SUPPLEMENTARY TEXT: ADAPTIVE CLINICAL TRIALS AND SAMPLE SIZE DETERMINATION IN THE PRESENCE OF MEASUREMENT ERROR AND HETEROGENEITY

Hassan Farooq[1], Sajid Ali[2,*], Ismail Shah[3,*], Ibrahim A. Nafisah[4], and Mohammed M. A. Aalmazah[5]

[1,2]*Department of Statistics, Quaid-i-Azam University, Islamabad 45320, Pakistan*

[3]*Department of Statistical Sciences, University of Padua, 35121, Padova, Italy*

[4]*Department of Statistics and Operations Research, College of Sciences, King Saud University, Riyadh 11451, Saudi Arabia*

[5] *Department of Mathematics, College of Sciences and Arts (Muhyil), King Khalid University, Muhyil 61421, Saudi Arabia*

[1]*farooqhassan949@gmail.com,* [2]*sajidali@qau.edu.pk,* [3]*ismail.shah@unipd.it,* [4]*inafisah@ksu.edu.sa,*

[5]*mmalmazah@kku.edu.sa*

Corresponding Author

Abstract. Adaptive clinical trials offer a flexible approach for refining sample sizes during ongoing research to enhance their efficiency. This study delves into improving sample size recalculation through resampling techniques, employing measurement error and mixed distribution models. The research employs diverse sample size recalculation strategies standard simulation, R1, and R2 approaches where R1 considers the mean and R2 employs both mean and standard deviation as summary locations. These strategies are tested against observed conditional power (OCP), restricted observed conditional power (ROCP), promising zone (PZ), and group sequential design (GSD). The key findings indicate that the R1 approach, capitalizing on mean as a summary location, outperforms standard recalculations without resampling as it mitigates variability in recalculated sample sizes across effect sizes. The OCP exhibits superior performance within the R1 approach compared to ROCP, PZ, and GSD due to enhanced conditional power. However, a tendency to inflate the initial stage's sample size is observed in the R1 approach, prompting the development of the R2 approach that considers mean and standard deviation. The ROCP in the R2 approach demonstrates robust performance across most effect sizes, although GSD retains superiority within the R2 approach due to its sample size boundary. Notably, sample size recalculation designs perform worse than R1 for specific effect sizes, attributed to inefficiencies in approaching target sample sizes.

Keywords: Adaptive clinical trials;Group sequential design; Measurement error in trials; Promising zone; Observed conditional power; Sample size determination; Restricted observed conditional power; Mixture distribution

S.1. Additional Results

From Table S1 for the standard sample size recalculation scenario without resampling, the GSD consistently demonstrates superior performance across all Δ. This is particularly evident for $\Delta = 0$, where the GSD achieves the highest conditional performance score of 0.7788, outperforming the other approaches. The OCP approach, however, shows a lower performance score of 0.4915, indicating that it is less effective in this context. The R1 approach significantly enhances the performance of the OCP, ROCP, and PZ methods across all the considered Δ. For instance, the OCP score improves from 0.4915 in the standard simulation to 0.6533 with the R1 approach, reflecting a notable reduction in variability of the recalculated sample sizes. This trend is consistent across all designs, with the R1 approach yielding higher scores than the original versions, particularly for the ROCP and PZ designs. In contrast, the R2 approach, while showing improvements over the original versions, does not consistently outperform the R1 approach. For

example, the OCP score with the R2 approach is 0.5081, which is lower than the R1 score of 0.6533. This suggests that while the R2 approach provides some benefits, it does not achieve the same level of performance enhancement as the R1 approach for smaller effect sizes.

Notably, the sample sizes recalculated using the R1 approach are not excessively high or low, maintaining a balance that contributes to its superior performance. In contrast, the R2 approach tends to push sample sizes closer to the maximum possible threshold, which may limit its effectiveness. This tendency to approach the maximum sample size can lead to less flexibility in adapting to the observed data, potentially resulting in suboptimal performance scores. As the true underlying effect size increases, the performance scores for the OCP and PZ approaches with the R1 and R2 methods converge, particularly for $\Delta \geq 0.3$. For instance, at $\Delta = 0.5$, the OCP score with the R1 approach is 0.6615, while the R2 approach yields a score of 0.5850. This indicates that the R1 approach remains superior for smaller effect sizes, while the R2 approach begins to show competitive performance for larger effect sizes.

TABLE S1. CPS when both treatment and control generated from MDM with $\pi = 0.2$

Δ	Design	Standard Simulation	R1 Approach	R2 Approach
0	OCP	0.4915	0.6533	0.5081
0	ROCP	0.7317	0.8232	0.6604
0	PZ	0.7209	0.7618	0.6685
0	GSD	0.7788	0.7764	0.7196
0.1	OCP	0.4233	0.6208	0.4719
0.1	ROCP	0.6424	0.7952	0.6243
0.1	PZ	0.6477	0.7319	0.6344
0.1	GSD	0.7283	0.7463	0.7463
0.2	OCP	0.3867	0.5850	0.4332
0.2	ROCP	0.5560	0.7644	0.5846
0.2	PZ	0.5831	0.6992	0.5966
0.2	GSD	0.6868	0.7126	0.7126
0.3	OCP	0.4135	0.6334	0.6919
0.3	ROCP	0.3519	0.5572	0.6231
0.3	PZ	0.4236	0.6036	0.6519
0.3	GSD	0.4766	0.6098	0.6098
0.4	OCP	0.4635	0.6831	0.7549
0.4	ROCP	0.4834	0.7031	0.6857
0.4	PZ	0.5298	0.7443	0.6982
0.4	GSD	0.5807	0.6098	0.7549
0.5	OCP	0.4568	0.6615	0.5850
0.5	ROCP	0.5148	0.7330	0.6655
0.5	PZ	0.5309	0.7131	0.6753
0.5	GSD	0.6044	0.7224	0.7224

TABLE S2. Standard Simulation MDM with $\pi = 0.2$

Δ	Design	$\mathbb{E}[CP_\Delta^{RA}]$	$e_{\text{CP}}(\Delta)$	$Var(CP_\Delta^{RA})$	$v_{CP}(\Delta)$	$SCP(\Delta)$	$\mathbb{E}[CN_\Delta^{RA}]$	$e_{CN}(\Delta)$	$Var(CN_\Delta^{RA})$	$v_{CN}(\Delta)$	$SCN(\Delta)$	$CPS(\Delta)$
0	OCP	0.1281	0.8943	0.0601	0.5098	0.7020	119.7084	0.5353	5330.5861	0.0266	0.2809	0.4915
0	ROCP	0.0775	0.9462	0.0534	0.5380	0.7421	61.3434	0.9244	1306.0210	0.5182	0.7213	0.7317
0	PZ	0.0884	0.9349	0.0455	0.5734	0.7542	78.1511	0.8123	1074.0334	0.5631	0.6877	0.7209
0	GSD	0.0733	0.9505	0.0286	0.6616	0.8061	74.5450	0.8364	624.8555	0.6667	0.7515	0.7788
0.1	OCP	0.2104	0.8099	0.0883	0.4058	0.6078	138.4681	0.4102	4893.7000	0.0673	0.2388	0.4233
0.1	ROCP	0.1405	0.8816	0.0888	0.4042	0.6429	69.1081	0.8726	1950.8900	0.4111	0.6419	0.6424
0.1	PZ	0.1555	0.8662	0.0765	0.4468	0.6565	88.5158	0.7432	1218.2630	0.5346	0.6389	0.6477
0.1	GSD	0.1289	0.8934	0.0498	0.5537	0.7236	82.2650	0.7849	572.2770	0.6811	0.7330	0.7283
0.2	OCP	0.3027	0.7152	0.1109	0.3339	0.5246	143.7470	0.3750	4330.0815	0.1227	0.2488	0.3867
0.2	ROCP	0.2244	0.7955	0.1239	0.2960	0.5458	78.9470	0.8070	2560.4807	0.3254	0.5662	0.5560
0.2	PZ	0.2385	0.7810	0.1068	0.3464	0.5637	95.5725	0.6962	1356.7091	0.5089	0.6025	0.5831
0.2	GSD	0.1979	0.8226	0.0719	0.4638	0.6432	86.1150	0.7592	501.5069	0.7014	0.7303	0.6868
0.3	OCP	0.3442	0.5325	0.1238	0.2964	0.4145	131.8613	0.7043	4349.2243	0.1207	0.4125	0.4135
0.3	ROCP	0.2781	0.4647	0.1401	0.2516	0.3581	83.8313	0.3841	2701.2750	0.3071	0.3456	0.3519
0.3	PZ	0.2894	0.4763	0.1256	0.2913	0.3838	95.2890	0.4605	1602.6634	0.4662	0.4634	0.4236
0.3	GSD	0.2403	0.4259	0.0877	0.4079	0.4169	83.8650	0.3844	546.4664	0.6883	0.5363	0.4766
0.4	OCP	0.3236	0.5114	0.1312	0.2755	0.3935	111.4431	0.9241	4133.9923	0.1428	0.5334	0.4635
0.4	ROCP	0.2777	0.4644	0.1409	0.2493	0.3568	81.4731	0.8761	2423.0173	0.3437	0.6099	0.4834
0.4	PZ	0.2850	0.4718	0.1320	0.2733	0.3726	87.9450	0.9193	1671.4819	0.4549	0.6871	0.5298
0.4	GSD	0.2407	0.4263	0.0971	0.3767	0.4015	77.7350	0.8512	617.5815	0.6687	0.7599	0.5807
0.5	OCP	0.2431	0.4288	0.1228	0.2991	0.3640	87.3695	0.8468	3144.5100	0.2524	0.5496	0.4568
0.5	ROCP	0.2180	0.4031	0.1240	0.2957	0.3494	73.3445	0.9403	1892.1480	0.4200	0.6802	0.5148
0.5	PZ	0.2217	0.4069	0.1208	0.3049	0.3559	75.8211	0.9238	1474.1891	0.4881	0.7060	0.5309
0.5	GSD	0.1900	0.3743	0.0920	0.3933	0.3838	68.5700	0.9722	583.7135	0.6779	0.8250	0.6044

