

Supplementary materials for “Design and analysis of partially randomized preference trials with propensity score stratification”

Yumin Wang, Fan Li, Ondrej Blaha, Can Meng, Denise Esserman

Tables for mean effect estimates (bias) (referenced in section 4.2.1 of main paper)

Table S.1: Absolute and percent relative bias based on 10,000 simulations with different propensity score stratification (PSS) strategies. The true effect parameter is $\Delta\pi = 0.4$, the nominal type I error rate is fixed at $\alpha = 0.05$, the number of strata varies in $s \in \{5, 8, 10\}$, allocation proportion to the choice arm varies in $\theta \in \{0.3, 0.5, 0.7\}$, the preference rate for treatment 1 varies in $\phi \in \{0.3, 0.5, 0.7\}$. Results were summarized under the homogeneous effect model.

Strata	θ	ϕ	Absolute Bias (Percent Relative Bias)				Sample size
			No PSS	PSS (X_1, X_2)	PSS (X_1)	PSS (X_2)	
5	0.3	0.3	0.108 (27.00)	0.013 (3.25)	0.038 (9.50)	0.090 (22.50)	1535
		0.5	0.092 (23.00)	0.017 (4.25)	0.038 (9.50)	0.081 (20.25)	1066
		0.7	0.111 (27.75)	0.024 (6.00)	0.049 (12.25)	0.101 (25.25)	1565
	0.5	0.3	0.105 (26.25)	0.009 (2.25)	0.035 (8.75)	0.087 (21.75)	1315
		0.5	0.090 (22.50)	0.014 (3.50)	0.035 (8.75)	0.078 (19.50)	878
		0.7	0.106 (26.50)	0.015 (3.75)	0.040 (10.00)	0.092 (23.00)	1370
	0.7	0.3	0.108 (27.00)	0.012 (3.00)	0.037 (9.25)	0.088 (22.00)	1700
		0.5	0.089 (22.25)	0.011 (2.75)	0.031 (7.75)	0.074 (18.50)	1075
		0.7	0.109 (27.25)	0.015 (3.75)	0.039 (9.75)	0.092 (23.00)	1720
8	0.3	0.3	0.109 (27.25)	0.011 (2.75)	0.037 (9.25)	0.090 (22.50)	1510
		0.5	0.094 (23.50)	0.019 (4.75)	0.041 (10.25)	0.084 (21.00)	1055
		0.7	0.113 (28.25)	0.028 (7.00)	0.055 (13.75)	0.106 (26.50)	1585
	0.5	0.3	0.110 (27.50)	0.010 (2.50)	0.036 (9.00)	0.088 (22.00)	1340
		0.5	0.094 (23.50)	0.016 (4.00)	0.038 (9.50)	0.082 (20.50)	878
		0.7	0.108 (27.00)	0.018 (4.50)	0.044 (11.00)	0.096 (24.00)	1335
	0.7	0.3	0.108 (27.00)	0.008 (2.00)	0.034 (8.50)	0.086 (21.50)	1680
		0.5	0.090 (22.50)	0.008 (2.00)	0.030 (7.50)	0.074 (18.50)	1090
		0.7	0.109 (27.25)	0.013 (3.25)	0.039 (9.75)	0.090 (22.50)	1705
10	0.3	0.3	0.109 (27.25)	0.010 (2.50)	0.037 (9.25)	0.091 (22.75)	1532
		0.5	0.092 (23.00)	0.021 (5.25)	0.043 (10.75)	0.087 (21.75)	1055
		0.7	0.109 (27.25)	0.030 (7.50)	0.056 (14.00)	0.110 (27.50)	1568
	0.5	0.3	0.108 (27.00)	0.007 (1.75)	0.034 (8.50)	0.088 (22.00)	1337
		0.5	0.090 (22.50)	0.013 (3.25)	0.035 (8.75)	0.079 (19.75)	880
		0.7	0.108 (27.00)	0.019 (4.75)	0.046 (11.50)	0.099 (24.75)	1370
	0.7	0.3	0.107 (26.75)	0.005 (1.25)	0.032 (8.00)	0.084 (21.00)	1665
		0.5	0.090 (22.50)	0.009 (2.25)	0.031 (7.75)	0.076 (19.00)	1090
		0.7	0.110 (27.50)	0.015 (3.75)	0.042 (10.50)	0.094 (23.50)	1715

Table S.2: Absolute and percent relative bias based on 10,000 simulations with different propensity score stratification (PSS) strategies. The true effect parameter is $\Delta\nu = 0.4$, the nominal type I error rate is fixed at $\alpha = 0.05$, the number of strata varies in $s \in \{5, 8, 10\}$, allocation proportion to the choice arm varies in $\theta \in \{0.3, 0.5, 0.7\}$, the preference rate for treatment 1 varies in $\phi \in \{0.3, 0.5, 0.7\}$. Results were summarized under the homogeneous effect model.

Strata	θ	ϕ	Absolute Bias (Percent Relative Bias)				Sample size
			No PSS	PSS (X_1, X_2)	PSS (X_1)	PSS (X_2)	
5	0.3	0.3	0.045 (11.25)	0.004 (1.00)	0.014 (3.50)	0.036 (9.00)	1520
		0.5	0.001 (0.25)	0.007 (1.75)	0.007 (1.75)	0.007 (1.75)	1070
		0.7	0.046 (11.50)	0.016 (4.00)	0.026 (6.50)	0.047 (11.75)	1590
	0.5	0.3	0.045 (11.25)	0.006 (1.50)	0.015 (3.75)	0.036 (9.00)	1335
		0.5	0.001 (0.25)	0.003 (0.75)	0.004 (1.00)	0.003 (0.75)	880
		0.7	0.043 (10.75)	0.011 (2.75)	0.021 (5.25)	0.042 (10.50)	1370
	0.7	0.3	0.043 (10.75)	0.004 (1.00)	0.013 (3.25)	0.034 (8.50)	1718
		0.5	0.001 (0.25)	0.001 (0.25)	0.002 (0.50)	0.002 (0.50)	1100
		0.7	0.044 (11.00)	0.010 (2.50)	0.020 (5.00)	0.040 (10.00)	1710
8	0.3	0.3	0.042 (10.50)	0.001 (0.25)	0.010 (2.50)	0.031 (7.75)	1565
		0.5	0.000 (0.00)	0.012 (3.00)	0.012 (3.00)	0.012 (3.00)	1048
		0.7	0.044 (11.00)	0.021 (5.25)	0.032 (8.00)	0.052 (13.00)	1600
	0.5	0.3	0.041 (10.25)	0.001 (0.25)	0.010 (2.50)	0.031 (7.75)	1325
		0.5	0.001 (0.25)	0.008 (2.00)	0.008 (2.00)	0.008 (2.00)	883
		0.7	0.044 (11.00)	0.015 (3.75)	0.026 (6.50)	0.046 (11.50)	1345
	0.7	0.3	0.043 (10.75)	0.001 (0.25)	0.011 (2.75)	0.032 (8.00)	1678
		0.5	0.002 (0.50)	0.006 (1.50)	0.007 (1.75)	0.006 (1.50)	1092
		0.7	0.044 (11.00)	0.008 (2.00)	0.020 (5.00)	0.041 (10.25)	1715
10	0.3	0.3	0.044 (11.00)	0.002 (0.50)	0.010 (2.50)	0.033 (8.25)	1530
		0.5	0.000 (0.00)	0.015 (3.75)	0.015 (3.75)	0.015 (3.75)	1052
		0.7	0.045 (11.25)	0.027 (6.75)	0.038 (9.50)	0.059 (14.75)	1610
	0.5	0.3	0.046 (11.50)	0.003 (0.75)	0.013 (3.25)	0.035 (8.75)	1360
		0.5	0.001 (0.25)	0.009 (2.25)	0.009 (2.25)	0.008 (2.00)	875
		0.7	0.043 (10.75)	0.017 (4.25)	0.028 (7.00)	0.050 (12.50)	1375
	0.7	0.3	0.043 (10.75)	0.000 (0.00)	0.011 (2.75)	0.032 (8.00)	1675
		0.5	0.001 (0.25)	0.005 (1.25)	0.005 (1.25)	0.006 (1.50)	1085
		0.7	0.045 (11.25)	0.012 (3.00)	0.023 (5.75)	0.044 (11.00)	1710

Table S.3: Absolute and percent relative bias based on 10,000 simulations with different propensity score stratification (PSS) strategies. The true effect parameter is $\Delta\tau = 0.4$, the nominal type I error rate is fixed at $\alpha = 0.05$, the number of strata varies in $s \in \{5, 8, 10\}$, allocation proportion to the choice arm varies in $\theta \in \{0.3, 0.5, 0.7\}$, the preference rate for treatment 1 varies in $\phi \in \{0.3, 0.5, 0.7\}$. Results were summarized under the homogeneous effect model.

Strata	θ	ϕ	Absolute Bias (Percent Relative Bias)				Sample size
			No PSS	PSS (X_1, X_2)	PSS (X_1)	PSS (X_2)	
5	0.3	0.3	0.000 (0.00)	0.000 (0.00)	0.001 (0.25)	0.000 (0.00)	314
		0.5	0.002 (0.50)	0.002 (0.50)	0.002 (0.50)	0.001 (0.25)	319
		0.7	0.002 (0.50)	0.002 (0.50)	0.003 (0.75)	0.003 (0.75)	317
	0.5	0.3	0.002 (0.50)	0.001 (0.25)	0.001 (0.25)	0.001 (0.25)	449
		0.5	0.000 (0.00)	0.001 (0.25)	0.000 (0.00)	0.000 (0.00)	451
		0.7	0.001 (0.25)	0.000 (0.00)	0.000 (0.00)	0.000 (0.00)	453
	0.7	0.3	0.001 (0.25)	0.000 (0.00)	0.000 (0.00)	0.000 (0.00)	787
		0.5	0.001 (0.25)	0.001 (0.25)	0.001 (0.25)	0.001 (0.25)	785
		0.7	0.001 (0.25)	0.001 (0.25)	0.000 (0.00)	0.000 (0.00)	773
8	0.3	0.3	0.002 (0.50)	0.001 (0.25)	0.001 (0.25)	0.001 (0.25)	312
		0.5	0.001 (0.25)	0.001 (0.25)	0.001 (0.25)	0.001 (0.25)	319
		0.7	0.004 (1.00)	0.004 (1.00)	0.004 (1.00)	0.004 (1.00)	314
	0.5	0.3	0.001 (0.25)	0.001 (0.25)	0.002 (0.50)	0.002 (0.50)	452
		0.5	0.001 (0.25)	0.000 (0.00)	0.001 (0.25)	0.000 (0.00)	449
		0.7	0.001 (0.25)	0.001 (0.25)	0.000 (0.00)	0.000 (0.00)	453
	0.7	0.3	0.000 (0.00)	0.000 (0.00)	0.001 (0.25)	0.000 (0.00)	770
		0.5	0.001 (0.25)	0.000 (0.00)	0.000 (0.00)	0.000 (0.00)	795
		0.7	0.001 (0.25)	0.000 (0.00)	0.001 (0.25)	0.001 (0.25)	790
10	0.3	0.3	0.002 (0.50)	0.002 (0.50)	0.003 (0.75)	0.003 (0.75)	316
		0.5	0.000 (0.00)	0.000 (0.00)	0.001 (0.25)	0.001 (0.25)	318
		0.7	0.001 (0.25)	0.002 (0.50)	0.001 (0.25)	0.001 (0.25)	315
	0.5	0.3	0.000 (0.00)	0.000 (0.00)	0.000 (0.00)	0.000 (0.00)	450
		0.5	0.001 (0.25)	0.002 (0.50)	0.002 (0.50)	0.001 (0.25)	460
		0.7	0.002 (0.50)	0.002 (0.50)	0.002 (0.50)	0.002 (0.50)	456
	0.7	0.3	0.003 (0.75)	0.002 (0.50)	0.001 (0.25)	0.002 (0.50)	775
		0.5	0.001 (0.25)	0.001 (0.25)	0.000 (0.00)	0.001 (0.25)	775
		0.7	0.001 (0.25)	0.000 (0.00)	0.000 (0.00)	0.001 (0.25)	768

Table S.4: Absolute and percent relative bias based on 10,000 simulations with different propensity score stratification (PSS) strategies. The true effect parameter is $\Delta\pi = 0.4$, the nominal type I error rate is fixed at $\alpha = 0.05$, the number of strata varies in $s \in \{5, 8, 10\}$, allocation proportion to the choice arm varies in $\theta \in \{0.3, 0.5, 0.7\}$, the preference rate for treatment 1 varies in $\phi \in \{0.3, 0.5, 0.7\}$. Results were summarized under the heterogeneous effect model.

