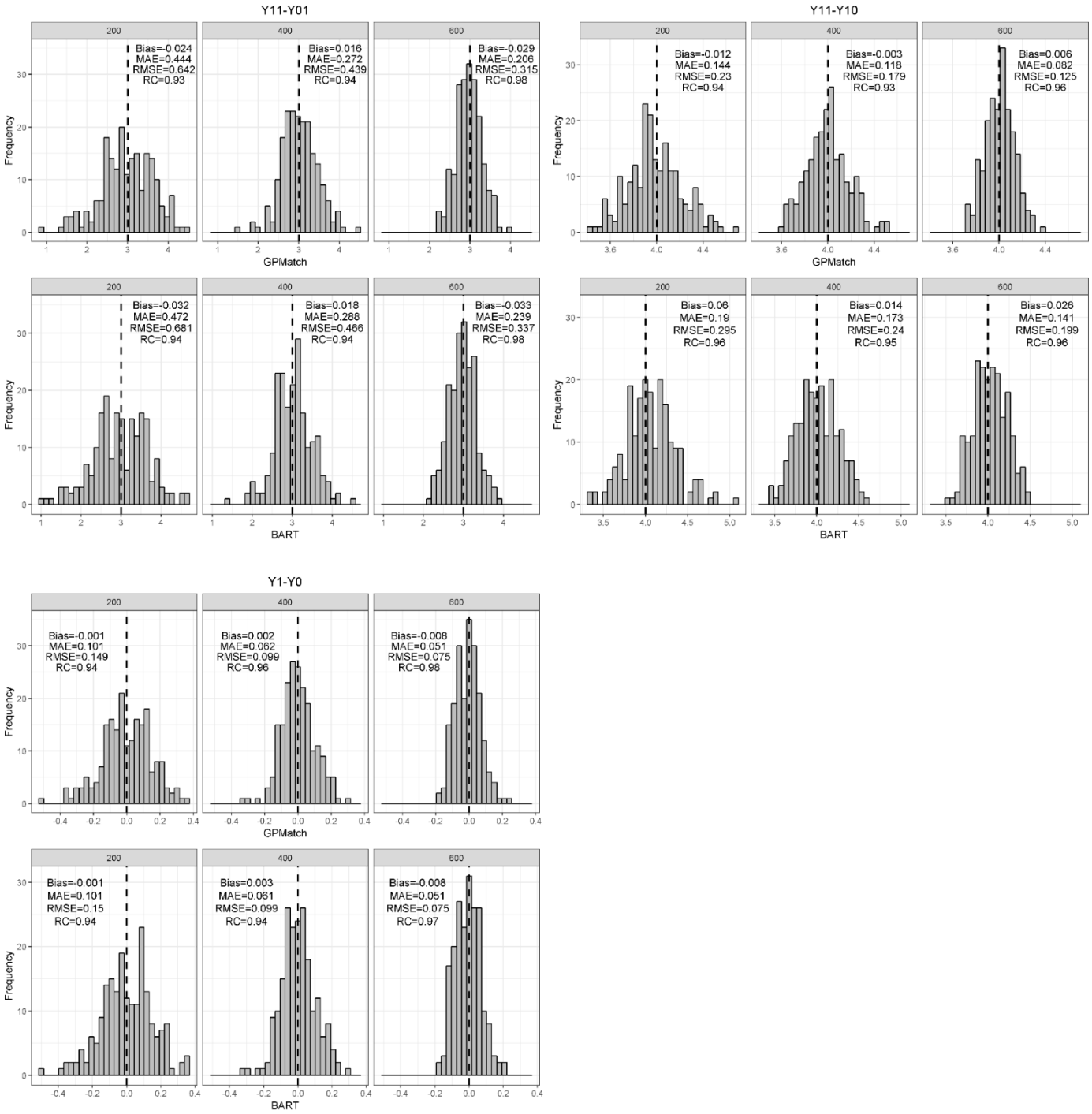


Y11-Y10






Y01-Y00
200
400
600
Bias=0.012
MAE=0.175
RMSE=0.251
RC=0.94
Bias=-0.007
MAE=0.105
RMSE=0.163
RC=0.94
Bias=0.003
MAE=0.088
RMSE=0.133
RC=0.96
Frequency
GPMatch
Bias=0.1
MAE=0.235
RMSE=0.346
RC=0.95
Bias=0.023
MAE=0.15
RMSE=0.217
RC=0.96
Bias=0.033
MAE=0.131
RMSE=0.206
RC=0.94
BART
Y10-Y00
Bias=0
MAE=0.343
RMSE=0.515
RC=0.94
Bias=0.012
MAE=0.22
RMSE=0.334
RC=0.96
Bias=-0.032
MAE=0.165
RMSE=0.253
RC=0.97
Bias=0.008
MAE=0.39
RMSE=0.569
RC=0.96
Bias=0.027
MAE=0.235
RMSE=0.364
RC=0.95
Bias=-0.027
MAE=0.187
RMSE=0.262
RC=0.99
Y11-Y00
Bias=-0.012
MAE=0.363
RMSE=0.592
RC=0.94
Bias=0.009
MAE=0.275
RMSE=0.406
RC=0.92
Bias=-0.026
MAE=0.211
RMSE=0.297
RC=0.97
Bias=0.068
MAE=0.459
RMSE=0.657
RC=0.94
Bias=0.041
MAE=0.281
RMSE=0.45
RC=0.94
Bias=0
MAE=0.232
RMSE=0.347
RC=0.96
Y10-Y01
Bias=-0.012
MAE=0.376
RMSE=0.581
RC=0.92
Bias=0.019
MAE=0.247