TABLE S3. R1 approach when both treatment and control generated from MDM with $\pi = 0.2$

Δ	Design	$\mathbb{E}[CP_\Delta^{RA}]$	$e_{\text{CP}}(\Delta)$	$Var(CP_\Delta^{RA})$	$v_{CP}(\Delta)$	$SCP(\Delta)$	$\mathbb{E}[CN_\Delta^{RA}]$	$e_{CN}(\Delta)$	$Var(CN_\Delta^{RA})$	$v_{CN}(\Delta)$	$SCN(\Delta)$	$CPS(\Delta)$
0	OCP	0.2038	0.8166	0.0677	0.4797	0.6481	134.4892	0.4367	80.4917	0.8804	0.6586	0.6533
0	ROCP	0.1146	0.9081	0.0319	0.6429	0.7755	73.3875	0.8441	58.9043	0.8977	0.8709	0.8232
0	PZ	0.1527	0.8690	0.0455	0.5736	0.7213	103.8805	0.6408	7.4070	0.9637	0.8023	0.7618
0	GSD	0.1493	0.8725	0.0470	0.5666	0.7196	100.0000	0.6667	0.0000	1.0000	0.8333	0.7764
0.1	OCP	0.2641	0.7548	0.0842	0.4196	0.5872	134.2718	0.4382	94.3748	0.8705	0.6543	0.6208
0.1	ROCP	0.1558	0.8659	0.0438	0.5816	0.7237	75.0236	0.8332	55.9920	0.9002	0.8667	0.7952
0.1	PZ	0.2025	0.8180	0.0599	0.5107	0.6643	103.4544	0.6436	11.3032	0.9552	0.7994	0.7319
0.1	GSD	0.1997	0.8208	0.0630	0.4979	0.6593	100.0000	0.6667	0.0000	1.0000	0.8333	0.7463
0.2	OCP	0.3518	0.6648	0.0979	0.3743	0.5196	133.2481	0.4450	116.7995	0.8559	0.6505	0.5850
0.2	ROCP	0.2167	0.8034	0.0558	0.5278	0.6656	77.0112	0.8199	49.2223	0.9065	0.8632	0.76438
0.2	PZ	0.2754	0.7431	0.0733	0.4586	0.6009	102.7049	0.6486	16.1384	0.9464	0.7975	0.6992
0.2	GSD	0.2740	0.7446	0.0787	0.4391	0.5918	100.0000	0.6667	0.0000	1.0000	0.8333	0.7126
0.3	OCP	0.4420	0.6328	0.1022	0.3605	0.4967	131.2461	0.7002	143.7828	0.8401	0.7702	0.6334
0.3	ROCP	0.2844	0.4712	0.0644	0.4924	0.4818	78.6553	0.3496	40.1107	0.9156	0.6326	0.5572
0.3	PZ	0.3537	0.5422	0.0811	0.4306	0.4864	101.7114	0.5033	21.4904	0.9382	0.7208	0.6036
0.3	GSD	0.3548	0.5434	0.0888	0.4041	0.4737	100.0000	0.4919	0.0000	1.0000	0.7460	0.6098
0.4	OCP	0.5257	0.7186	0.0973	0.3763	0.5475	128.7447	0.8087	165.3979	0.8285	0.8186	0.6831
0.4	ROCP	0.3515	0.5400	0.0675	0.4803	0.5101	79.8428	0.8652	30.0680	0.9269	0.8961	0.7031
0.4	PZ	0.4292	0.6197	0.0817	0.4284	0.5241	100.6080	0.9963	25.4502	0.9327	0.9645	0.7443
0.4	GSD	0.4339	0.6245	0.0913	0.3958	0.5101	100.0000	0.9996	0.0000	1.0000	0.9998	0.7550
0.5	OCP	0.6065	0.8015	0.0805	0.4327	0.6171	126.0441	0.5890	176.9650	0.8226	0.7058	0.6615
0.5	ROCP	0.4179	0.6081	0.0632	0.4972	0.5526	80.8842	0.8901	19.8175	0.9406	0.9154	0.7340
0.5	PZ	0.5031	0.6954	0.0727	0.4607	0.5782	99.4476	0.7663	27.5174	0.9301	0.8482	0.7131
0.5	GSD	0.5115	0.7041	0.0833	0.4228	0.5635	100.0000	0.7626	0.0000	1.0000	0.8813	0.7224

TABLE S4. R2 Approach when both treatment and control generated from MDM with $\pi = 0.2$

Δ	Design	$\mathbb{E}[CP_\Delta^{RA}]$	$e_{\text{CP}}(\Delta)$	$Var(CP_\Delta^{RA})$	$v_{CP}(\Delta)$	$SCP(\Delta)$	$\mathbb{E}[CN_\Delta^{RA}]$	$e_{CN}(\Delta)$	$Var(CN_\Delta^{RA})$	$v_{CN}(\Delta)$	$SCN(\Delta)$	$CPS(\Delta)$
0	OCP	0.2759	0.7427	0.1059	0.3493	0.5460	197.8366	0.0144	30.7784	0.9260	0.4702	0.5081
0	ROCP	0.1992	0.8213	0.0739	0.4565	0.6389	119.5254	0.5365	167.9548	0.8272	0.6819	0.6604
0	PZ	0.2027	0.8178	0.0735	0.4578	0.6378	129.5864	0.4694	28.5264	0.9288	0.6991	0.6685
0	GSD	0.1493	0.8725	0.0470	0.5666	0.7196	100.0000	0.6667	0.0000	1.0000	0.8333	0.7764
0.1	OCP	0.3501	0.6666	0.1249	0.2933	0.4799	197.0347	0.0198	47.8374	0.9078	0.4638	0.4719
0.1	ROCP	0.2619	0.7570	0.0943	0.3859	0.5715	122.2188	0.5185	151.6825	0.8358	0.6772	0.6243
0.1	PZ	0.2655	0.7533	0.0935	0.3886	0.5709	130.5314	0.4631	25.6757	0.9324	0.6978	0.6344
0.1	GSD	0.1997	0.8208	0.0630	0.4979	0.6593	100.0000	0.6667	0.0000	1.0000	0.8333	0.7463
0.2	OCP	0.4565	0.5574	0.1376	0.2580	0.4077	195.6488	0.0290	70.2135	0.8883	0.4586	0.4332
0.2	ROCP	0.3542	0.6624	0.1123	0.3298	0.4961	125.3646	0.4976	128.6256	0.8488	0.6732	0.5846
0.2	PZ	0.3575	0.6590	0.1109	0.3330	0.4965	131.6081	0.4559	21.8617	0.9377	0.6968	0.5966
0.2	GSD	0.2740	0.7446	0.0787	0.4391	0.5918	100.0000	0.6667	0.0000	1.0000	0.8333	0.7126
0.3	OCP	0.5617	0.7556	0.1361	0.2622	0.5089	193.7528	0.8830	99.9559	0.8667	0.8749	0.6919
0.3	ROCP	0.4507	0.6417	0.1198	0.3079	0.4748	127.7526	0.6770	101.0471	0.8660	0.7715	0.6231
0.3	PZ	0.4535	0.6446	0.1179	0.3131	0.4789	132.2468	0.7069	18.2562	0.9430	0.8250	0.6519
0.3	GSD	0.3548	0.5434	0.0888	0.4041	0.4737	100.0000	0.4919	0.0000	1.0000	0.7450	0.6098
0.4	OCP	0.6551	0.8513	0.1222	0.3009	0.5761	191.5124	0.3903	123.9879	0.8515	0.6209	0.5985
0.4	ROCP	0.5416	0.7349	0.1161	0.3185	0.5267	129.4534	0.8040	73.7423	0.8855	0.8448	0.6857
0.4	PZ	0.5437	0.7372	0.1142	0.3243	0.5307	132.5736	0.7832	15.1457	0.9481	0.8657	0.6982
0.4	GSD	0.4339	0.6245	0.0913	0.3958	0.5101	100.0000	0.9996	0.0000	1.0000	0.9998	0.7549
0.5	OCP	0.7441	0.9427	0.0938	0.3875	0.6651	189.1384	0.1684	141.3249	0.8415	0.5049	0.5850
0.5	ROCP	0.6295	0.8252	0.0984	0.3727	0.5989	130.8889	0.5567	48.4140	0.9072	0.7320	0.6655
0.5	PZ	0.6311	0.8268	0.0965	0.3787	0.6027	132.7533	0.5443	13.2298	0.9515	0.7479	0.6753
0.5	GSD	0.5115	0.7041	0.0833	0.4228	0.5635	100.0000	0.7626	0.0000	1.0000	0.8813	0.7224

S.2. BOXPLOTS OF CONDITIONAL POWER FOR MIXTURE AND MEM

S.2.1. **Mixture Model Boxplots.** The box plots for second stage conditional power and re-calculated sample size for standard simulation, R1 approach, and R2 approach are illustrated in Figures S1, S2, S3, S4, S5, S6, S7, S8, S9, S10, S11, and S12, respectively.