Strata	θ	ϕ	Absolute Bias (Percent Relative Bias)				Sample size
			No PSS	PSS (X_1, X_2)	PSS (X_1)	PSS (X_2)	
5	0.3	0.3	0.207 (51.75)	0.030 (7.50)	0.075 (18.75)	0.299 (74.75)	1880
		0.5	0.270 (67.50)	0.036 (9.00)	0.036 (9.00)	0.338 (84.50)	1302
		0.7	0.443 (110.75)	0.058 (14.50)	0.007 (1.75)	0.513 (128.25)	1930
	0.5	0.3	0.206 (51.50)	0.027 (6.75)	0.078 (19.50)	0.297 (74.25)	1630
		0.5	0.275 (68.75)	0.036 (9.00)	0.034 (8.50)	0.341 (85.25)	1093
		0.7	0.443 (110.75)	0.055 (13.75)	0.010 (2.50)	0.508 (127.00)	1715
	0.7	0.3	0.211 (52.75)	0.026 (6.50)	0.078 (19.50)	0.301 (75.25)	2100
		0.5	0.279 (69.75)	0.034 (8.50)	0.037 (9.25)	0.342 (85.50)	1350
		0.7	0.450 (112.50)	0.052 (13.00)	0.011 (2.75)	0.510 (127.50)	2185
8	0.3	0.3	0.206 (51.50)	0.020 (5.00)	0.092 (23.00)	0.320 (80.00)	1870
		0.5	0.274 (68.50)	0.029 (7.25)	0.047 (11.75)	0.362 (90.50)	1300
		0.7	0.444 (111.00)	0.041 (10.25)	0.026 (6.50)	0.536 (134.00)	1945
	0.5	0.3	0.208 (52.00)	0.017 (4.25)	0.091 (22.75)	0.319 (79.75)	1660
		0.5	0.275 (68.75)	0.025 (6.25)	0.050 (12.50)	0.359 (89.75)	1088
		0.7	0.443 (110.75)	0.036 (9.00)	0.031 (7.75)	0.528 (132.00)	1670
	0.7	0.3	0.209 (52.25)	0.012 (3.00)	0.096 (24.00)	0.319 (79.75)	2090
		0.5	0.279 (69.75)	0.021 (5.25)	0.053 (13.25)	0.357 (89.25)	1370
		0.7	0.450 (112.50)	0.032 (8.00)	0.034 (8.50)	0.526 (131.50)	2140
10	0.3	0.3	0.207 (51.75)	0.019 (4.75)	0.093 (23.25)	0.324 (81.00)	1878
		0.5	0.273 (68.25)	0.026 (6.50)	0.050 (12.50)	0.365 (91.25)	1288
		0.7	0.445 (111.25)	0.039 (9.75)	0.030 (7.50)	0.548 (137.00)	1910
	0.5	0.3	0.207 (51.75)	0.015 (3.75)	0.097 (24.25)	0.320 (80.00)	1642
		0.5	0.276 (69.00)	0.023 (5.75)	0.051 (12.75)	0.362 (90.50)	1090
		0.7	0.445 (111.25)	0.033 (8.25)	0.035 (8.75)	0.534 (133.50)	1690
	0.7	0.3	0.211 (52.75)	0.012 (3.00)	0.098 (24.50)	0.322 (80.50)	2085
		0.5	0.279 (69.75)	0.018 (4.50)	0.057 (14.25)	0.359 (89.75)	1360
		0.7	0.451 (112.75)	0.029 (7.25)	0.038 (9.50)	0.530 (132.50)	2160

Table S.5: Absolute and percent relative bias based on 10,000 simulations with different propensity score stratification (PSS) strategies. The true effect parameter is $\Delta\nu = 0.4$, the nominal type I error rate is fixed at $\alpha = 0.05$, the number of strata varies in $s \in \{5, 8, 10\}$, allocation proportion to the choice arm varies in $\theta \in \{0.3, 0.5, 0.7\}$, the preference rate for treatment 1 varies in $\phi \in \{0.3, 0.5, 0.7\}$. Results were summarized under the heterogeneous effect model.

Strata	θ	ϕ	Absolute Bias (Percent Relative Bias)				Sample size
			No PSS	PSS (X_1, X_2)	PSS (X_1)	PSS (X_2)	
5	0.3	0.3	0.117 (29.25)	0.008 (2.00)	0.088 (22.00)	0.050 (12.50)	1870
		0.5	0.199 (49.75)	0.025 (6.25)	0.063 (15.75)	0.175 (43.75)	1300
		0.7	0.375 (93.75)	0.047 (11.75)	0.058 (14.50)	0.381 (95.25)	1935
	0.5	0.3	0.083 (20.75)	0.006 (1.50)	0.084 (21.00)	0.020 (5.00)	1642
		0.5	0.163 (40.75)	0.018 (4.50)	0.054 (13.50)	0.138 (34.50)	1085
		0.7	0.337 (84.25)	0.040 (10.00)	0.050 (12.50)	0.341 (85.25)	1700
	0.7	0.3	0.048 (12.00)	0.000 (0.00)	0.074 (18.50)	0.013 (3.25)	2120
		0.5	0.131 (32.75)	0.016 (4.00)	0.048 (12.00)	0.107 (26.75)	1368
		0.7	0.308 (77.00)	0.038 (9.50)	0.044 (11.00)	0.308 (77.00)	2165
8	0.3	0.3	0.117 (29.25)	0.004 (1.00)	0.090 (22.50)	0.036 (9.00)	1910
		0.5	0.199 (49.75)	0.018 (4.50)	0.058 (14.50)	0.174 (43.50)	1265
		0.7	0.374 (93.50)	0.034 (8.50)	0.047 (11.75)	0.389 (97.25)	1940
	0.5	0.3	0.082 (20.50)	0.001 (0.25)	0.083 (20.75)	0.005 (1.25)	1628
		0.5	0.164 (41.00)	0.014 (3.50)	0.052 (13.00)	0.138 (34.50)	1085
		0.7	0.336 (84.00)	0.027 (6.75)	0.036 (9.00)	0.347 (86.75)	1655
	0.7	0.3	0.048 (12.00)	0.000 (0.00)	0.077 (19.25)	0.027 (6.75)	2070
		0.5	0.131 (32.75)	0.010 (2.50)	0.044 (11.00)	0.104 (26.00)	1368
		0.7	0.308 (77.00)	0.024 (6.00)	0.030 (7.50)	0.312 (78.00)	2135
10	0.3	0.3	0.119 (29.75)	0.003 (0.75)	0.091 (22.75)	0.036 (9.00)	1875
		0.5	0.199 (49.75)	0.016 (4.00)	0.057 (14.25)	0.177 (44.25)	1275
		0.7	0.374 (93.50)	0.032 (8.00)	0.043 (10.75)	0.398 (99.50)	1940
	0.5	0.3	0.080 (20.00)	0.000 (0.00)	0.083 (20.75)	0.003 (0.75)	1665
		0.5	0.166 (41.50)	0.014 (3.50)	0.051 (12.75)	0.141 (35.25)	1075
		0.7	0.338 (84.50)	0.026 (6.50)	0.035 (8.75)	0.352 (88.00)	1690
	0.7	0.3	0.050 (12.50)	0.001 (0.25)	0.080 (20.00)	0.026 (6.50)	2085
		0.5	0.131 (32.75)	0.009 (2.25)	0.043 (10.75)	0.105 (26.25)	1345
		0.7	0.303 (75.75)	0.018 (4.50)	0.025 (6.25)	0.310 (77.50)	2145

Table S.6: Absolute and percent relative bias based on 10,000 simulations with different propensity score stratification (PSS) strategies. The true effect parameter is $\Delta\tau = 0.4$, the nominal type I error rate is fixed at $\alpha = 0.05$, the number of strata varies in $s \in \{5, 8, 10\}$, allocation proportion to the choice arm varies in $\theta \in \{0.3, 0.5, 0.7\}$, the preference rate for treatment 1 varies in $\phi \in \{0.3, 0.5, 0.7\}$. Results were summarized under the heterogeneous effect model.

Strata	θ	ϕ	Absolute Bias (Percent Relative Bias)				Sample size
			No PSS	PSS (X_1, X_2)	PSS (X_1)	PSS (X_2)	
5	0.3	0.3	0.025 (6.25)	0.001 (0.25)	0.010 (2.50)	0.017 (4.25)	385
		0.5	0.023 (5.75)	0.002 (0.50)	0.011 (2.75)	0.017 (4.25)	400
		0.7	0.022 (5.50)	0.002 (0.50)	0.009 (2.25)	0.014 (3.50)	401
	0.5	0.3	0.045 (11.25)	0.005 (1.25)	0.019 (4.75)	0.033 (8.25)	555
		0.5	0.038 (9.50)	0.001 (0.25)	0.016 (4.00)	0.026 (6.50)	572
		0.7	0.039 (9.75)	0.007 (1.75)	0.021 (5.25)	0.026 (6.50)	577
	0.7	0.3	0.063 (15.75)	0.009 (2.25)	0.028 (7.00)	0.047 (11.75)	968
		0.5	0.055 (13.75)	0.004 (1.00)	0.022 (5.50)	0.038 (9.50)	992
		0.7	0.051 (12.75)	0.002 (0.50)	0.022 (5.50)	0.035 (8.75)	984
8	0.3	0.3	0.025 (6.25)	0.001 (0.25)	0.010 (2.50)	0.017 (4.25)	384
		0.5	0.023 (5.75)	0.001 (0.25)	0.009 (2.25)	0.015 (3.75)	399
		0.7	0.023 (5.75)	0.002 (0.50)	0.009 (2.25)	0.014 (3.50)	397
	0.5	0.3	0.043 (10.75)	0.001 (0.25)	0.016 (4.00)	0.029 (7.25)	562
		0.5	0.040 (10.00)	0.003 (0.75)	0.017 (4.25)	0.025 (6.25)	564
		0.7	0.038 (9.50)	0.002 (0.50)	0.017 (4.25)	0.023 (5.75)	578
	0.7	0.3	0.058 (14.50)	0.000 (0.00)	0.020 (5.00)	0.039 (9.75)	965
		0.5	0.057 (14.25)	0.002 (0.50)	0.023 (5.75)	0.038 (9.50)	1005
		0.7	0.051 (12.75)	0.003 (0.75)	0.022 (5.50)	0.030 (7.50)	1017
10	0.3	0.3	0.027 (6.75)	0.002 (0.50)	0.010 (2.50)	0.019 (4.75)	388
		0.5	0.025 (6.25)	0.001 (0.25)	0.010 (2.50)	0.016 (4.00)	399
		0.7	0.023 (5.75)	0.001 (0.25)	0.010 (2.50)	0.014 (3.50)	396
	0.5	0.3	0.042 (10.50)	0.001 (0.25)	0.016 (4.00)	0.027 (6.75)	557
		0.5	0.043 (10.75)	0.003 (0.75)	0.017 (4.25)	0.028 (7.00)	570
		0.7	0.035 (8.75)	0.001 (0.25)	0.015 (3.75)	0.021 (5.25)	577
	0.7	0.3	0.062 (15.50)	0.003 (0.75)	0.023 (5.75)	0.041 (10.25)	968
		0.5	0.057 (14.25)	0.001 (0.25)	0.023 (5.75)	0.037 (9.25)	980
		0.7	0.050 (12.50)	0.001 (0.25)	0.021 (5.25)	0.030 (7.50)	975

Tables for variance ratio and mean squared error comparison (referenced in section 4.2.2 of main paper)

Table S.7: Variance ratio (empirical/predicted) and mean squared error for preference effect based on 10,000 simulations with different propensity score stratification (PSS) strategies. The true effect parameter is $\Delta\pi = 0.4$, the predicted variance is 0.020, the nominal type I error rate is fixed at $\alpha = 0.05$, the number of strata varies in $s \in \{5, 8, 10\}$, allocation proportion to the choice arm varies in $\theta \in \{0.3, 0.5, 0.7\}$, the preference rate for treatment 1 varies in $\phi \in \{0.3, 0.5, 0.7\}$. Results were summarized under the homogeneous effect model.