RMSE=0.38
RC=0.96
Bias=-0.036
MAE=0.185
RMSE=0.281
RC=0.98
Bias=-0.092
MAE=0.418
RMSE=0.655
RC=0.94
Bias=0.004
MAE=0.263
RMSE=0.419
RC=0.94
Bias=-0.059
MAE=0.194
RMSE=0.318
RC=0.97



5. Observational Study Adaptive Treatment Subgroup Treatment Effect

This simulation implemented a modified simulation design setting used in Schulte's et. al. (2017). It considered a 3-level categorical baseline covariates $X = -5,0,5$, with multinomial distribution, $X \sim Multinomial\left(\frac{1}{3},\frac{1}{3},\frac{1}{3}\right)$. The causal treatment effect at both stages varies by the baseline covariates.

$$A_0 \sim Bernoulli(expit(0.3 - 0.05X))$$

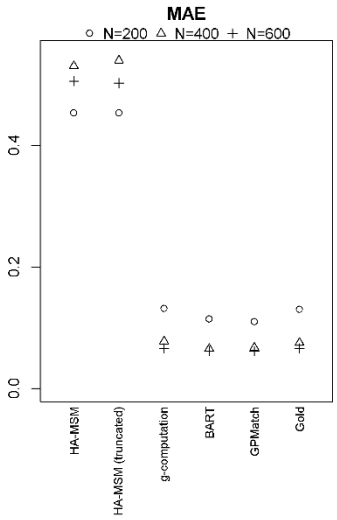
$$L_1(a_0) \sim N(0.75X - 0.75a_0 - 0.25a_0X, 1)$$

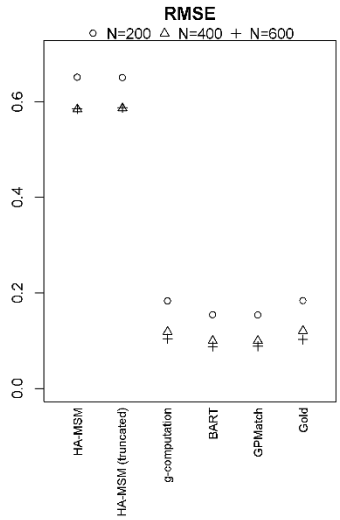
$$L_1 = A_0L_1(1) + (1 - A_0)L_1(0)$$

$$A_1 \sim Bernoulli(expit(0.05X + 0.2A_0 - 0.05L_1 - 0.1L_1 * A_0 - 0.01L_1^2))$$