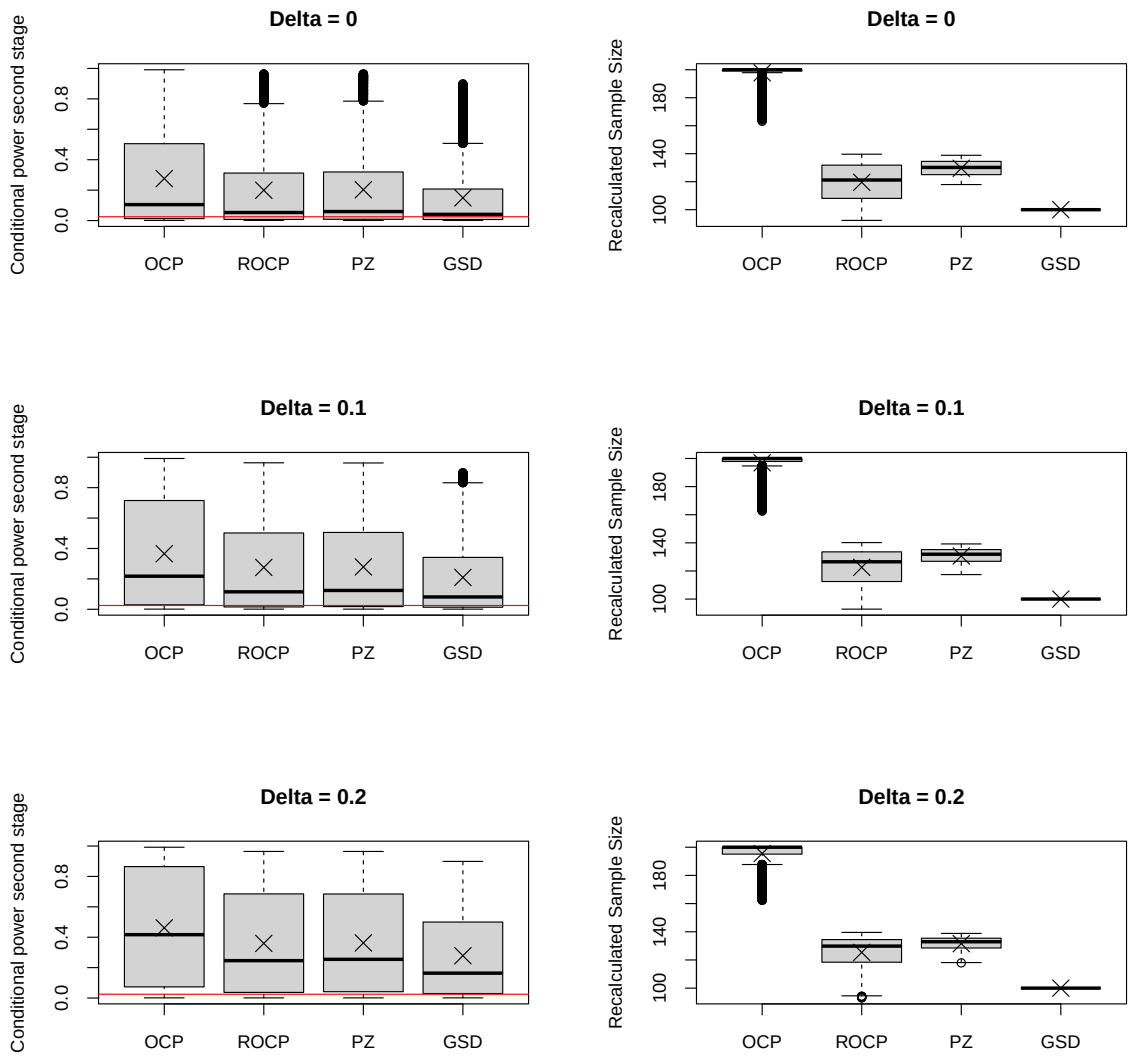


FIGURE S1. Standard simulation when both treatment and control generated from MDM with p.mix=0.5

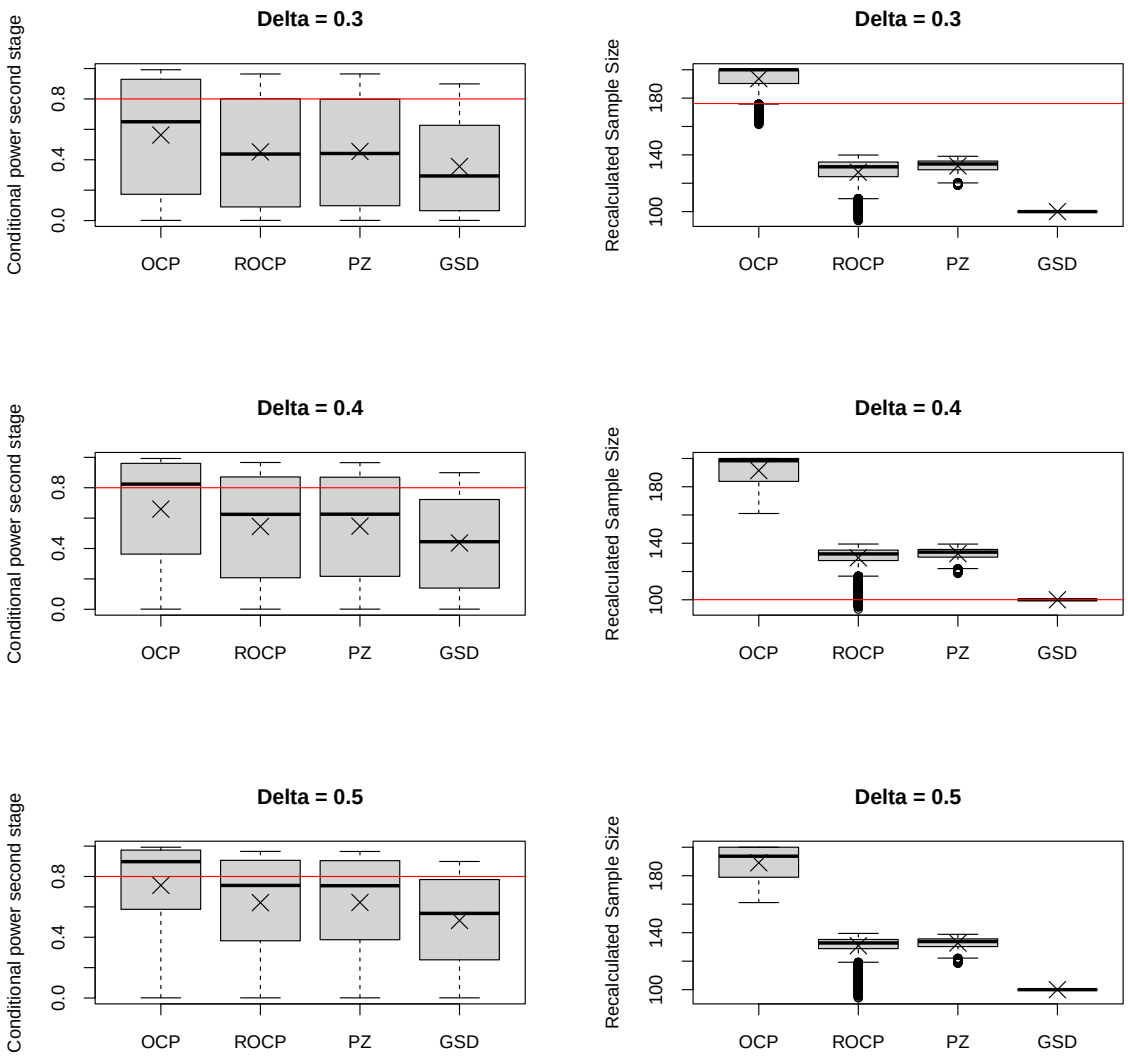


FIGURE S2. Standard simulation when both treatment and control generated from MDM with p.mix=0.5

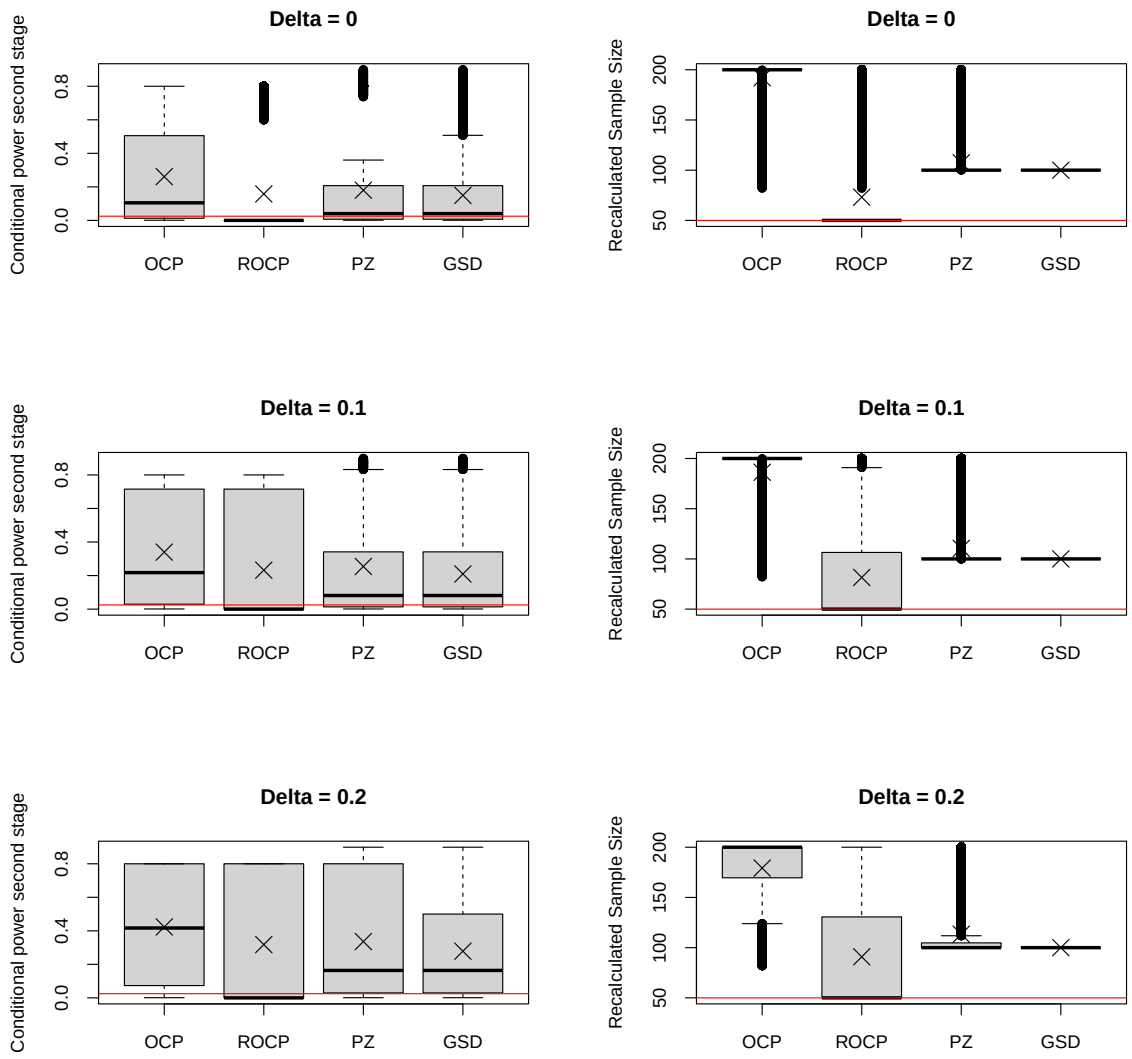


FIGURE S3. R1 Approach when both treatment and control generated from MDM with p.mix=0.5