Strata	θ	ϕ	Variance Ratio			Mean Squared Error			Sample size	
			PSS (X_1, X_2)	PSS (X_1)	PSS (X_2)	PSS (X_1, X_2)	PSS (X_1)	PSS (X_2)		
5	0.3	0.3	1.050	1.050	1.000	0.021	0.022	0.028	1535	
		0.5	1.050	1.050	1.000	0.021	0.022	0.027	1066	
		0.7	1.150	1.100	1.100	0.024	0.024	0.032	1565	
	0.5	0.3	1.050	1.050	1.000	0.021	0.022	0.028	1315	
		0.5	1.050	1.050	1.050	0.021	0.022	0.027	878	
		0.7	1.050	1.050	1.050	0.021	0.023	0.029	1370	
	0.7	0.3	1.000	0.950	0.950	0.020	0.020	0.027	1700	
		0.5	1.000	1.000	0.950	0.020	0.021	0.024	1075	
		0.7	1.100	1.100	1.050	0.022	0.024	0.029	1720	
8	0.3	0.3	1.100	1.100	1.100	0.022	0.023	0.030	1510	
		0.5	1.100	1.100	1.050	0.022	0.024	0.028	1055	
		0.7	1.200	1.200	1.150	0.025	0.027	0.034	1585	
	0.5	0.3	1.050	1.050	1.050	0.021	0.022	0.029	1340	
		0.5	1.100	1.100	1.050	0.022	0.023	0.028	878	
		0.7	1.100	1.100	1.100	0.022	0.024	0.031	1335	
	0.7	0.3	1.000	1.000	1.000	0.020	0.021	0.027	1680	
		0.5	1.050	1.050	1.000	0.021	0.022	0.025	1090	
		0.7	1.100	1.100	1.100	0.022	0.024	0.030	1705	
	10	0.3	0.3	1.200	1.200	1.100	0.024	0.025	0.030	1532
			0.5	1.150	1.150	1.100	0.023	0.025	0.030	1055
			0.7	1.250	1.250	1.200	0.026	0.028	0.036	1568
0.5		0.3	1.100	1.100	1.050	0.022	0.023	0.029	1337	
		0.5	1.100	1.100	1.100	0.022	0.023	0.028	880	
		0.7	1.150	1.150	1.100	0.023	0.025	0.032	1370	
0.7		0.3	1.050	1.050	1.000	0.021	0.022	0.027	1665	
		0.5	1.100	1.100	1.050	0.022	0.023	0.027	1090	
		0.7	1.100	1.100	1.050	0.022	0.024	0.030	1715	

Table S.8: Variance ratio (empirical/predicted) and mean squared error for selection effect based on 10,000 simulations with different propensity score stratification (PSS) strategies. The true effect parameter is $\Delta\nu = 0.4$, the predicted variance is 0.020, the nominal type I error rate is fixed at $\alpha = 0.05$, the number of strata varies in $s \in \{5, 8, 10\}$, allocation proportion to the choice arm varies in $\theta \in \{0.3, 0.5, 0.7\}$, the preference rate for treatment 1 varies in $\phi \in \{0.3, 0.5, 0.7\}$. Results were summarized under the homogeneous effect model.

Strata	θ	ϕ	Variance Ratio			Mean Squared Error			Sample size	
			PSS (X_1, X_2)	PSS (X_1)	PSS (X_2)	PSS (X_1, X_2)	PSS (X_1)	PSS (X_2)		
5	0.3	0.3	1.050	1.050	1.050	0.021	0.021	0.022	1520	
		0.5	1.050	1.050	1.050	0.021	0.021	0.021	1070	
		0.7	1.050	1.100	1.050	0.021	0.023	0.023	1590	
	0.5	0.3	1.000	1.000	1.000	0.020	0.020	0.021	1335	
		0.5	1.000	1.050	1.000	0.020	0.021	0.020	880	
		0.7	1.050	1.100	1.100	0.021	0.022	0.024	1370	
	0.7	0.3	0.950	0.950	0.900	0.019	0.019	0.019	1718	
		0.5	1.000	1.000	1.000	0.020	0.020	0.020	1100	
		0.7	1.050	1.050	1.050	0.021	0.021	0.023	1710	
8	0.3	0.3	1.050	1.050	1.050	0.021	0.021	0.022	1565	
		0.5	1.100	1.150	1.100	0.022	0.023	0.022	1048	
		0.7	1.150	1.150	1.150	0.023	0.024	0.026	1600	
	0.5	0.3	1.050	1.050	1.050	0.021	0.021	0.022	1325	
		0.5	1.000	1.050	1.050	0.020	0.021	0.021	883	
		0.7	1.100	1.150	1.100	0.022	0.024	0.024	1345	
	0.7	0.3	1.000	1.000	1.000	0.020	0.020	0.021	1678	
		0.5	1.050	1.050	1.000	0.021	0.021	0.020	1092	
		0.7	1.100	1.100	1.050	0.022	0.022	0.023	1715	
	10	0.3	0.3	1.150	1.150	1.100	0.023	0.023	0.023	1530
			0.5	1.150	1.150	1.100	0.023	0.023	0.022	1052
			0.7	1.200	1.250	1.150	0.025	0.026	0.026	1610
0.5		0.3	1.050	1.100	1.100	0.021	0.022	0.023	1360	
		0.5	1.100	1.100	1.100	0.022	0.022	0.022	875	
		0.7	1.150	1.150	1.150	0.023	0.024	0.026	1375	
0.7		0.3	1.050	1.050	1.000	0.021	0.021	0.021	1675	
		0.5	1.050	1.050	1.000	0.021	0.021	0.020	1085	
		0.7	1.050	1.100	1.050	0.021	0.023	0.023	1710	

Table S.9: Variance ratio (empirical/predicted) and mean squared error for treatment effect based on 10,000 simulations with different propensity score stratification (PSS) strategies. The true effect parameter is $\Delta\tau = 0.4$, the predicted variance is 0.020, the nominal type I error rate is fixed at $\alpha = 0.05$, the number of strata varies in $s \in \{5, 8, 10\}$, allocation proportion to the choice arm varies in $\theta \in \{0.3, 0.5, 0.7\}$, the preference rate for treatment 1 varies in $\phi \in \{0.3, 0.5, 0.7\}$. Results were summarized under the homogeneous effect model.

Strata	θ	ϕ	Variance Ratio			Mean Squared Error			Sample size
			PSS (X_1, X_2)	PSS (X_1)	PSS (X_2)	PSS (X_1, X_2)	PSS (X_1)	PSS (X_2)	
5	0.3	0.3	1.000	1.000	1.000	0.020	0.020	0.020	314
		0.5	1.000	1.000	1.000	0.020	0.020	0.020	319
		0.7	1.000	1.050	1.050	0.020	0.021	0.021	317
	0.5	0.3	1.000	1.050	1.000	0.020	0.021	0.020	449
		0.5	1.000	1.050	1.000	0.020	0.021	0.020	451
		0.7	1.000	1.000	1.000	0.020	0.020	0.020	453
	0.7	0.3	1.050	1.050	1.000	0.021	0.021	0.020	787
		0.5	1.000	1.000	0.950	0.020	0.020	0.019	785
		0.7	1.000	1.000	1.000	0.020	0.020	0.020	773
8	0.3	0.3	1.050	1.050	1.050	0.021	0.021	0.021	312
		0.5	1.000	1.050	1.050	0.020	0.021	0.021	319
		0.7	1.000	1.050	1.050	0.020	0.021	0.021	314
	0.5	0.3	1.050	1.050	1.000	0.021	0.021	0.020	452
		0.5	1.050	1.050	1.050	0.021	0.021	0.021	449
		0.7	1.050	1.050	1.000	0.021	0.021	0.020	453
	0.7	0.3	1.050	1.050	1.000	0.021	0.021	0.020	770
		0.5	1.050	1.050	1.000	0.021	0.021	0.020	795
		0.7	1.000	1.000	0.950	0.020	0.020	0.019	790
10	0.3	0.3	1.050	1.050	1.050	0.021	0.021	0.021	316
		0.5	1.000	1.050	1.050	0.020	0.021	0.021	318
		0.7	1.000	1.050	1.050	0.020	0.021	0.021	315
	0.5	0.3	1.050	1.050	1.050	0.021	0.021	0.021	450
		0.5	1.050	1.050	1.000	0.021	0.021	0.020	460
		0.7	1.050	1.050	1.000	0.021	0.021	0.020	456
	0.7	0.3	1.100	1.100	1.050	0.022	0.022	0.021	775
		0.5	1.100	1.100	1.050	0.022	0.022	0.021	775
		0.7	1.100	1.100	1.050	0.022	0.022	0.021	768

Table S.10: Variance ratio (empirical/predicted) and mean squared error for preference effect based on 10,000 simulations with different propensity score stratification (PSS) strategies. The true effect parameter is $\Delta\pi = 0.4$, the predicted variance is 0.020, the nominal type I error rate is fixed at $\alpha = 0.05$, the number of strata varies in $s \in \{5, 8, 10\}$, allocation proportion to the choice arm varies in $\theta \in \{0.3, 0.5, 0.7\}$, the preference rate for treatment 1 varies in $\phi \in \{0.3, 0.5, 0.7\}$. Results were summarized under the heterogeneous effect model.

Strata	θ	ϕ	Variance Ratio			Mean Squared Error			Sample size
			PSS (X_1, X_2)	PSS (X_1)	PSS (X_2)	PSS (X_1, X_2)	PSS (X_1)	PSS (X_2)	
5	0.3	0.3	0.900	1.050	1.000	0.019	0.027	0.109	1880
		0.5	0.900	0.950	1.100	0.019	0.020	0.136	1302
		0.7	0.900	1.000	1.300	0.021	0.020	0.289	1930
	0.5	0.3	0.850	0.950	0.950	0.018	0.025	0.107	1630
		0.5	0.850	0.950	1.100	0.018	0.020	0.138	1093
		0.7	0.850	0.900	1.200	0.020	0.018	0.282	1715
	0.7	0.3	0.850	0.950	0.900	0.018	0.025	0.109	2100
		0.5	0.850	0.950	1.000	0.018	0.020	0.137	1350
		0.7	0.900	0.900	1.150	0.021	0.018	0.283	2185
8	0.3	0.3	0.950	1.050	1.000	0.019	0.029	0.122	1870
		0.5	0.900	1.000	1.100	0.019	0.022	0.153	1300
		0.7	0.950	1.000	1.350	0.021	0.021	0.314	1945
	0.5	0.3	0.850	1.000	0.950	0.017	0.028	0.121	1660
		0.5	0.900	0.950	1.100	0.019	0.022	0.151	1088
		0.7	0.950	0.950	1.250	0.020	0.020	0.304	1670
	0.7	0.3	0.900	0.950	0.950	0.018	0.028	0.121	2090
		0.5	0.900	0.950	1.000	0.018	0.022	0.147	1370
		0.7	0.900	0.950	1.200	0.019	0.020	0.301	2140
10	0.3	0.3	0.950	1.050	1.050	0.019	0.030	0.126	1878
		0.5	0.950	1.050	1.150	0.020	0.024	0.156	1288
		0.7	1.000	1.050	1.500	0.022	0.022	0.330	1910
	0.5	0.3	0.950	1.050	1.000	0.019	0.030	0.122	1642
		0.5	0.900	1.000	1.100	0.019	0.023	0.153	1090
		0.7	0.950	1.000	1.300	0.020	0.021	0.311	1690
	0.7	0.3	0.900	0.950	0.950	0.018	0.029	0.123	2085
		0.5	0.900	0.950	1.050	0.018	0.022	0.150	1360
		0.7	0.900	0.950	1.200	0.019	0.020	0.305	2160

Table S.11: Variance ratio (empirical/predicted) and mean squared error for selection effect based on 10,000 simulations with different propensity score stratification (PSS) strategies. The true effect parameter is $\Delta\nu = 0.4$, the predicted variance is 0.020, the nominal type I error rate is fixed at $\alpha = 0.05$, the number of strata varies in $s \in \{5, 8, 10\}$, allocation proportion to the choice arm varies in $\theta \in \{0.3, 0.5, 0.7\}$, the preference rate for treatment 1 varies in $\phi \in \{0.3, 0.5, 0.7\}$. Results were summarized under the heterogeneous effect model.