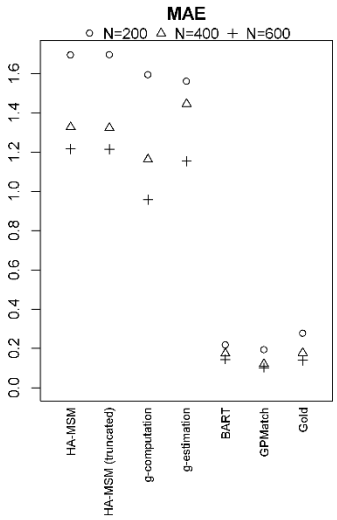
$$Y(a_0, a_1)|L_1(a_0) \sim N(3 + 0.5a_0 + 0.4a_0X - L_1(a_0) - L_1(a_0)^2 + 2a_1 - a_0a_1 + a_1L_1(a_0), 1)$$

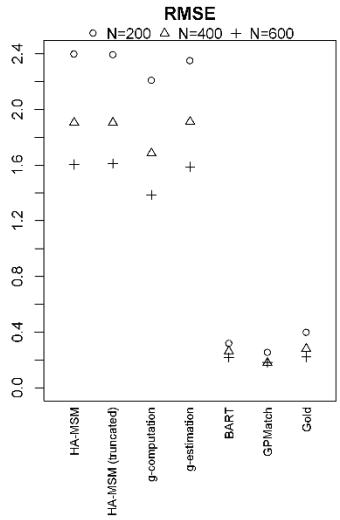
Y1-Y0



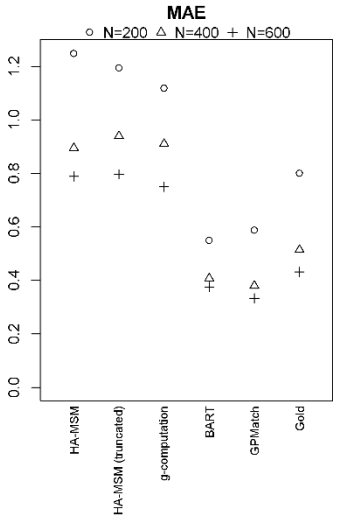