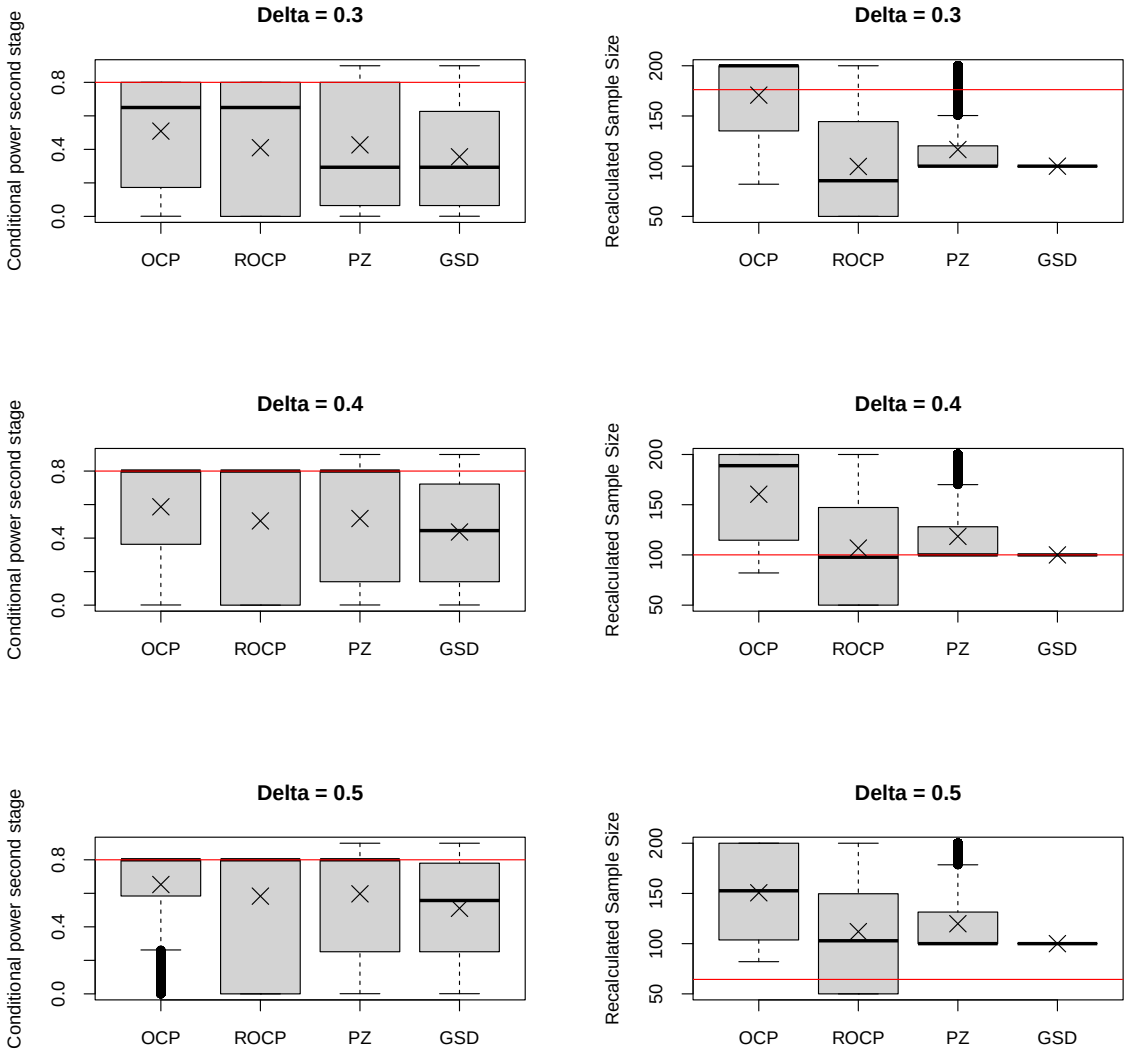


FIGURE S4. R1 Approach when both treatment and control generated from MDM with p.mix=0.5

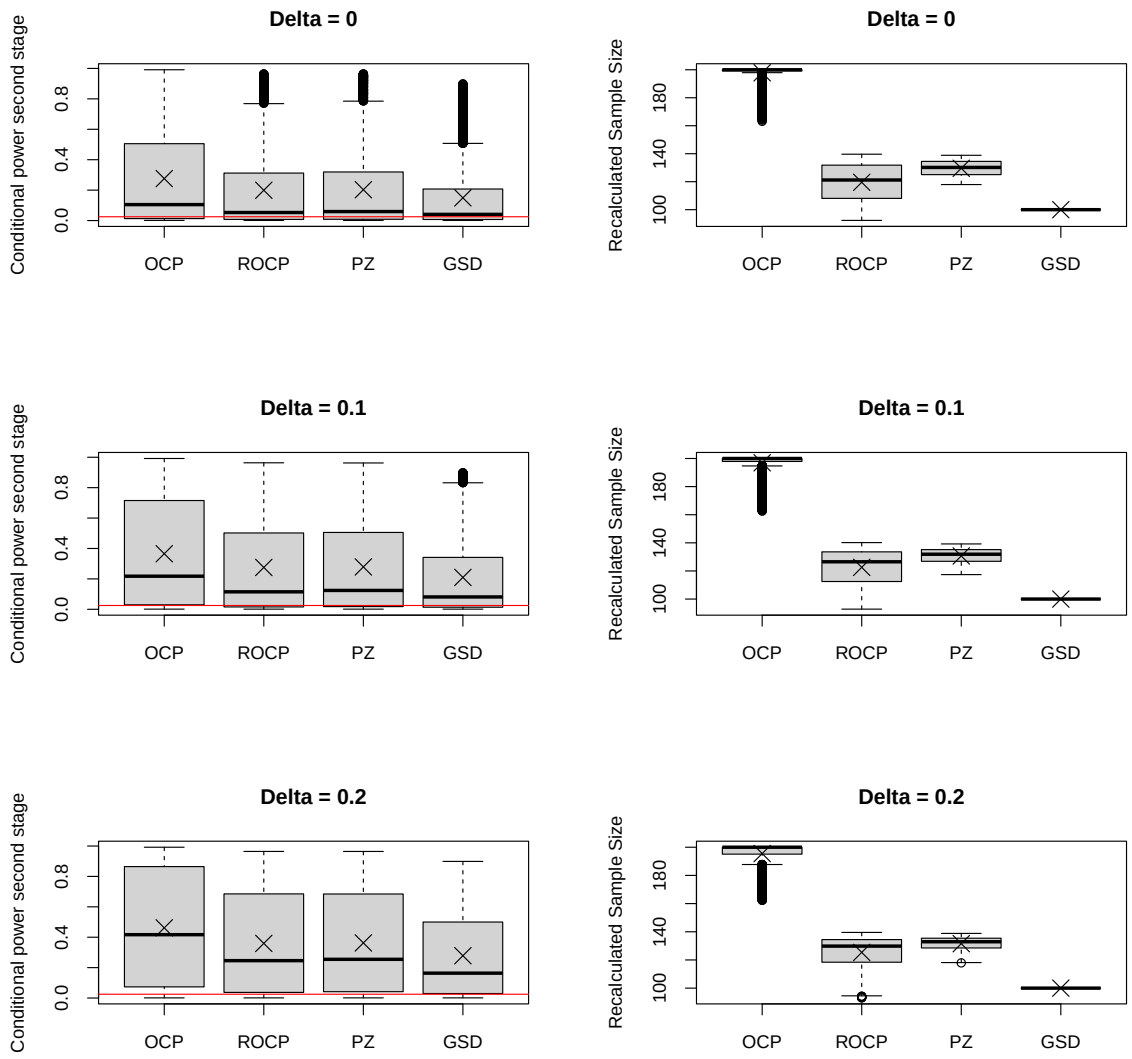


FIGURE S5. R2 Approach when both treatment and control generated from MDM with p.mix=0.5