Strata	θ	ϕ	Variance Ratio			Mean Squared Error			Sample size	
			PSS (X_1, X_2)	PSS (X_1)	PSS (X_2)	PSS (X_1, X_2)	PSS (X_1)	PSS (X_2)		
5	0.3	0.3	1.000	1.000	1.050	0.020	0.028	0.024	1870	
		0.5	1.050	0.950	1.100	0.022	0.023	0.053	1300	
		0.7	1.050	0.950	1.350	0.023	0.022	0.172	1935	
	0.5	0.3	0.950	0.950	0.950	0.019	0.026	0.019	1642	
		0.5	1.050	0.950	1.150	0.021	0.022	0.042	1085	
		0.7	1.000	0.950	1.250	0.022	0.022	0.141	1700	
	0.7	0.3	0.950	0.900	0.950	0.019	0.023	0.019	2120	
		0.5	1.000	0.950	1.050	0.020	0.021	0.032	1368	
		0.7	0.950	0.900	1.150	0.020	0.020	0.118	2165	
8	0.3	0.3	1.000	1.000	1.000	0.020	0.028	0.021	1910	
		0.5	1.050	1.000	1.200	0.021	0.023	0.054	1265	
		0.7	1.050	1.000	1.400	0.022	0.022	0.179	1940	
	0.5	0.3	1.000	1.000	1.000	0.020	0.027	0.020	1628	
		0.5	1.050	1.000	1.150	0.021	0.023	0.042	1085	
		0.7	1.050	0.950	1.300	0.022	0.020	0.146	1655	
	0.7	0.3	1.000	0.950	0.950	0.020	0.025	0.020	2070	
		0.5	1.000	0.900	1.050	0.020	0.020	0.032	1368	
		0.7	1.000	0.950	1.200	0.021	0.020	0.121	2135	
	10	0.3	0.3	1.100	1.100	1.100	0.022	0.030	0.023	1875
			0.5	1.100	1.000	1.200	0.022	0.023	0.055	1275
			0.7	1.100	1.050	1.450	0.023	0.023	0.187	1940
0.5		0.3	1.000	0.950	1.000	0.020	0.026	0.020	1665	
		0.5	1.100	1.000	1.150	0.022	0.023	0.043	1075	
		0.7	1.050	0.900	1.250	0.022	0.019	0.149	1690	
0.7		0.3	1.000	0.950	0.950	0.020	0.025	0.020	2085	
		0.5	1.050	0.950	1.050	0.021	0.021	0.032	1345	
		0.7	1.050	0.950	1.200	0.021	0.020	0.120	2145	

Table S.12: Variance ratio (empirical/predicted) and mean squared error for treatment effect based on 10,000 simulations with different propensity score stratification (PSS) strategies. The true effect parameter is $\Delta\tau = 0.4$, the predicted variance is 0.020, the nominal type I error rate is fixed at $\alpha = 0.05$, the number of strata varies in $s \in \{5, 8, 10\}$, allocation proportion to the choice arm varies in $\theta \in \{0.3, 0.5, 0.7\}$, the preference rate for treatment 1 varies in $\phi \in \{0.3, 0.5, 0.7\}$. Results were summarized under the heterogeneous effect model.

Strata	θ	ϕ	Variance Ratio			Mean Squared Error			Sample size
			PSS (X_1, X_2)	PSS (X_1)	PSS (X_2)	PSS (X_1, X_2)	PSS (X_1)	PSS (X_2)	
5	0.3	0.3	1.050	0.950	1.000	0.021	0.019	0.020	385
		0.5	1.050	0.950	1.100	0.021	0.019	0.022	400
		0.7	1.000	0.900	1.100	0.020	0.018	0.022	401
	0.5	0.3	1.000	0.950	1.000	0.020	0.019	0.021	555
		0.5	1.000	0.900	1.000	0.020	0.018	0.021	572
		0.7	1.000	0.900	1.100	0.020	0.018	0.023	577
	0.7	0.3	1.000	0.900	0.950	0.020	0.019	0.021	968
		0.5	1.000	0.900	1.000	0.020	0.018	0.021	992
		0.7	1.050	0.900	1.050	0.021	0.018	0.022	984
8	0.3	0.3	1.000	0.950	1.000	0.020	0.019	0.020	384
		0.5	1.000	0.900	1.050	0.020	0.018	0.021	399
		0.7	1.000	0.900	1.100	0.020	0.018	0.022	397
	0.5	0.3	1.000	0.950	0.950	0.020	0.019	0.020	562
		0.5	1.000	0.900	1.050	0.020	0.018	0.022	564
		0.7	1.000	0.900	1.050	0.020	0.018	0.022	578
	0.7	0.3	1.050	0.950	0.950	0.021	0.019	0.021	965
		0.5	1.050	0.950	1.000	0.021	0.020	0.021	1005
		0.7	1.000	0.900	1.000	0.020	0.018	0.021	1017
10	0.3	0.3	1.000	0.950	1.000	0.020	0.019	0.020	388
		0.5	1.000	0.950	1.050	0.020	0.019	0.021	399
		0.7	1.050	0.950	1.150	0.021	0.019	0.023	396
	0.5	0.3	1.050	1.000	1.000	0.021	0.020	0.021	557
		0.5	1.050	0.950	1.050	0.021	0.019	0.022	570
		0.7	1.050	0.950	1.100	0.021	0.019	0.022	577
	0.7	0.3	1.050	0.950	0.950	0.021	0.020	0.021	968
		0.5	1.050	0.950	1.050	0.021	0.020	0.022	980
		0.7	1.050	0.950	1.050	0.021	0.019	0.022	975

Tables for power comparison (referenced in section 4.2.3 of main paper)

Table S.13: Empirical power for selection effect based on 10,000 simulations with different propensity score stratification (PSS) strategies compared with predicted power. The true effect parameter is $\Delta\nu = 0.4$, the nominal type I error rate is fixed at $\alpha = 0.05$, the number of strata varies in $s \in \{5, 8, 10\}$, allocation proportion to the choice arm varies in $\theta \in \{0.3, 0.5, 0.7\}$, the preference rate for treatment 1 varies in $\phi \in \{0.3, 0.5, 0.7\}$. Results were summarized under the homogeneous effect model.

Strata	θ	ϕ	Predicted Power		Empirical Power		Sample size
			PSS	PSS (X_1, X_2)	PSS (X_1)	PSS (X_2)	
5	0.3	0.3	0.800	0.763	0.743	0.706	1520
		0.5	0.800	0.794	0.796	0.808	1070
		0.7	0.800	0.809	0.832	0.873	1590
	0.5	0.3	0.800	0.784	0.765	0.720	1335
		0.5	0.800	0.794	0.793	0.800	880
		0.7	0.800	0.800	0.820	0.864	1370
	0.7	0.3	0.800	0.809	0.792	0.759	1718
		0.5	0.800	0.795	0.803	0.813	1100
		0.7	0.800	0.800	0.821	0.865	1710
8	0.3	0.3	0.800	0.760	0.740	0.710	1565
		0.5	0.800	0.771	0.775	0.790	1048
		0.7	0.800	0.796	0.820	0.863	1600
	0.5	0.3	0.800	0.778	0.755	0.712	1325
		0.5	0.800	0.791	0.789	0.800	883
		0.7	0.800	0.787	0.814	0.854	1345
	0.7	0.3	0.800	0.784	0.769	0.737	1678
		0.5	0.800	0.796	0.799	0.806	1092
		0.7	0.800	0.786	0.810	0.858	1715
10	0.3	0.3	0.800	0.730	0.707	0.677	1530
		0.5	0.800	0.769	0.771	0.789	1052
		0.7	0.800	0.781	0.807	0.857	1610
	0.5	0.3	0.800	0.767	0.743	0.698	1360
		0.5	0.800	0.771	0.776	0.784	875
		0.7	0.800	0.786	0.807	0.851	1375
	0.7	0.3	0.800	0.778	0.758	0.720	1675
		0.5	0.800	0.777	0.782	0.795	1085
		0.7	0.800	0.788	0.813	0.861	1710

Table S.14: Empirical power for treatment effect based on 10,000 simulations with different propensity score stratification (PSS) strategies compared with predicted power. The true effect parameter is $\Delta\tau = 0.4$, the nominal type I error rate is fixed at $\alpha = 0.05$, the number of strata varies in $s \in \{5, 8, 10\}$, allocation proportion to the choice arm varies in $\theta \in \{0.3, 0.5, 0.7\}$, the preference rate for treatment 1 varies in $\phi \in \{0.3, 0.5, 0.7\}$. Results were summarized under the homogeneous effect model.

Strata	θ	ϕ	Predicted Power		Empirical Power		Sample size
			PSS	PSS (X_1, X_2)	PSS (X_1)	PSS (X_2)	
5	0.3	0.3	0.800	0.796	0.796	0.800	314
		0.5	0.800	0.798	0.797	0.798	319
		0.7	0.800	0.802	0.804	0.807	317
	0.5	0.3	0.800	0.788	0.791	0.796	449
		0.5	0.800	0.788	0.788	0.798	451
		0.7	0.800	0.796	0.798	0.809	453
	0.7	0.3	0.800	0.793	0.800	0.814	787
		0.5	0.800	0.792	0.796	0.811	785
		0.7	0.800	0.785	0.790	0.805	773
8	0.3	0.3	0.800	0.784	0.786	0.787	312
		0.5	0.800	0.792	0.795	0.796	319
		0.7	0.800	0.778	0.777	0.776	314
	0.5	0.3	0.800	0.784	0.793	0.796	452
		0.5	0.800	0.776	0.780	0.784	449
		0.7	0.800	0.782	0.785	0.796	453
	0.7	0.3	0.800	0.775	0.780	0.799	770
		0.5	0.800	0.784	0.790	0.806	795
		0.7	0.800	0.784	0.795	0.813	790
10	0.3	0.3	0.800	0.784	0.786	0.791	316
		0.5	0.800	0.782	0.784	0.786	318
		0.7	0.800	0.784	0.782	0.785	315
	0.5	0.3	0.801	0.777	0.777	0.789	450
		0.5	0.800	0.778	0.778	0.790	460
		0.7	0.800	0.775	0.782	0.785	456
	0.7	0.3	0.800	0.760	0.771	0.793	775
		0.5	0.800	0.762	0.770	0.786	775
		0.7	0.800	0.758	0.770	0.792	768

Table S.15: Empirical power for preference effect based on 10,000 simulations with different propensity score stratification (PSS) strategies compared with predicted power. The true effect parameter is $\Delta\pi = 0.4$, the nominal type I error rate is fixed at $\alpha = 0.05$, the number of strata varies in $s \in \{5, 8, 10\}$, allocation proportion to the choice arm varies in $\theta \in \{0.3, 0.5, 0.7\}$, the preference rate for treatment 1 varies in $\phi \in \{0.3, 0.5, 0.7\}$. Results were summarized under the heterogeneous effect model.