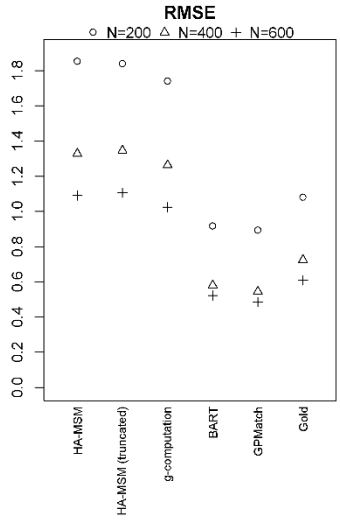
Y01-Y00



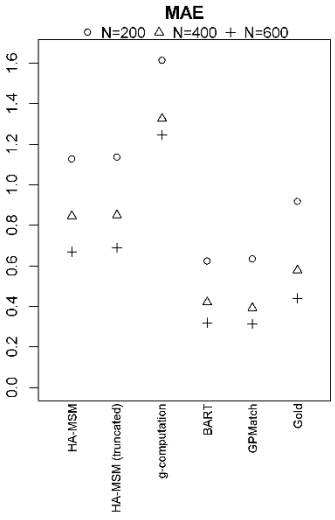


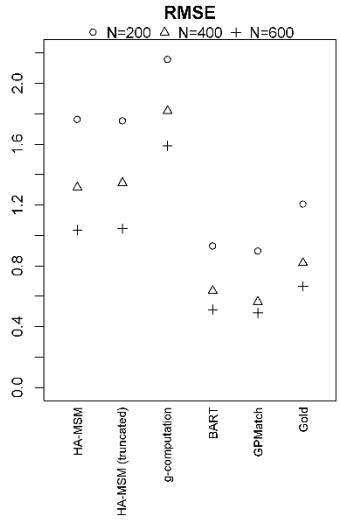
Y10-Y00



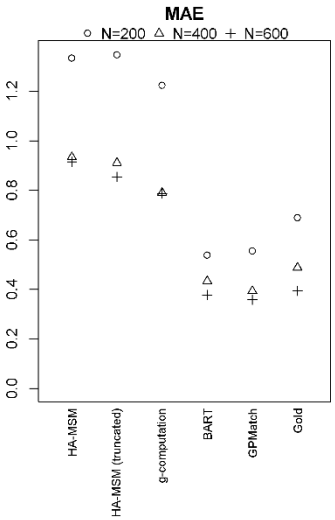


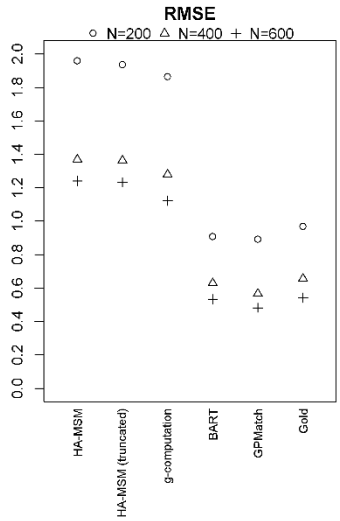
Y11-Y00