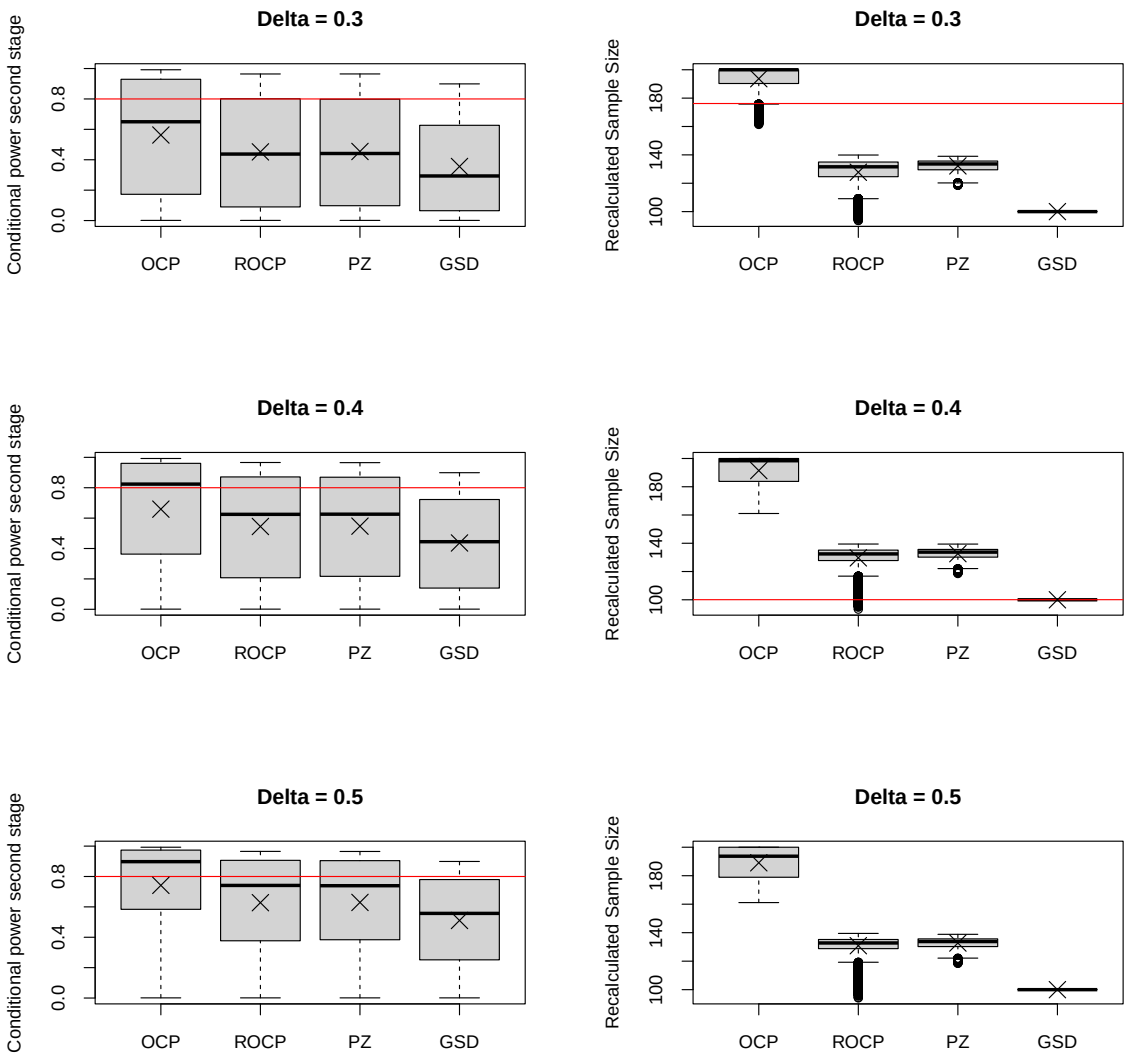


FIGURE S6. R2 Approach when both treatment and control generated from MDM with p.mix=0.5

S.2.2. **MEM Boxplots.** The box plots for second stage conditional power and recalculated sample size for standard simulation, R1 approach, and R2 approach are illustrated in Figures S13, S14, S15, S16, and S17, respectively. The box plots for the second stage conditional power and recalculated sample size for the standard simulation, R1 approach, and R2 approach are illustrated in Figures S18, and S19, respectively.

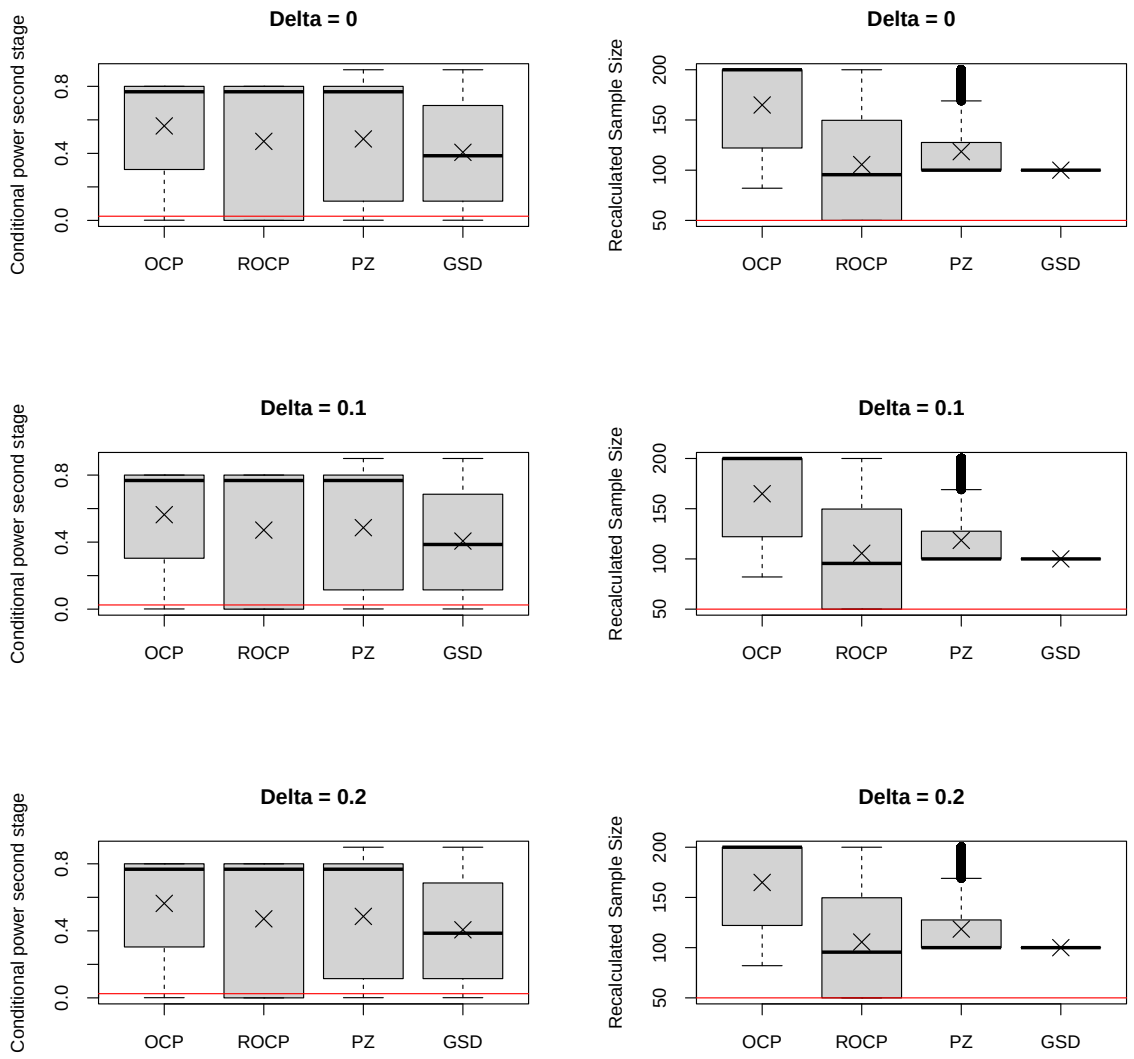


FIGURE S7. Standard simulation when both treatment and control generated from MDM with p.mix=0.1

From Figures S20-S25, the R2 approach consistently outperforms R1 and the standard method across all levels of Δ and design configurations with $\gamma_{loc} = 0.5$. The R1 approach also shows improved performance over the standard simulation, indicating the effectiveness of the refinements made in R1 and especially R2. As Δ increases from 0 to 0.5, the R2 approach maintains superior performance, demonstrating robustness to changes in Δ. The advantage of using R2 becomes more apparent as Δ increases, particularly in the ROCP and PZ designs. The GSD consistently achieves the highest scores across all approaches and Δ levels, with a marginal difference between R1 and R2 at some levels. The PZ and ROCP show substantial improvements under the R1 and R2 approaches, particularly under R2, highlighting their sensitivity to the method used. Thus,