Strata	θ	ϕ	Predicted Power		Empirical Power		Sample size
			PSS	PSS (X_1, X_2)	PSS (X_1)	PSS (X_2)	
5	0.3	0.3	0.800	0.845	0.631	0.998	1880
		0.5	0.800	0.864	0.740	0.998	1302
		0.7	0.800	0.894	0.808	1.000	1930
	0.5	0.3	0.800	0.855	0.635	0.999	1630
		0.5	0.800	0.875	0.751	0.999	1093
		0.7	0.800	0.907	0.823	1.000	1715
	0.7	0.3	0.800	0.863	0.656	0.999	2100
		0.5	0.800	0.871	0.754	0.999	1350
		0.7	0.800	0.904	0.822	1.000	2185
8	0.3	0.3	0.800	0.808	0.564	0.998	1870
		0.5	0.800	0.844	0.706	0.999	1300
		0.7	0.800	0.856	0.755	1.000	1945
	0.5	0.3	0.800	0.836	0.593	0.999	1660
		0.5	0.800	0.846	0.707	0.999	1088
		0.7	0.800	0.862	0.753	1.000	1670
	0.7	0.3	0.800	0.831	0.590	0.999	2090
		0.5	0.800	0.847	0.713	0.999	1370
		0.7	0.800	0.862	0.760	1.000	2140
10	0.3	0.3	0.800	0.799	0.541	0.998	1878
		0.5	0.800	0.827	0.672	0.998	1288
		0.7	0.800	0.836	0.722	1.000	1910
	0.5	0.3	0.800	0.811	0.560	0.998	1642
		0.5	0.800	0.833	0.689	0.999	1090
		0.7	0.800	0.848	0.741	1.000	1690
	0.7	0.3	0.800	0.826	0.584	0.999	2085
		0.5	0.800	0.832	0.697	1.000	1360
		0.7	0.800	0.852	0.743	1.000	2160

Table S.16: Empirical power for selection effect based on 10,000 simulations with different propensity score stratification (PSS) strategies compared with predicted power. The true effect parameter is $\Delta\nu = 0.4$, the nominal type I error rate is fixed at $\alpha = 0.05$, the number of strata varies in $s \in \{5, 8, 10\}$, allocation proportion to the choice arm varies in $\theta \in \{0.3, 0.5, 0.7\}$, the preference rate for treatment 1 varies in $\phi \in \{0.3, 0.5, 0.7\}$. Results were summarized under the heterogeneous effect model.

Strata	θ	ϕ	Predicted Power		Empirical Power		Sample size
			PSS	PSS (X_1, X_2)	PSS (X_1)	PSS (X_2)	
5	0.3	0.3	0.800	0.790	0.928	0.869	1870
		0.5	0.800	0.837	0.917	0.969	1300
		0.7	0.800	0.869	0.913	0.999	1935
	0.5	0.3	0.800	0.806	0.936	0.842	1642
		0.5	0.800	0.822	0.908	0.947	1085
		0.7	0.800	0.868	0.914	0.998	1700
	0.7	0.3	0.800	0.815	0.938	0.810	2120
		0.5	0.800	0.822	0.908	0.939	1368
		0.7	0.800	0.870	0.912	0.998	2165
8	0.3	0.3	0.800	0.773	0.928	0.855	1910
		0.5	0.800	0.798	0.893	0.960	1265
		0.7	0.800	0.834	0.893	0.999	1940
	0.5	0.3	0.800	0.786	0.926	0.802	1628
		0.5	0.800	0.802	0.893	0.942	1085
		0.7	0.800	0.830	0.879	0.997	1655
	0.7	0.3	0.800	0.790	0.929	0.764	2070
		0.5	0.800	0.810	0.900	0.936	1368
		0.7	0.800	0.836	0.882	0.996	2135
10	0.3	0.3	0.800	0.748	0.905	0.828	1875
		0.5	0.800	0.786	0.889	0.961	1275
		0.7	0.800	0.820	0.875	0.998	1940
	0.5	0.3	0.800	0.778	0.925	0.798	1665
		0.5	0.800	0.794	0.891	0.941	1075
		0.7	0.800	0.828	0.887	0.998	1690
	0.7	0.3	0.800	0.792	0.928	0.767	2085
		0.5	0.800	0.799	0.891	0.928	1345
		0.7	0.800	0.818	0.870	0.996	2145

Table S.17: Empirical power for treatment effect based on 10,000 simulations with different propensity score stratification (PSS) strategies compared with predicted power. The true effect parameter is $\Delta\tau = 0.4$, the nominal type I error rate is fixed at $\alpha = 0.05$, the number of strata varies in $s \in \{5, 8, 10\}$, allocation proportion to the choice arm varies in $\theta \in \{0.3, 0.5, 0.7\}$, the preference rate for treatment 1 varies in $\phi \in \{0.3, 0.5, 0.7\}$. Results were summarized under the heterogeneous effect model.

Strata	θ	ϕ	Predicted Power		Empirical Power		Sample size
			PSS	PSS (X_1, X_2)	PSS (X_1)	PSS (X_2)	
5	0.3	0.3	0.800	0.794	0.807	0.767	385
		0.5	0.801	0.796	0.817	0.752	400
		0.7	0.800	0.793	0.826	0.737	401
	0.5	0.3	0.801	0.793	0.795	0.739	555
		0.5	0.800	0.799	0.804	0.747	572
		0.7	0.801	0.784	0.800	0.716	577
	0.7	0.3	0.800	0.785	0.779	0.723	968
		0.5	0.800	0.786	0.791	0.728	992
		0.7	0.800	0.786	0.792	0.705	984
8	0.3	0.3	0.801	0.795	0.805	0.763	384
		0.5	0.800	0.797	0.821	0.755	399
		0.7	0.800	0.795	0.821	0.730	397
	0.5	0.3	0.801	0.793	0.798	0.752	562
		0.5	0.800	0.784	0.792	0.735	564
		0.7	0.800	0.797	0.811	0.724	578
	0.7	0.3	0.800	0.785	0.777	0.734	965
		0.5	0.800	0.787	0.789	0.732	1005
		0.7	0.800	0.787	0.800	0.729	1017
10	0.3	0.3	0.800	0.792	0.802	0.759	388
		0.5	0.800	0.791	0.807	0.748	399
		0.7	0.801	0.787	0.812	0.721	396
	0.5	0.3	0.801	0.782	0.787	0.750	557
		0.5	0.800	0.778	0.788	0.730	570
		0.7	0.800	0.787	0.808	0.719	577
	0.7	0.3	0.800	0.780	0.774	0.729	968
		0.5	0.800	0.767	0.768	0.705	980
		0.7	0.800	0.773	0.785	0.706	975

Tables for type I error rate comparison (referenced in section 4.2.4 of main paper)

Table S.18: Empirical type I error for selection effect based on 10,000 simulations with different propensity score stratification (PSS) strategies. The true effect parameter is $\Delta\nu = 0$, the nominal type I error rate is fixed at $\alpha = 0.05$, the number of strata varies in $s \in \{5, 8, 10\}$, allocation proportion to the choice arm varies in $\theta \in \{0.3, 0.5, 0.7\}$, the preference rate for treatment 1 varies in $\phi \in \{0.3, 0.5, 0.7\}$. Results were summarized under the homogeneous effect model.

Strata	θ	ϕ	Empirical Type I error rate				Sample size
			No PSS	PSS (X_1, X_2)	PSS (X_1)	PSS (X_2)	
5	0.3	0.3	0.054	0.043	0.045	0.050	1520
		0.5	0.052	0.048	0.048	0.051	1070
		0.7	0.059	0.049	0.049	0.055	1590
	0.5	0.3	0.058	0.046	0.049	0.054	1335
		0.5	0.052	0.053	0.052	0.052	880
		0.7	0.056	0.046	0.047	0.054	1370
	0.7	0.3	0.061	0.049	0.050	0.058	1718
		0.5	0.051	0.049	0.049	0.049	1100
		0.7	0.057	0.046	0.047	0.055	1710
8	0.3	0.3	0.059	0.043	0.044	0.053	1565
		0.5	0.051	0.047	0.047	0.046	1048
		0.7	0.057	0.045	0.043	0.052	1600
	0.5	0.3	0.056	0.048	0.049	0.051	1325
		0.5	0.053	0.050	0.051	0.052	883
		0.7	0.059	0.046	0.046	0.053	1345
	0.7	0.3	0.057	0.047	0.050	0.053	1678
		0.5	0.049	0.047	0.049	0.048	1092
		0.7	0.062	0.048	0.050	0.054	1715
10	0.3	0.3	0.057	0.041	0.040	0.049	1530
		0.5	0.050	0.042	0.044	0.045	1052
		0.7	0.058	0.040	0.044	0.049	1610
	0.5	0.3	0.062	0.047	0.049	0.053	1360
		0.5	0.051	0.045	0.049	0.047	875
		0.7	0.057	0.044	0.047	0.050	1375
	0.7	0.3	0.056	0.045	0.046	0.051	1675
		0.5	0.051	0.048	0.048	0.048	1085
		0.7	0.060	0.049	0.049	0.054	1710

Table S.19: Empirical type I error for treatment effect based on 10,000 simulations with different propensity score stratification (PSS) strategies. The true effect parameter is $\Delta\tau = 0$, the nominal type I error rate is fixed at $\alpha = 0.05$, the number of strata varies in $s \in \{5, 8, 10\}$, allocation proportion to the choice arm varies in $\theta \in \{0.3, 0.5, 0.7\}$, the preference rate for treatment 1 varies in $\phi \in \{0.3, 0.5, 0.7\}$. Results were summarized under the homogeneous effect model.

Strata	θ	ϕ	Empirical Type I error rate				Sample size
			No PSS	PSS (X_1, X_2)	PSS (X_1)	PSS (X_2)	
5	0.3	0.3	0.047	0.047	0.048	0.050	314
		0.5	0.046	0.047	0.045	0.045	319
		0.7	0.047	0.049	0.050	0.049	317
	0.5	0.3	0.050	0.048	0.049	0.048	449
		0.5	0.053	0.052	0.050	0.053	451
		0.7	0.051	0.050	0.048	0.050	453
	0.7	0.3	0.046	0.047	0.048	0.048	787
		0.5	0.054	0.054	0.055	0.054	785
		0.7	0.050	0.048	0.048	0.050	773
8	0.3	0.3	0.048	0.047	0.047	0.049	312
		0.5	0.052	0.050	0.049	0.050	319
		0.7	0.050	0.049	0.050	0.050	314
	0.5	0.3	0.052	0.052	0.054	0.051	452
		0.5	0.049	0.049	0.047	0.048	449
		0.7	0.049	0.050	0.050	0.051	453
	0.7	0.3	0.045	0.047	0.047	0.045	770
		0.5	0.048	0.050	0.050	0.048	795
		0.7	0.052	0.049	0.052	0.053	790
10	0.3	0.3	0.052	0.052	0.051	0.049	316
		0.5	0.047	0.047	0.047	0.050	318
		0.7	0.049	0.047	0.047	0.048	315
	0.5	0.3	0.051	0.049	0.050	0.052	450
		0.5	0.050	0.049	0.051	0.049	460
		0.7	0.051	0.050	0.050	0.052	456
	0.7	0.3	0.050	0.049	0.047	0.052	775
		0.5	0.047	0.046	0.049	0.047	775
		0.7	0.055	0.054	0.052	0.056	768

Table S.20: Empirical type I error for preference effect based on 10,000 simulations with different propensity score stratification (PSS) strategies. The true effect parameter is $\Delta\pi = 0$, the nominal type I error rate is fixed at $\alpha = 0.05$, the number of strata varies in $s \in \{5, 8, 10\}$, allocation proportion to the choice arm varies in $\theta \in \{0.3, 0.5, 0.7\}$, the preference rate for treatment 1 varies in $\phi \in \{0.3, 0.5, 0.7\}$. Results were summarized under the heterogeneous effect model.

Strata	θ	ϕ	Empirical Type I error rate				Sample size
			No PSS	PSS (X_1, X_2)	PSS (X_1)	PSS (X_2)	
5	0.3	0.3	0.307	0.036	0.077	0.546	1880
		0.5	0.456	0.039	0.060	0.623	1302
		0.7	0.822	0.047	0.051	0.904	1930
	0.5	0.3	0.301	0.037	0.082	0.552	1630
		0.5	0.463	0.038	0.058	0.637	1093
		0.7	0.846	0.044	0.048	0.923	1715
	0.7	0.3	0.336	0.041	0.084	0.601	2100
		0.5	0.498	0.041	0.064	0.660	1350
		0.7	0.857	0.051	0.051	0.928	2185
8	0.3	0.3	0.301	0.034	0.094	0.568	1870
		0.5	0.452	0.034	0.062	0.652	1300
		0.7	0.826	0.036	0.053	0.918	1945
	0.5	0.3	0.316	0.033	0.094	0.609	1660
		0.5	0.466	0.037	0.073	0.655	1088
		0.7	0.836	0.036	0.058	0.923	1670
	0.7	0.3	0.328	0.034	0.101	0.642	2090
		0.5	0.499	0.038	0.070	0.699	1370
		0.7	0.860	0.040	0.058	0.935	2140
10	0.3	0.3	0.300	0.032	0.103	0.556	1878
		0.5	0.448	0.032	0.065	0.646	1288
		0.7	0.818	0.033	0.054	0.913	1910
	0.5	0.3	0.314	0.036	0.104	0.597	1642
		0.5	0.467	0.031	0.066	0.663	1090
		0.7	0.834	0.036	0.059	0.919	1690
	0.7	0.3	0.331	0.034	0.109	0.639	2085
		0.5	0.495	0.034	0.073	0.688	1360
		0.7	0.859	0.038	0.062	0.935	2160

Table S.21: Empirical type I error for selection effect based on 10,000 simulations with different propensity score stratification (PSS) strategies. The true effect parameter is $\Delta\nu = 0$, the nominal type I error rate is fixed at $\alpha = 0.05$, the number of strata varies in $s \in \{5, 8, 10\}$, allocation proportion to the choice arm varies in $\theta \in \{0.3, 0.5, 0.7\}$, the preference rate for treatment 1 varies in $\phi \in \{0.3, 0.5, 0.7\}$. Results were summarized under the heterogeneous effect model.