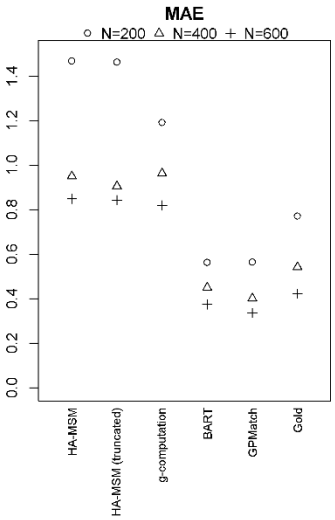


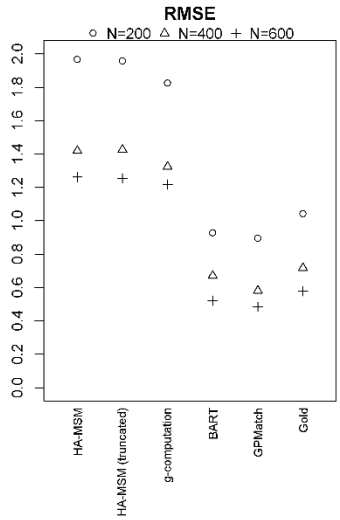
Y10-Y01



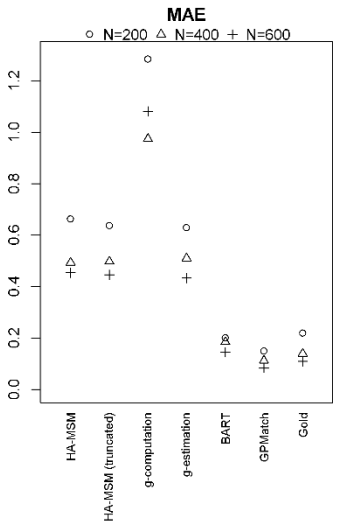


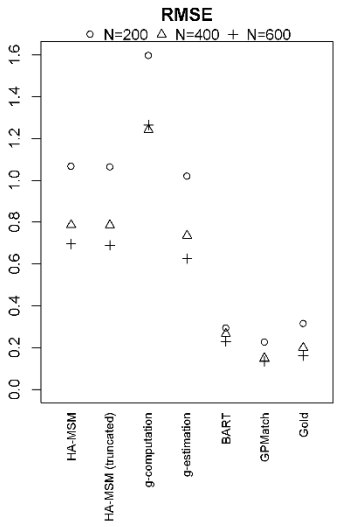
Y11-Y01

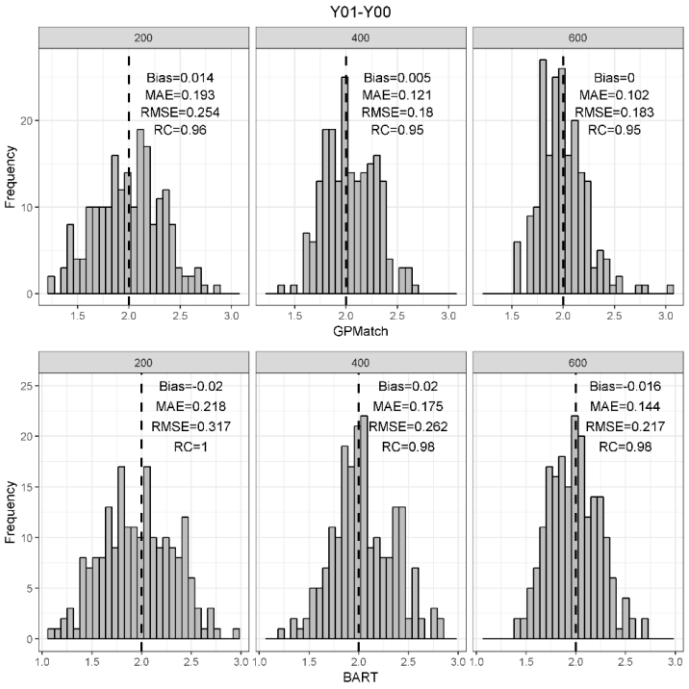