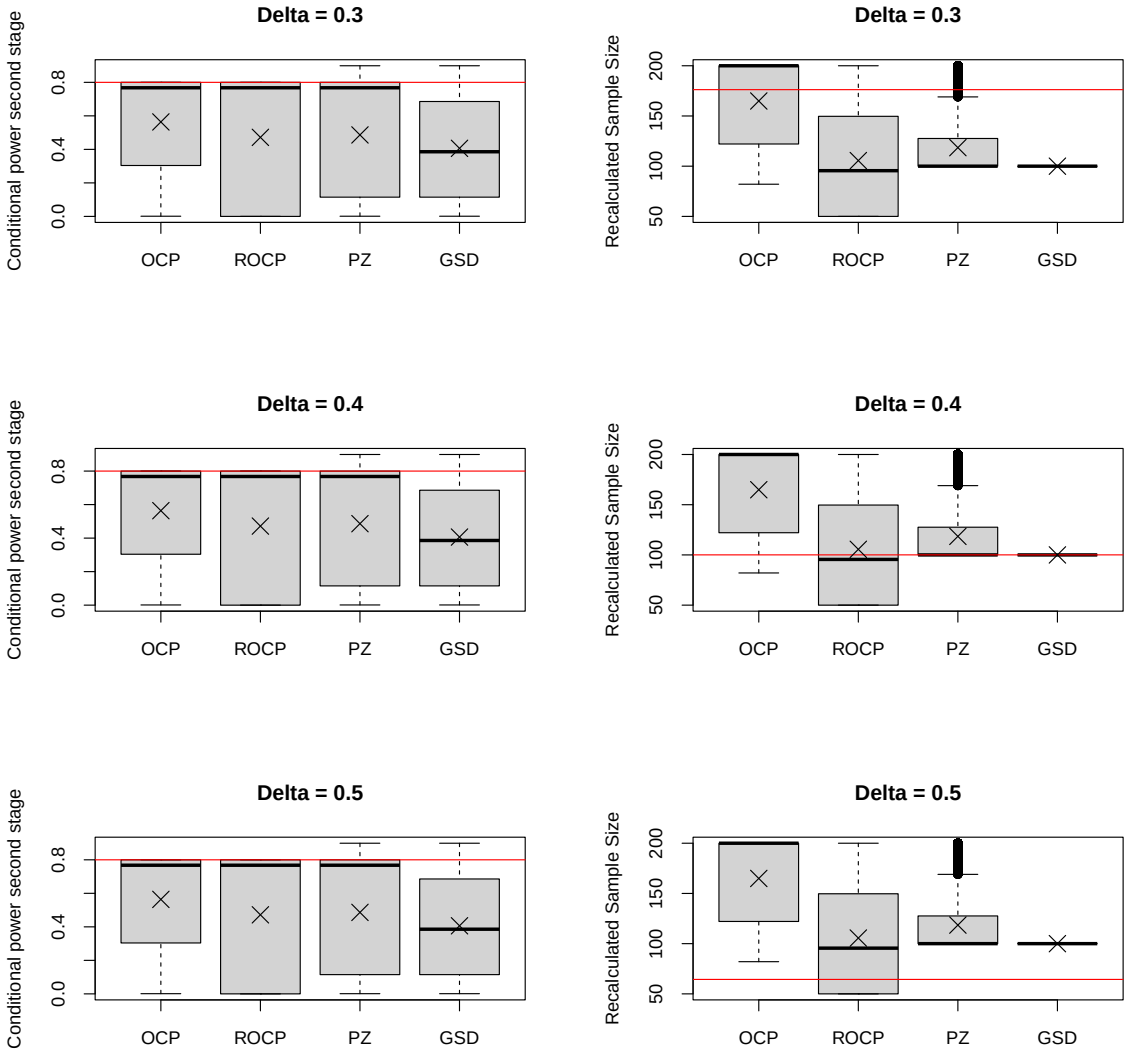


FIGURE S8. Standard simulation when both treatment and control generated from MDM with p.mix=0.1

with equal weight on location and variance in the scoring mechanism, the R2 approach demonstrates the highest effectiveness across design configurations and Δ levels. This suggests that R2 is a robust and superior approach in contexts where location and other factors are equally weighted.

From Figures S26-S31, assuming $\gamma_{\text{loc}} = 0.2$, the designs consistently show a similar trend across all Δ values, and the CPS does not change drastically. The OCP shows the lowest robustness, particularly under the standard approach, where its CPS is the lowest. This indicates that OCP is more sensitive to the unequal weight on the location parameter and performs less consistently across different Δ values. The ROCP and PZ show more stable performance compared to OCP.

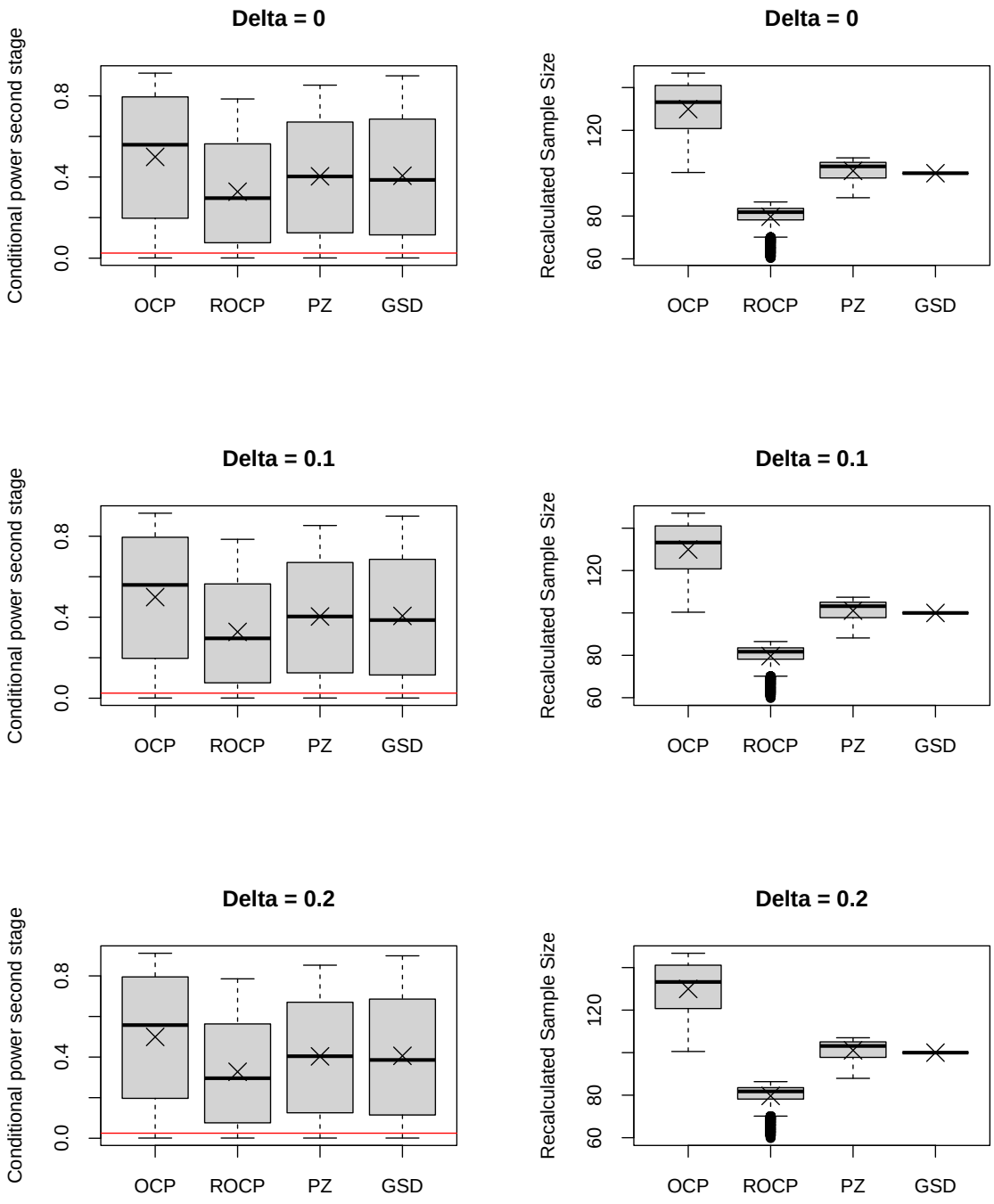


FIGURE S9. R1 Approach when both treatment and control generated from MDM with p.mix=0.1

Their CPS values tend to be closer to the higher performing designs and exhibit relatively stable performance across the different Δ values, particularly under R1 and R2 approaches. The performance of the designs remains largely consistent with increasing Δ, suggesting that the

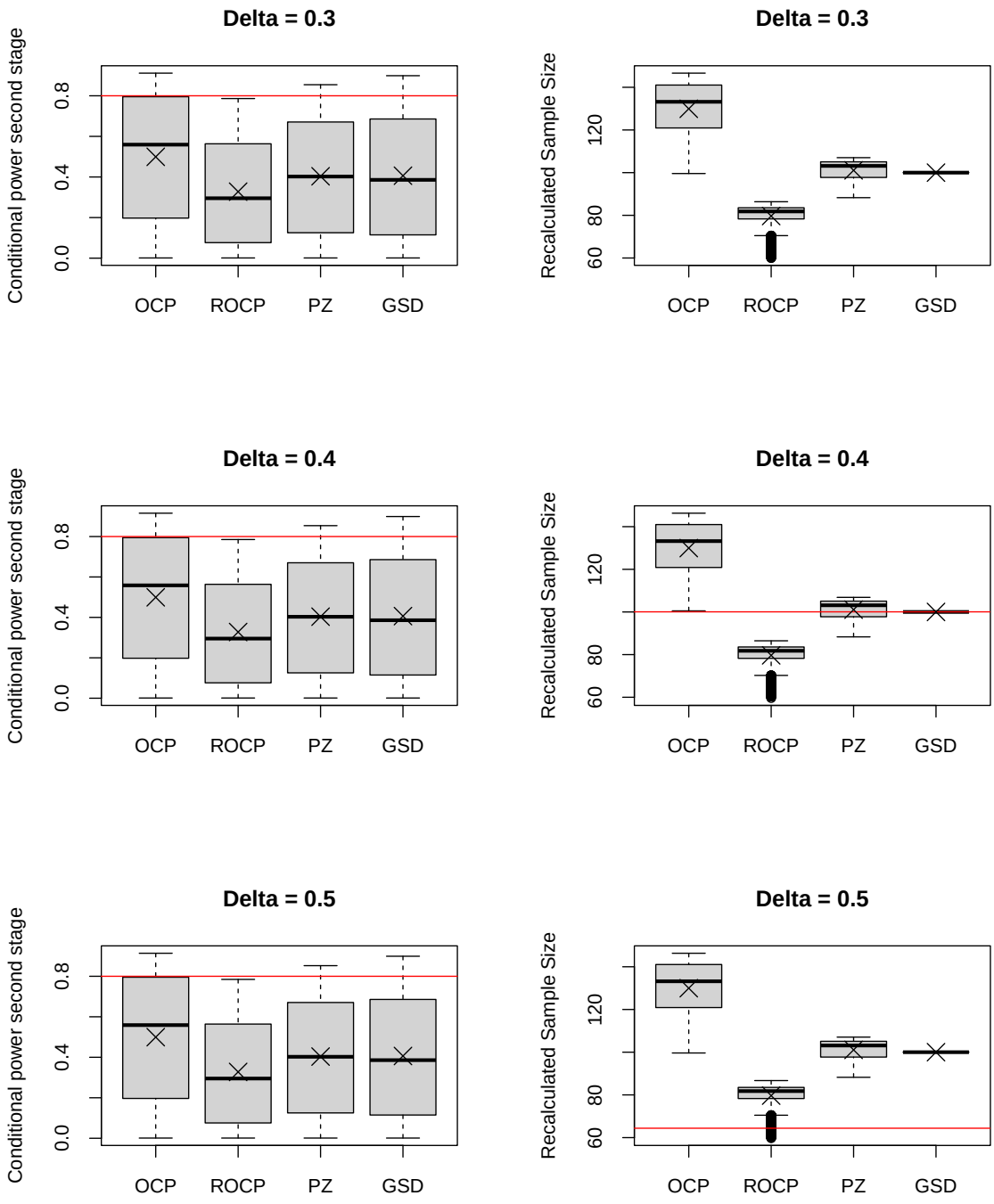


FIGURE S10. R1 Approach when both treatment and control generated from MDM with p=0.1

relative differences between the designs remain intact despite changes in Δ. The designs that perform better in lower Δ values continue to perform better at higher Δ values, and vice versa. The unequal weights for the location parameter ($\gamma_{\text{loc}} = 0.2$) highlight the advantages of the R1