Strata	θ	ϕ	Empirical Type I error rate				Sample size
			No PSS	PSS (X_1, X_2)	PSS (X_1)	PSS (X_2)	
5	0.3	0.3	0.128	0.040	0.091	0.063	1870
		0.5	0.260	0.049	0.067	0.196	1300
		0.7	0.670	0.051	0.063	0.650	1935
	0.5	0.3	0.090	0.041	0.087	0.050	1642
		0.5	0.198	0.049	0.061	0.149	1085
		0.7	0.603	0.054	0.062	0.588	1700
	0.7	0.3	0.064	0.044	0.088	0.047	2120
		0.5	0.147	0.053	0.067	0.108	1368
		0.7	0.535	0.055	0.060	0.525	2165
8	0.3	0.3	0.132	0.039	0.083	0.052	1910
		0.5	0.256	0.048	0.057	0.179	1265
		0.7	0.679	0.044	0.050	0.650	1940
	0.5	0.3	0.098	0.045	0.090	0.046	1628
		0.5	0.196	0.053	0.064	0.140	1085
		0.7	0.582	0.048	0.051	0.561	1655
	0.7	0.3	0.065	0.046	0.079	0.052	2070
		0.5	0.148	0.054	0.063	0.103	1368
		0.7	0.538	0.050	0.053	0.525	2135
10	0.3	0.3	0.135	0.041	0.087	0.049	1875
		0.5	0.260	0.046	0.060	0.179	1275
		0.7	0.675	0.047	0.051	0.646	1940
	0.5	0.3	0.096	0.043	0.086	0.048	1665
		0.5	0.198	0.046	0.060	0.137	1075
		0.7	0.593	0.047	0.052	0.570	1690
	0.7	0.3	0.065	0.043	0.081	0.051	2085
		0.5	0.147	0.050	0.059	0.099	1345
		0.7	0.531	0.051	0.055	0.519	2145

Table S.22: Empirical type I error for treatment effect based on 10,000 simulations with different propensity score stratification (PSS) strategies. The true effect parameter is $\Delta\tau = 0$, the nominal type I error rate is fixed at $\alpha = 0.05$, the number of strata varies in $s \in \{5, 8, 10\}$, allocation proportion to the choice arm varies in $\theta \in \{0.3, 0.5, 0.7\}$, the preference rate for treatment 1 varies in $\phi \in \{0.3, 0.5, 0.7\}$. Results were summarized under the heterogeneous effect model.

Strata	θ	ϕ	Empirical Type I error rate				Sample size
			No PSS	PSS (X_1, X_2)	PSS (X_1)	PSS (X_2)	
5	0.3	0.3	0.051	0.047	0.049	0.050	385
		0.5	0.055	0.050	0.049	0.053	400
		0.7	0.054	0.048	0.049	0.049	401
	0.5	0.3	0.062	0.050	0.051	0.057	555
		0.5	0.056	0.048	0.049	0.052	572
		0.7	0.057	0.051	0.051	0.051	577
	0.7	0.3	0.074	0.049	0.054	0.063	968
		0.5	0.068	0.049	0.054	0.061	992
		0.7	0.065	0.049	0.056	0.056	984
8	0.3	0.3	0.053	0.050	0.049	0.052	384
		0.5	0.053	0.052	0.052	0.049	399
		0.7	0.052	0.050	0.050	0.052	397
	0.5	0.3	0.063	0.054	0.054	0.057	562
		0.5	0.058	0.053	0.053	0.055	564
		0.7	0.057	0.047	0.050	0.053	578
	0.7	0.3	0.073	0.050	0.052	0.061	965
		0.5	0.066	0.045	0.050	0.055	1005
		0.7	0.062	0.047	0.050	0.053	1017
10	0.3	0.3	0.052	0.047	0.051	0.051	388
		0.5	0.050	0.050	0.049	0.048	399
		0.7	0.050	0.049	0.050	0.047	396
	0.5	0.3	0.063	0.047	0.051	0.053	557
		0.5	0.059	0.048	0.052	0.056	570
		0.7	0.056	0.049	0.049	0.050	577
	0.7	0.3	0.072	0.048	0.055	0.057	968
		0.5	0.064	0.045	0.052	0.055	980
		0.7	0.062	0.049	0.055	0.053	975

Tables for Bias, Variance, Power, Type I Error comparison when sample size is approximately half the original sample size (referenced in section 4.3.1 of main paper)

Table S.23: Absolute and percent relative bias based on 10,000 simulations with different propensity score stratification (PSS) strategies. The true effect parameter is $\Delta\pi = 0.6$, the nominal type I error rate is fixed at $\alpha = 0.05$, the number of strata varies in $s \in \{5, 8, 10\}$, allocation proportion to the choice arm varies in $\theta \in \{0.3, 0.5, 0.7\}$, the preference rate for treatment 1 varies in $\phi \in \{0.3, 0.5, 0.7\}$. Results were summarized under the homogeneous effect model.

Strata	ϕ	θ	Absolute Bias (Percent Relative Bias)				Sample size
			No PSS	PSS (X_1, X_2)	PSS (X_1)	PSS (X_2)	
5	0.3	0.3	0.109 (18.21)	0.018 (2.95)	0.044 (7.25)	0.097 (16.16)	716
		0.5	0.108 (17.96)	0.014 (2.32)	0.040 (6.72)	0.092 (15.34)	628
		0.7	0.109 (18.23)	0.015 (2.53)	0.040 (6.60)	0.092 (15.40)	825
	0.5	0.3	0.095 (15.81)	0.036 (5.98)	0.056 (9.38)	0.098 (16.35)	510
		0.5	0.095 (15.91)	0.026 (4.37)	0.048 (7.98)	0.092 (15.27)	426
		0.7	0.088 (14.71)	0.013 (2.14)	0.034 (5.69)	0.079 (13.10)	530
	0.7	0.3	0.117 (19.45)	0.054 (9.02)	0.079 (13.13)	0.130 (21.59)	760
		0.5	0.109 (18.23)	0.032 (5.31)	0.057 (9.54)	0.110 (18.26)	670
		0.7	0.107 (17.85)	0.021 (3.46)	0.046 (7.69)	0.097 (16.22)	845
8	0.3	0.3	0.108 (18.04)	0.017 (2.85)	0.047 (7.81)	0.101 (16.76)	705
		0.5	0.109 (18.17)	0.013 (2.18)	0.042 (6.94)	0.095 (15.81)	640
		0.7	0.109 (18.18)	0.011 (1.77)	0.037 (6.10)	0.090 (14.96)	820
	0.5	0.3	0.099 (16.51)	0.055 (9.11)	0.078 (12.98)	0.120 (20.08)	503
		0.5	0.095 (15.85)	0.035 (5.87)	0.056 (9.40)	0.103 (17.10)	425
		0.7	0.091 (15.11)	0.018 (2.95)	0.039 (6.56)	0.084 (13.94)	538
	0.7	0.3	0.108 (17.92)	0.070 (11.68)	0.096 (16.01)	0.149 (24.81)	770
		0.5	0.111 (18.56)	0.048 (8.05)	0.076 (12.69)	0.127 (21.18)	655
		0.7	0.109 (18.18)	0.027 (4.46)	0.053 (8.78)	0.105 (17.56)	840
10	0.3	0.3	0.109 (18.18)	0.021 (3.54)	0.049 (8.19)	0.107 (17.79)	715
		0.5	0.105 (17.56)	0.010 (1.70)	0.038 (6.35)	0.094 (15.67)	638
		0.7	0.103 (17.19)	0.006 (0.93)	0.032 (5.27)	0.086 (14.32)	810
	0.5	0.3	0.096 (15.96)	0.066 (11.03)	0.089 (14.82)	0.133 (22.18)	502
		0.5	0.089 (14.75)	0.036 (5.97)	0.060 (9.96)	0.103 (17.22)	428
		0.7	0.095 (15.87)	0.023 (3.75)	0.045 (7.58)	0.091 (15.23)	538
	0.7	0.3	0.112 (18.72)	0.097 (16.17)	0.126 (20.95)	0.180 (29.93)	760
		0.5	0.109 (18.25)	0.058 (9.66)	0.086 (14.31)	0.138 (22.99)	670
		0.7	0.112 (18.69)	0.032 (5.27)	0.058 (9.69)	0.113 (18.80)	845

Table S.24: Variance ratio (empirical/predicted) and mean squared error for preference effect based on 10,000 simulations with different propensity score stratification (PSS) strategies. The true effect parameter is $\Delta\pi = 0.6$, the predicted variance is 0.046, the nominal type I error rate is fixed at $\alpha = 0.05$, the number of strata varies in $s \in \{5, 8, 10\}$, allocation proportion to the choice arm varies in $\theta \in \{0.3, 0.5, 0.7\}$, the preference rate for treatment 1 varies in $\phi \in \{0.3, 0.5, 0.7\}$. Results were summarized under the homogeneous effect model.

Strata	ϕ	θ	Variance Ratio			Mean Squared Error			Sample size
			PSS (X_1, X_2)	PSS (X_1)	PSS (X_2)	PSS (X_1, X_2)	PSS (X_1)	PSS (X_2)	
5	0.3	0.3	1.133	1.115	1.049	0.052	0.053	0.058	716
		0.5	1.040	1.038	1.010	0.048	0.049	0.055	628
		0.7	0.958	0.951	0.922	0.044	0.045	0.051	825
	0.5	0.3	1.111	1.112	1.041	0.052	0.054	0.058	510
		0.5	1.088	1.085	1.067	0.051	0.052	0.057	426
		0.7	1.066	1.047	1.011	0.049	0.049	0.053	530
	0.7	0.3	1.272	1.247	1.181	0.061	0.064	0.071	760
		0.5	1.120	1.119	1.101	0.053	0.055	0.063	670
		0.7	1.093	1.076	1.040	0.051	0.052	0.057	845
8	0.3	0.3	1.367	1.330	1.240	0.063	0.063	0.067	705
		0.5	1.108	1.116	1.063	0.051	0.053	0.058	640
		0.7	1.011	0.993	0.952	0.047	0.047	0.052	820
	0.5	0.3	1.290	1.264	1.190	0.062	0.064	0.069	503
		0.5	1.188	1.182	1.134	0.056	0.058	0.063	425
		0.7	1.098	1.087	1.045	0.051	0.052	0.055	538
	0.7	0.3	1.471	1.492	1.399	0.073	0.078	0.087	770
		0.5	1.289	1.289	1.249	0.062	0.065	0.074	655
		0.7	1.152	1.149	1.123	0.054	0.056	0.063	840
10	0.3	0.3	1.496	1.479	1.404	0.069	0.070	0.076	715
		0.5	1.191	1.190	1.145	0.055	0.056	0.062	638
		0.7	1.036	1.033	0.983	0.048	0.048	0.053	810
	0.5	0.3	1.426	1.415	1.350	0.070	0.073	0.080	502
		0.5	1.212	1.216	1.151	0.057	0.059	0.064	428
		0.7	1.147	1.133	1.081	0.053	0.054	0.058	538
	0.7	0.3	1.695	1.695	1.678	0.087	0.094	0.109	760
		0.5	1.408	1.412	1.327	0.068	0.072	0.080	670
		0.7	1.217	1.212	1.174	0.057	0.059	0.067	845

Table S.25: Empirical power for preference effect based on 10,000 simulations with different propensity score stratification (PSS) strategies compared with predicted power. The true effect parameter is $\Delta\pi = 0.6$, the nominal type I error rate is fixed at $\alpha = 0.05$, the number of strata varies in $s \in \{5, 8, 10\}$, allocation proportion to the choice arm varies in $\theta \in \{0.3, 0.5, 0.7\}$, the preference rate for treatment 1 varies in $\phi \in \{0.3, 0.5, 0.7\}$. Results were summarized under the homogeneous effect model.