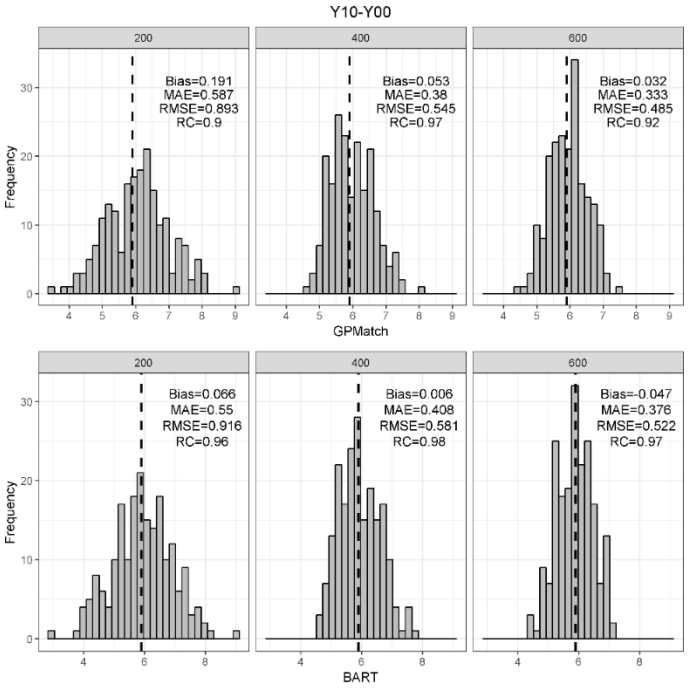


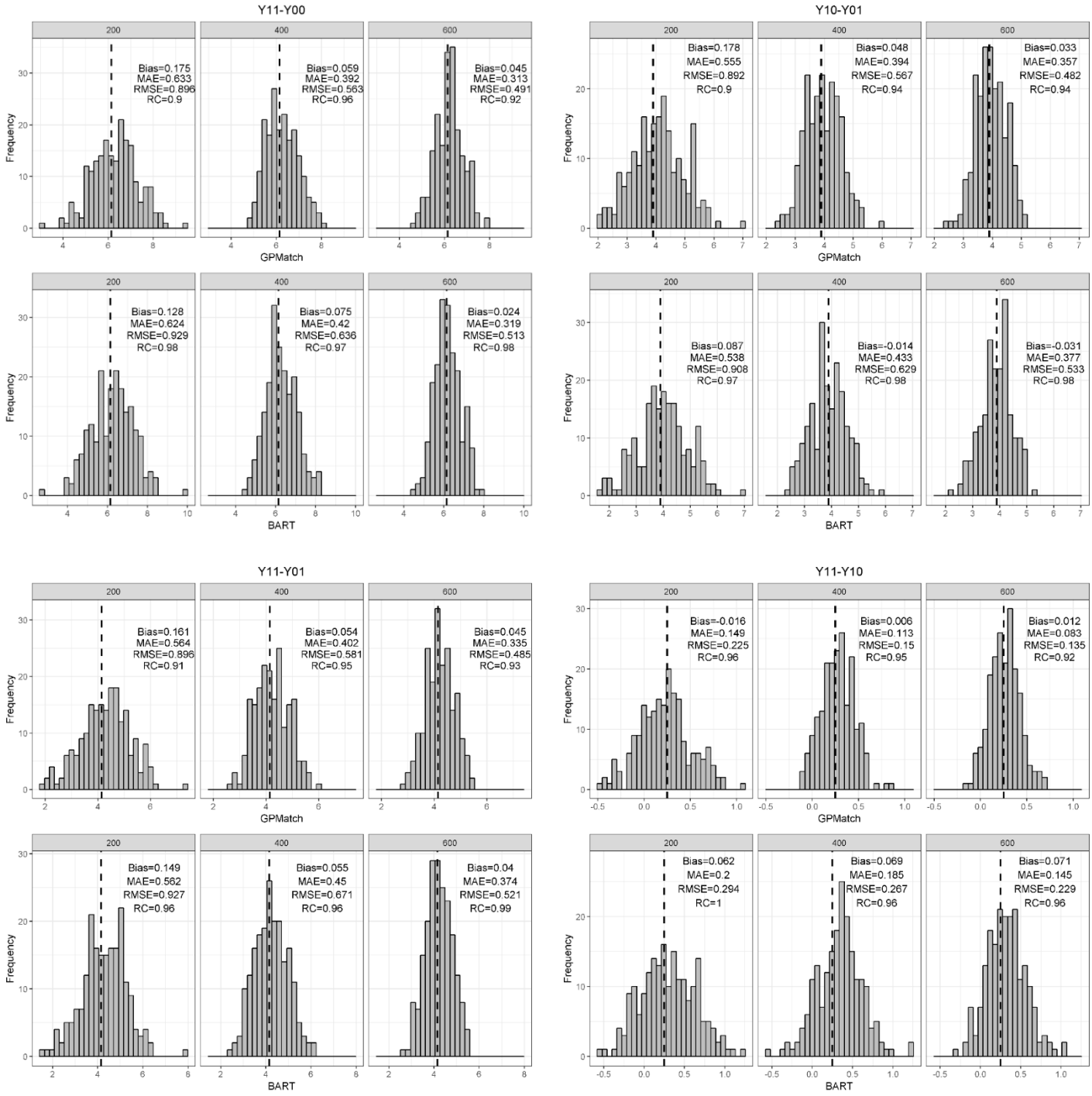
Y11-Y10

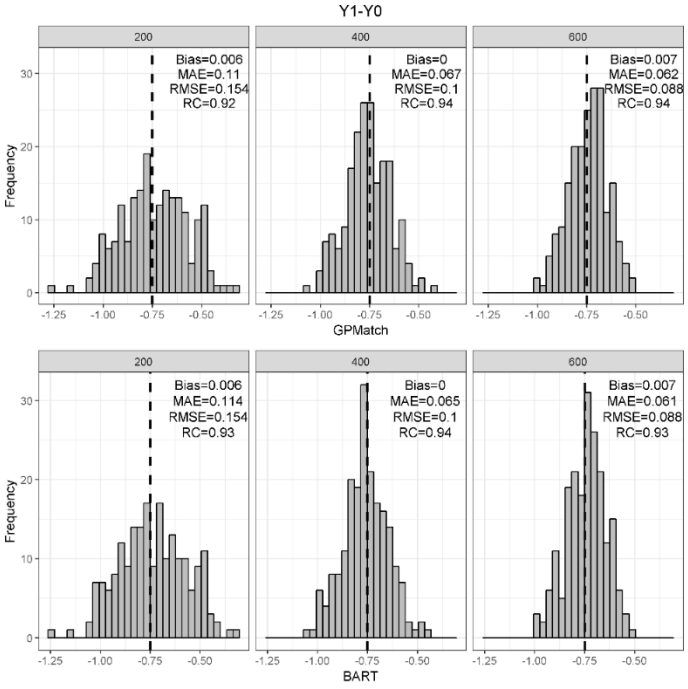

MAE
N=200 N=400 N=600
0.0
0.2
0.4
0.6
0.8
1.0
1.2
HA-MSM
HA-MSM (truncated)
g-computation
g-estimation
BART
GPMatch
Gold