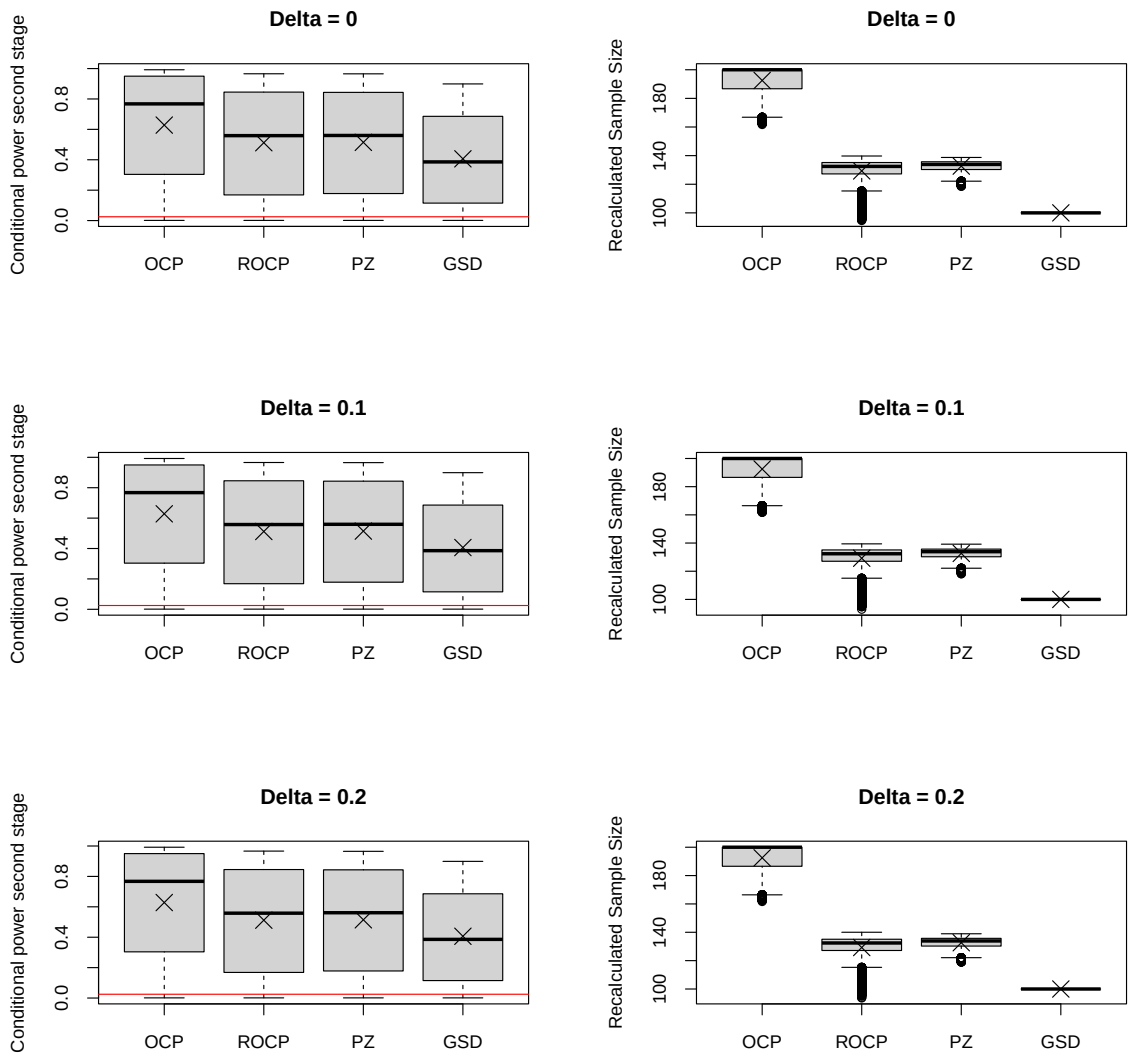


FIGURE S11. R2 Approach when both treatment and control generated from MDM with p=0.1

and R2 approaches, which show consistently a higher CPS as compared to the standard approach. The OCP is the least robust design strategy, with lower CPS values across all approaches and Δ values. In contrast, the ROCP and PZ exhibit relatively stable performance and better robustness. The overall trend suggests that designs like ROCP and PZ are more reliable in maintaining high CPS under the unequal weighting of location, while the OCP is more sensitive to the changes in location weight.

The CPS across all approaches (Standard, R1, and R2) is affected by the weighting of the location parameter (γ_{loc}). In particular, a large γ_{loc}, i.e., $\gamma_{\text{loc}} = 0.5$, has a significant influence on the CPS, which leads to different performance patterns compared to when $\gamma_{\text{loc}} = 0.2$. The R1

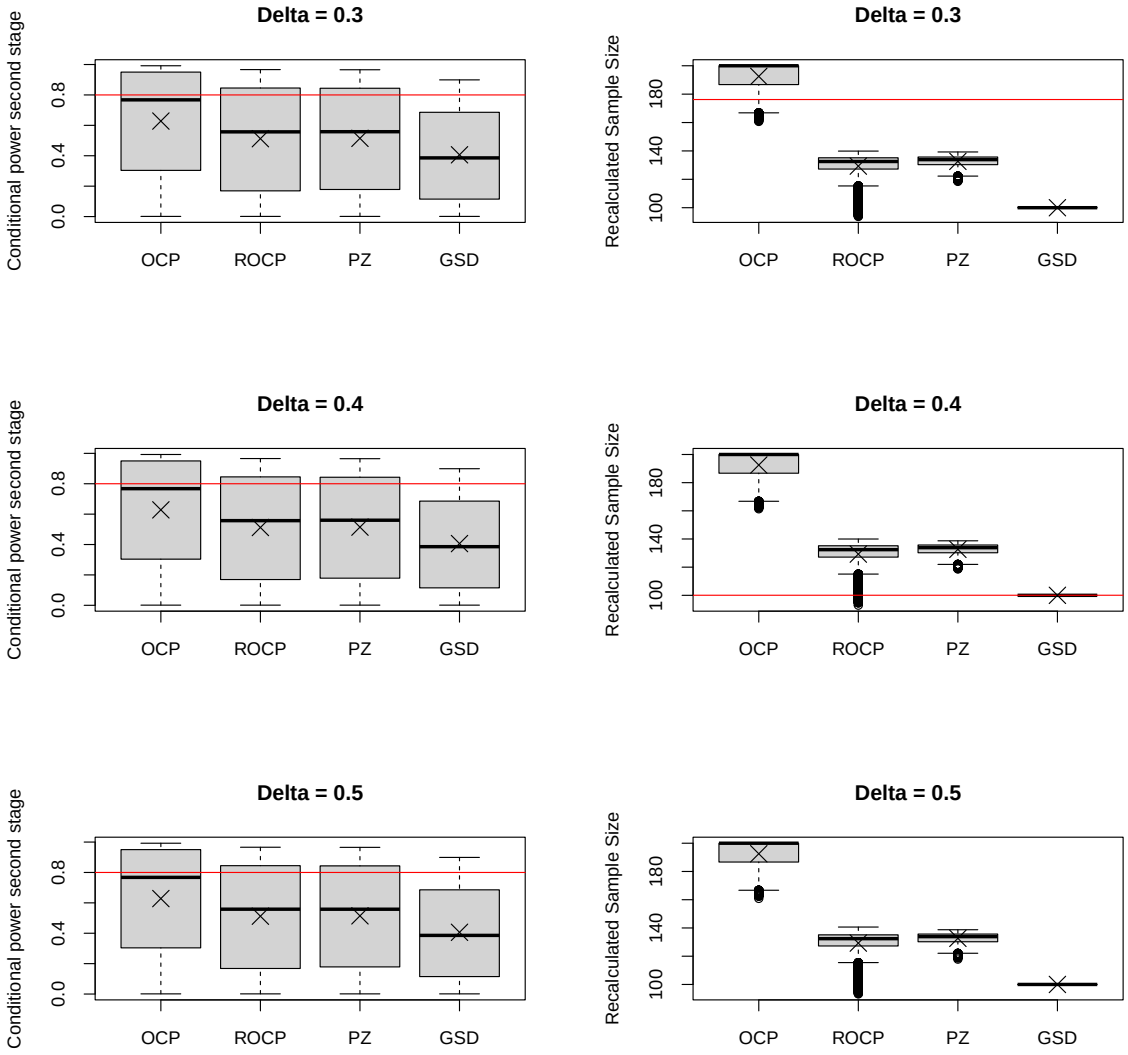


FIGURE S12. R2 Approach when both treatment and control generated from MDM with p=0.1

approach generally performs better than the standard approach, especially for higher values of Δ, i.e., $\Delta = 0.4$ and $\Delta = 0.5$. For example, in the GSD design at $\Delta = 0.5$, the R1 yields a score of 0.7212, compared to 0.685 for the standard approach, while the performance improvement of R1 is less significant. The difference in CPS between R1 and standard is smaller, especially for lower values of Δ, like $\Delta = 0.0$ (0.6679 vs. 0.3575 for the OCP design). The R2 approach outperforms R1 and the standard approach across all values of Δ. For instance, in the GSD design at $\Delta = 0.5$, the R2 achieves 0.7212, which is significantly higher than the standard's score of 0.685. For $\gamma_{\text{loc}} = 0.2$, the R2 still outperforms both standard and R1, but the overall score values are lower. For example, at $\Delta = 0.0$, the GSD design has a score of 0.7789 for both R1 and