Strata	ϕ	θ	Predicted Power		Empirical Power		Sample size
			PSS	PSS (X_1, X_2)	PSS (X_1)	PSS (X_2)	
5	0.3	0.3	0.800	0.766	0.799	0.867	716
		0.5	0.801	0.791	0.826	0.883	628
		0.7	0.800	0.831	0.857	0.913	825
	0.5	0.3	0.801	0.791	0.823	0.883	510
		0.5	0.800	0.799	0.824	0.878	426
		0.7	0.800	0.792	0.822	0.879	530
	0.7	0.3	0.800	0.796	0.831	0.898	760
		0.5	0.800	0.802	0.831	0.892	670
		0.7	0.800	0.798	0.830	0.895	845
8	0.3	0.3	0.800	0.696	0.736	0.814	705
		0.5	0.800	0.758	0.795	0.861	640
		0.7	0.801	0.794	0.830	0.895	820
	0.5	0.3	0.800	0.758	0.794	0.856	503
		0.5	0.800	0.770	0.799	0.861	425
		0.7	0.801	0.778	0.814	0.875	538
	0.7	0.3	0.800	0.752	0.791	0.866	770
		0.5	0.801	0.765	0.801	0.870	655
		0.7	0.801	0.771	0.806	0.878	840
10	0.3	0.3	0.800	0.693	0.733	0.802	715
		0.5	0.800	0.719	0.762	0.836	638
		0.7	0.800	0.766	0.807	0.879	810
	0.5	0.3	0.800	0.737	0.769	0.833	502
		0.5	0.801	0.740	0.777	0.845	428
		0.7	0.801	0.772	0.800	0.868	538
	0.7	0.3	0.800	0.738	0.775	0.847	760
		0.5	0.800	0.749	0.785	0.860	670
		0.7	0.800	0.764	0.804	0.869	845

Table S.26: Empirical type I error for preference effect based on 10,000 simulations with different propensity score stratification (PSS) strategies. The true effect parameter is $\Delta\pi = 0$, the nominal type I error rate is fixed at $\alpha = 0.05$, the number of strata varies in $s \in \{5, 8, 10\}$, allocation proportion to the choice arm varies in $\theta \in \{0.3, 0.5, 0.7\}$, the preference rate for treatment 1 varies in $\phi \in \{0.3, 0.5, 0.7\}$. Results were summarized under the homogeneous effect model.

Strata	ϕ	θ	Empirical Type I error rate				Sample size
			No PSS	PSS (X_1, X_2)	PSS (X_1)	PSS (X_2)	
5	0.3	0.3	0.090	0.040	0.044	0.062	716
		0.5	0.084	0.045	0.048	0.062	628
		0.7	0.087	0.050	0.054	0.069	825
	0.5	0.3	0.077	0.046	0.047	0.060	510
		0.5	0.075	0.049	0.051	0.061	426
		0.7	0.071	0.050	0.053	0.061	530
	0.7	0.3	0.074	0.040	0.041	0.058	760
		0.5	0.078	0.043	0.046	0.064	670
		0.7	0.081	0.043	0.049	0.066	845
8	0.3	0.3	0.080	0.032	0.035	0.048	705
		0.5	0.085	0.038	0.042	0.055	640
		0.7	0.083	0.046	0.047	0.062	820
	0.5	0.3	0.076	0.038	0.039	0.051	503
		0.5	0.077	0.046	0.047	0.060	425
		0.7	0.077	0.051	0.050	0.062	538
	0.7	0.3	0.074	0.034	0.036	0.049	770
		0.5	0.077	0.041	0.045	0.058	655
		0.7	0.084	0.046	0.053	0.069	840
10	0.3	0.3	0.087	0.032	0.034	0.045	715
		0.5	0.087	0.034	0.039	0.052	638
		0.7	0.086	0.049	0.050	0.059	810
	0.5	0.3	0.071	0.034	0.038	0.046	502
		0.5	0.074	0.039	0.041	0.053	428
		0.7	0.075	0.045	0.050	0.061	538
	0.7	0.3	0.070	0.028	0.031	0.043	760
		0.5	0.074	0.038	0.040	0.053	670
		0.7	0.077	0.045	0.047	0.062	845

Tables for Bias, Variance, Power, Type I Error comparison varying degree of positivity violation in PS model (referenced in section 4.3.2 of main paper)

Table S.27: Absolute and percent relative bias based on 10,000 simulations with different propensity score stratification (PSS) strategies. The true effect parameter is $\Delta\pi = 0.4$, the nominal type I error rate is fixed at $\alpha = 0.05$, the number of strata is fixed at 5, allocation proportion to the choice arm is fixed at $\theta = 0.5$, the preference rate for treatment 1 varies in $\phi \in \{0.3, 0.5, 0.7\}$, the multiplier γ in the PS model varies in $\gamma \in \{4, 6, 8\}$ to represent different degree of positivity violation. Results were summarized under the homogeneous effect model.

ϕ	γ	Absolute Bias (Percent Relative Bias)				Sample size
		No PSS	PSS (X_1, X_2)	PSS (X_1)	PSS (X_2)	
0.3	4	0.313 (78.21)	0.036 (8.92)	0.128 (32.12)	0.263 (65.63)	2060
	6	0.372 (92.88)	0.047 (11.69)	0.176 (43.93)	0.316 (78.88)	3320
	8	0.402 (100.51)	0.051 (12.80)	0.206 (51.54)	0.343 (85.75)	6400
0.5	4	0.263 (65.68)	0.034 (8.52)	0.111 (27.83)	0.222 (55.59)	1360
	6	0.313 (78.15)	0.042 (10.58)	0.150 (37.39)	0.266 (66.49)	2330
	8	0.339 (84.66)	0.048 (12.00)	0.176 (43.92)	0.290 (72.39)	4350
0.7	4	0.313 (78.29)	0.042 (10.49)	0.135 (33.68)	0.266 (66.53)	2140
	6	0.371 (92.64)	0.051 (12.69)	0.178 (44.60)	0.316 (79.01)	3400
	8	0.403 (100.69)	0.060 (14.90)	0.211 (52.65)	0.345 (86.16)	6800

Table S.28: Variance ratio (empirical/predicted) and mean squared error for preference effect based on 10,000 simulations with different propensity score stratification (PSS) strategies. The true effect parameter is $\Delta\pi = 0.4$, the predicted variance is 0.020, the nominal type I error rate is fixed at $\alpha = 0.05$, the number of strata is fixed at 5, allocation proportion to the choice arm is fixed at $\theta = 0.5$, the preference rate for treatment 1 varies in $\phi \in \{0.3, 0.5, 0.7\}$, the multiplier γ in the PS model varies in $\gamma \in \{4, 6, 8\}$ to represent different degree of positivity violation. Results were summarized under the homogeneous effect model.

ϕ	γ	Variance Ratio			Mean Squared Error			Sample size
		PSS (X_1, X_2)	PSS (X_1)	PSS (X_2)	PSS (X_1, X_2)	PSS (X_1)	PSS (X_2)	
0.3	4	1.042	0.902	0.653	0.022	0.035	0.082	2060
	6	1.107	0.773	0.409	0.024	0.046	0.108	3320
	8	1.113	0.565	0.217	0.025	0.054	0.122	6400
0.5	4	1.059	0.947	0.682	0.022	0.031	0.063	1360
	6	1.061	0.734	0.396	0.023	0.037	0.079	2330
	8	1.119	0.573	0.213	0.025	0.042	0.088	4350
0.7	4	1.077	0.953	0.687	0.023	0.037	0.085	2140
	6	1.197	0.830	0.438	0.027	0.048	0.109	3400
	8	1.159	0.581	0.222	0.027	0.056	0.123	6800

Table S.29: Empirical power for preference effect based on 10,000 simulations with different propensity score stratification (PSS) strategies compared with predicted power. The true effect parameter is $\Delta\pi = 0.4$, the nominal type I error rate is fixed at $\alpha = 0.05$, the number of strata is fixed at 5, allocation proportion to the choice arm is fixed at $\theta = 0.5$, the preference rate for treatment 1 varies in $\phi \in \{0.3, 0.5, 0.7\}$, the multiplier γ in the PS model varies in $\gamma \in \{4, 6, 8\}$ to represent different degree of positivity violation. Results were summarized under the homogeneous effect model.

ϕ	γ	Predicted Power		Empirical Power		Sample size
		PSS	PSS (X_1, X_2)	PSS (X_1)	PSS (X_2)	
0.3	4	0.800	0.852	0.969	1.000	2060
	6	0.800	0.843	0.994	1.000	3320
	8	0.801	0.849	1.000	1.000	6400
0.5	4	0.801	0.842	0.960	1.000	1360
	6	0.800	0.851	0.994	1.000	2330
	8	0.801	0.849	1.000	1.000	4350
0.7	4	0.800	0.856	0.974	1.000	2140
	6	0.801	0.843	0.995	1.000	3400
	8	0.800	0.867	1.000	1.000	6800

Table S.30: Empirical type I error for preference effect based on 10,000 simulations with different propensity score stratification (PSS) strategies. The true effect parameter is $\Delta\pi = 0$, the nominal type I error rate is fixed at $\alpha = 0.05$, the number of strata is fixed at 5, allocation proportion to the choice arm is fixed at $\theta = 0.5$, the preference rate for treatment 1 varies in $\phi \in \{0.3, 0.5, 0.7\}$, the multiplier γ in the PS model varies in $\gamma \in \{4, 6, 8\}$ to represent different degree of positivity violation. Results were summarized under the homogeneous effect model.

ϕ	γ	Empirical Type I error rate				Sample size
		No PSS	PSS (X_1, X_2)	PSS (X_1)	PSS (X_2)	
0.3	4	0.792	0.053	0.152	0.605	2060
	6	0.989	0.061	0.286	0.932	3320
	8	1.000	0.067	0.503	1.000	6400
0.5	4	0.648	0.054	0.119	0.472	1360
	6	0.957	0.060	0.227	0.848	2330
	8	1.000	0.062	0.387	0.993	4350
0.7	4	0.808	0.052	0.155	0.630	2140
	6	0.990	0.056	0.274	0.932	3400
	8	1.000	0.065	0.514	1.000	6800

Figure for propensity score overlap varying degree of positivity violation in PS model in simulation study (referenced in section 4.3.2 of main paper)