RMSE
N=200 N=400 N=600
0.0
0.2
0.4
0.6
0.8
1.0
1.2
1.4
1.6
HA-MSM
HA-MSM (truncated)
g-computation
g-estimation
BART
GPMatch
Gold


Y01-Y00
200
400
600
Bias=0.014
MAE=0.193
RMSE=0.254
RC=0.96
Bias=0.005
MAE=0.121
RMSE=0.18
RC=0.95
Bias=0
MAE=0.102
RMSE=0.183
RC=0.95
Frequency
GPMatch
Bias=-0.02
MAE=0.218
RMSE=0.317
RC=1
Bias=0.02
MAE=0.175
RMSE=0.262
RC=0.98
Bias=-0.016
MAE=0.144
RMSE=0.217
RC=0.98
BART


Y10-Y00
200
400
600
Bias=0.191
MAE=0.587
RMSE=0.893
RC=0.9
Bias=0.053
MAE=0.38
RMSE=0.545
RC=0.97
Bias=0.032
MAE=0.333
RMSE=0.485
RC=0.92
Frequency
GPMatch
Bias=0.066
MAE=0.55
RMSE=0.916
RC=0.96
Bias=0.006
MAE=0.408
RMSE=0.581
RC=0.98
Bias=-0.047
MAE=0.376
RMSE=0.522
RC=0.97
BART


Y11-Y00
200
400
600
Bias=0.175
MAE=0.633
RMSE=0.896
RC=0.9
Bias=0.059
MAE=0.392
RMSE=0.563
RC=0.96
Bias=0.045
MAE=0.313
RMSE=0.491
RC=0.92
Frequency
GPMatch
Bias=0.128
MAE=0.624
RMSE=0.929
RC=0.98
Bias=0.075
MAE=0.42
RMSE=0.636
RC=0.97
Bias=0.024
MAE=0.319
RMSE=0.513
RC=0.98
BART
Y10-Y01
Bias=0.178
MAE=0.555
RMSE=0.892
RC=0.9
Bias=0.048
MAE=0.394
RMSE=0.567
RC=0.94
Bias=0.033
MAE=0.357
RMSE=0.482
RC=0.94
Bias=0.087
MAE=0.538
RMSE=0.908
RC=0.97
Bias=-0.014
MAE=0.433
RMSE=0.629
RC=0.98
Bias=-0.031
MAE=0.377
RMSE=0.533
RC=0.98
Y11-Y01
Bias=0.161
MAE=0.564
RMSE=0.896
RC=0.91
Bias=0.054
MAE=0.402
RMSE=0.581
RC=0.95
Bias=0.045
MAE=0.335
RMSE=0.485
RC=0.93
Bias=0.149
MAE=0.562
RMSE=0.927
RC=0.96
Bias=0.055
MAE=0.45
RMSE=0.671
RC=0.96
Bias=0.04
MAE=0.374
RMSE=0.521
RC=0.99
Y11-Y10
Bias=-0.016
MAE=0.149
RMSE=0.225
RC=0.96
Bias=0.006
MAE=0.113
RMSE=0.15
RC=0.95
Bias=0.012
MAE=0.083
RMSE=0.135
RC=0.92
Bias=0.062
MAE=0.2
RMSE=0.294
RC=1
Bias=0.069
MAE=0.185
RMSE=0.267
RC=0.96
Bias=0.071
MAE=0.145
RMSE=0.229
RC=0.96


Y1-Y0
200
400
600
Bias=0.006
MAE=0.11
RMSE=0.154
RC=0.92
Bias=0
MAE=0.067
RMSE=0.1
RC=0.94
Bias=0.007
MAE=0.062
RMSE=0.088
RC=0.94
Frequency
GPMatch
Bias=0.006
MAE=0.114
RMSE=0.154
RC=0.93
Bias=0
MAE=0.065
RMSE=0.1
RC=0.94
Bias=0.007
MAE=0.061
RMSE=0.088
RC=0.93
BART