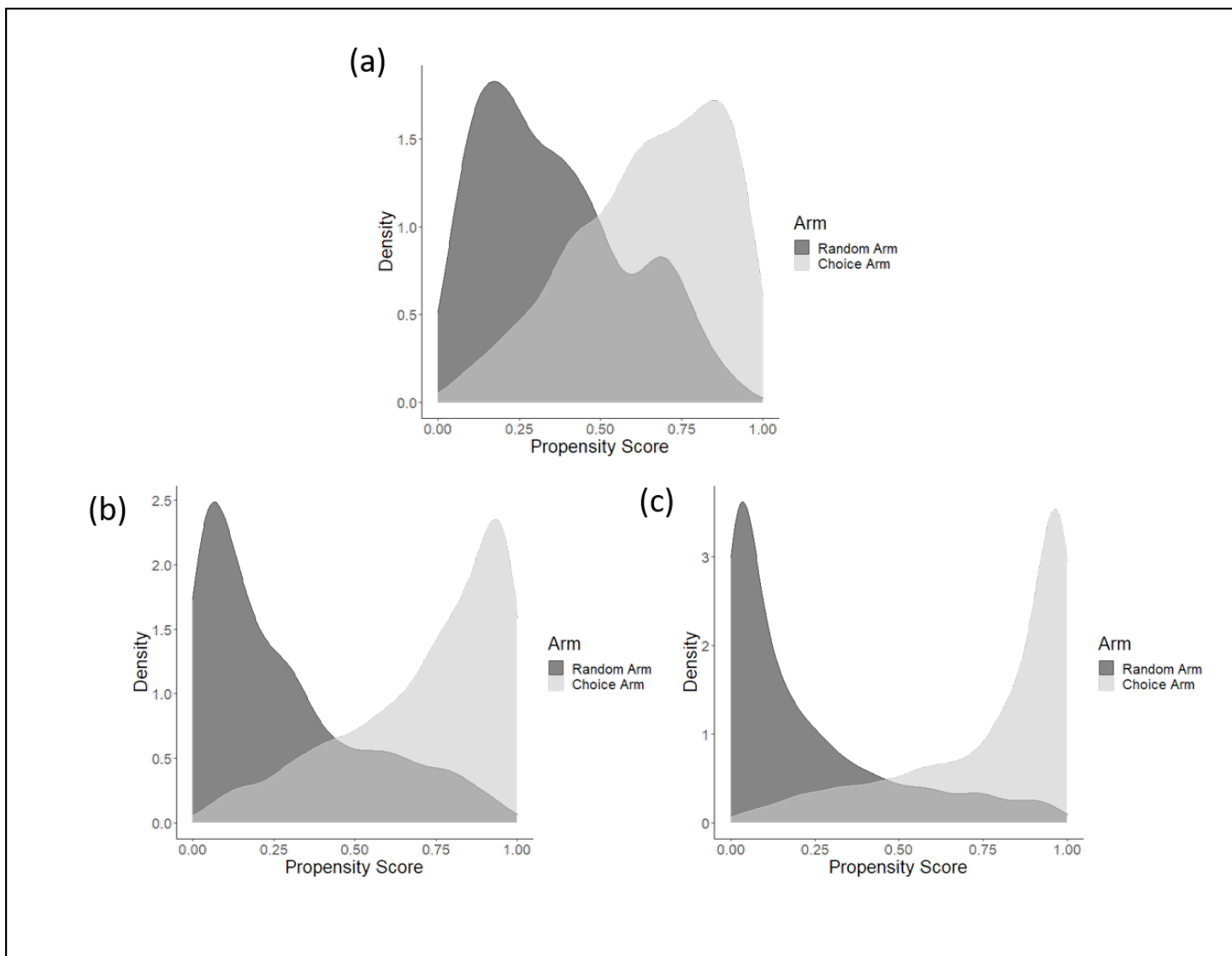


Figure S.1: Distributions of propensity score by choice arm and random arm in the simulation study with the multiplier γ in the PS model varies in $\gamma \in \{4, 6, 8\}$ to represent different degree of positivity violation: (a) $\gamma = 4$ (b) $\gamma = 6$ (c) $\gamma = 8$. The black bar indicate the random arm and the grey bar indicate the choice arm. Allocation proportion to the choice arm is fixed at $\theta = 0.5$.

Figure for examining the gradual impact of unmeasured confounders on absolute bias, variance ratio, empirical power and empirical type I error (referenced in section 4.3.3 of main paper)

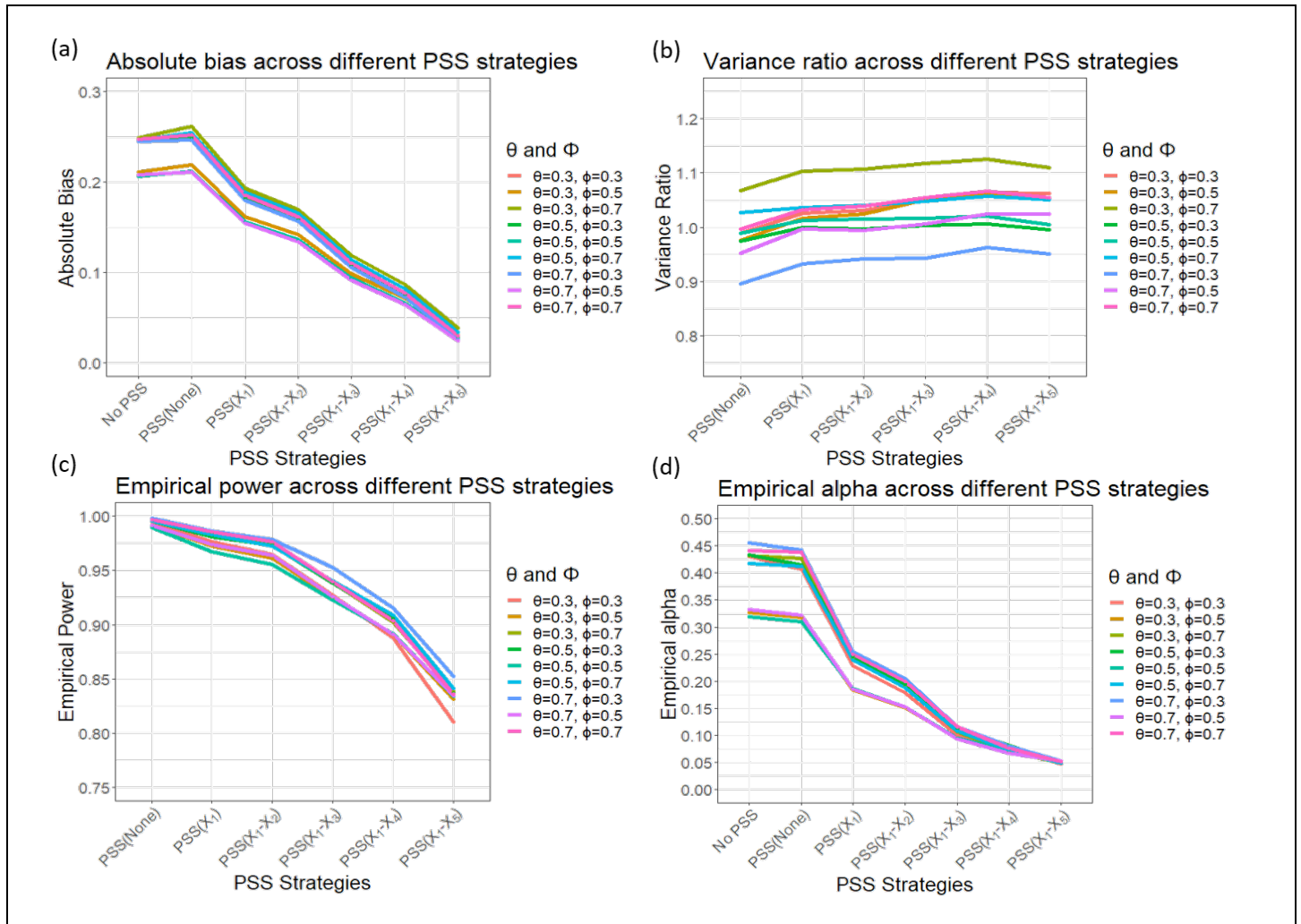


Figure S.2: The evaluation of the preference effect estimator with different degree of unmeasured confounders used for propensity score generation and stratification (PSS): (a) Absolute bias across different PSS strategies. (b) Variance ratio across different PSS strategies. (c) Empirical power across different PSS strategies. (d) Empirical type I error across different PSS strategies. There are in total five covariates (X_1 to X_5) and No PSS, PSS that used no covariates (PSS(None)), PSS that used covariates from forward selection are considered (e.g. PSS(X_1-X_5) is the PSS strategy that there is no unmeasured confounders, PSS(X_1-X_3) is the PSS strategy that X_4 and X_5 are left out as unmeasured confounders). For (a), (b), (c), the true effect parameter is $\Delta\pi = 0.4$, the predicted variance is 0.020, and the predicted power is 0.800. For (d), the true effect parameter is $\Delta\pi = 0$. For all figures, the nominal type I error rate is fixed at $\alpha = 0.05$, number of strata is fixed at 5, allocation proportion to the choice arm varies in $\theta \in \{0.3, 0.5, 0.7\}$, the preference rate for treatment 1 varies in $\phi \in \{0.3, 0.5, 0.7\}$, number of simulation is 10,000. Results were summarized under the homogeneous effect model.

Figure for propensity score overlap for the motivating example (referenced in section 5 of main paper)

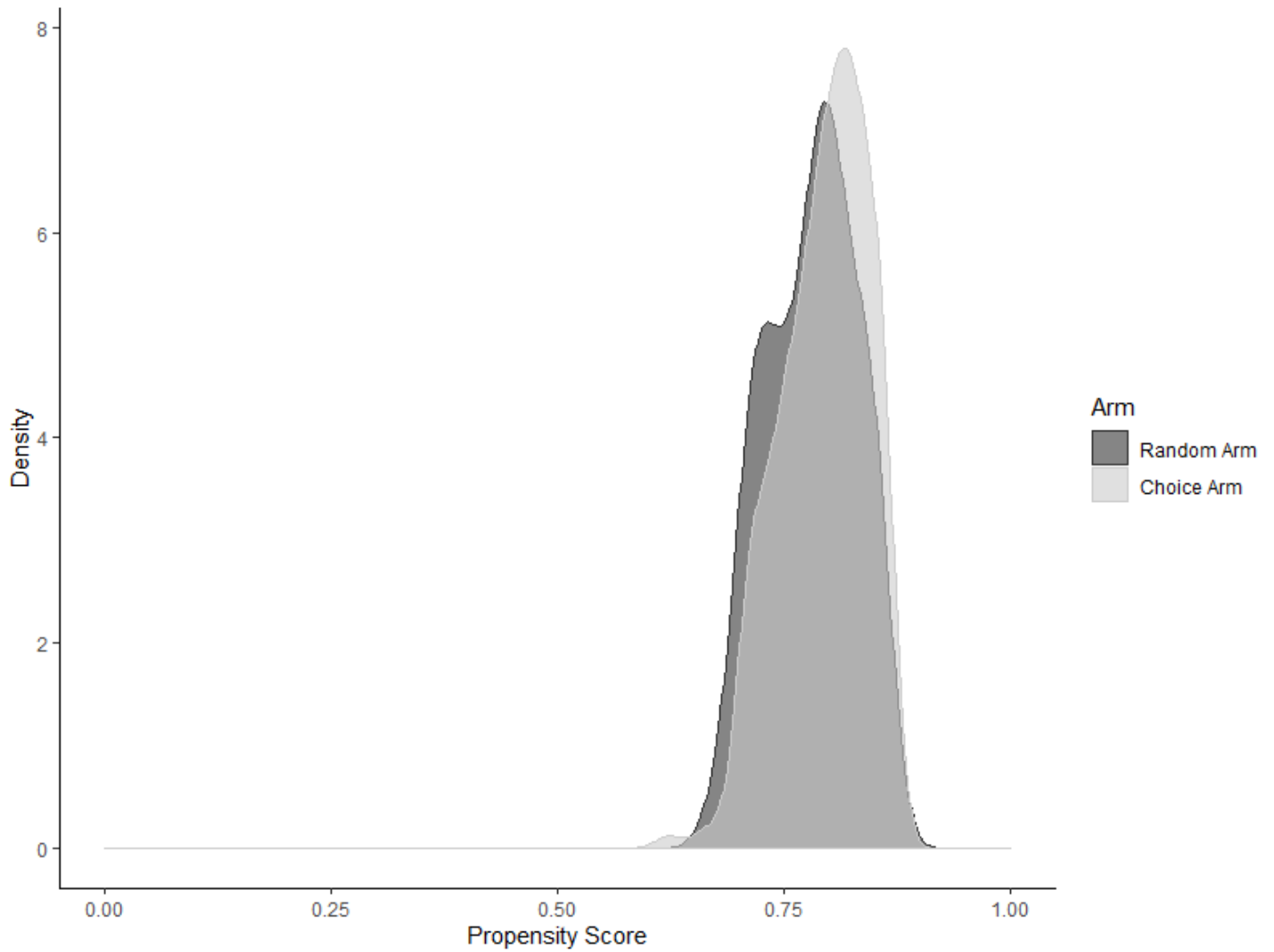


Figure S.3: Distributions of propensity score by choice arm and random arm for Harapan study. The black bar indicate the random arm and the grey bar indicate the choice arm. Patients' choice or random arm allocation indicator as response variable and continuous age as well as four dummy versions of age as predictors were included in the logistic model to generate propensity score.