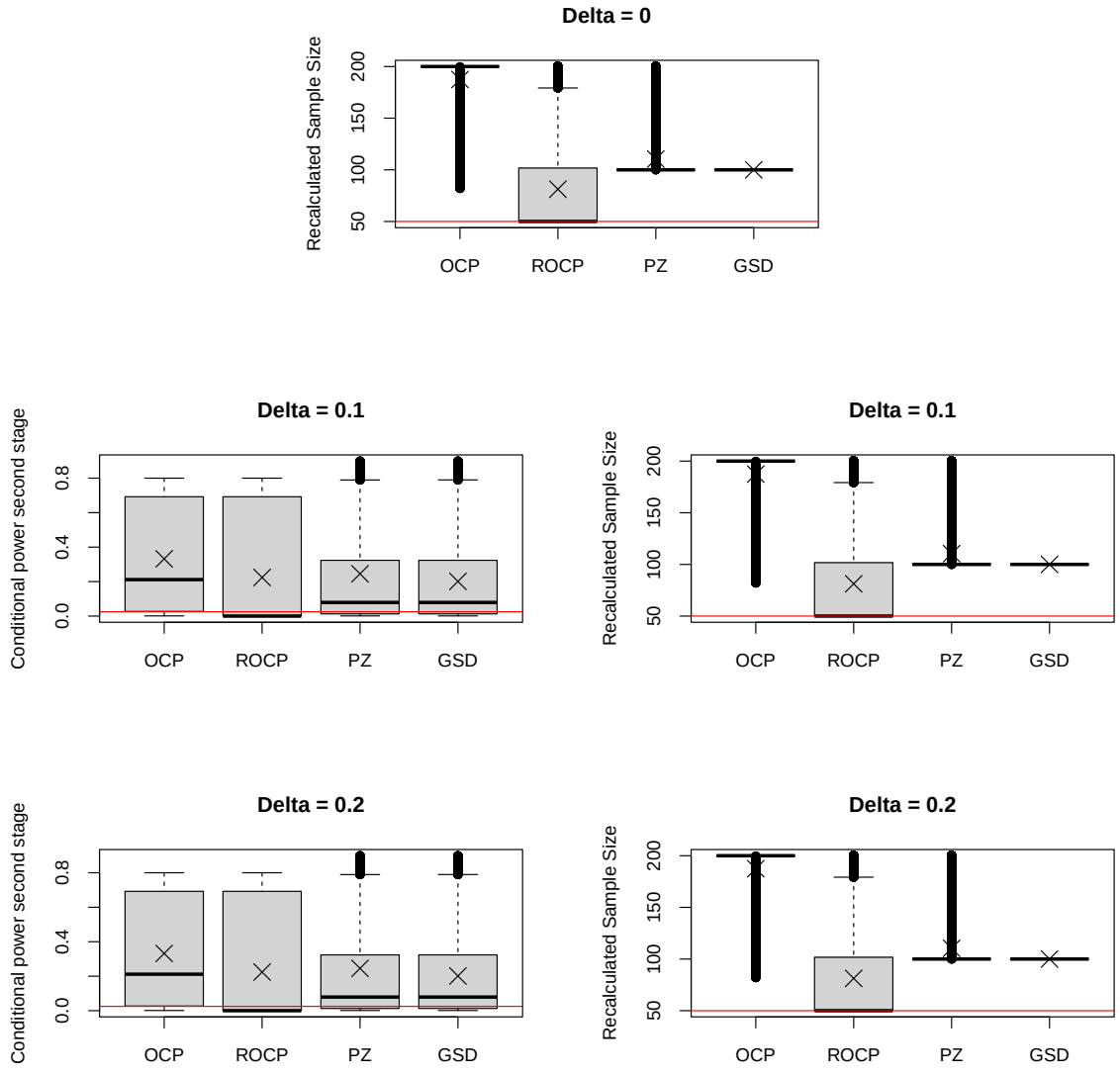


FIGURE S13. Standard simulation when both treatment and control generated from MEM

R2, which is lower than the score achieved under $\gamma_{\text{loc}} = 0.5$ (0.7764 for R1 and 0.7764 for R2 at $\Delta = 0.0$). The Standard approach shows the lowest scores across the board, but the differences between the standard and R1/R2 approaches are more noticeable with $\gamma_{\text{loc}} = 0.5$, particularly at higher Δ values. For example, at $\Delta = 0.5$, the standard approach's performance is significantly lower in the GSD design (0.685) compared to R1 (0.7212) and R2 (0.7212). With $\gamma_{\text{loc}} = 0.2$, the standard approach still shows the lowest CPS, but the difference in performance from R1 and R2 is narrower, especially for designs like the ROCP and PZ at higher Δ values.

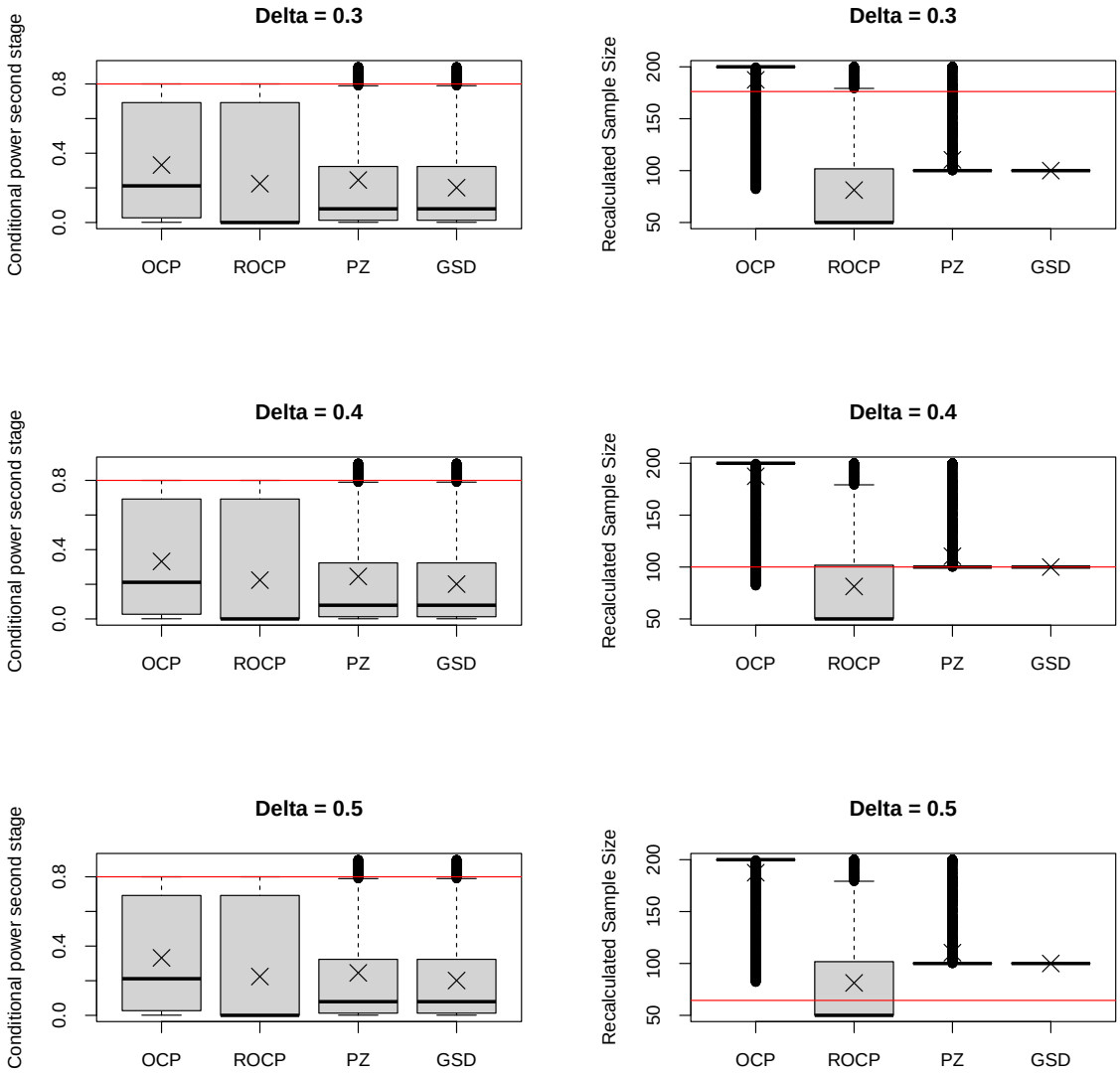


FIGURE S14. Standard simulation when both treatment and control generated from MEM

The choice of the location weight (γ_{loc}) has a significant impact on the performance of sample size recalculation approaches. With equal location weighting (w_{loc} = 0.5), the differences in performance between R1, R2, and standard are more substantial, with R2 generally providing the best results. In contrast, with unequal location weighting (γ_{loc} = 0.2), the approaches' performance becomes more comparable, and the overall conditional scores are lower but higher than the standard method. Therefore, the effectiveness of each approach is more pronounced when the location parameter is given equal weight in the score.

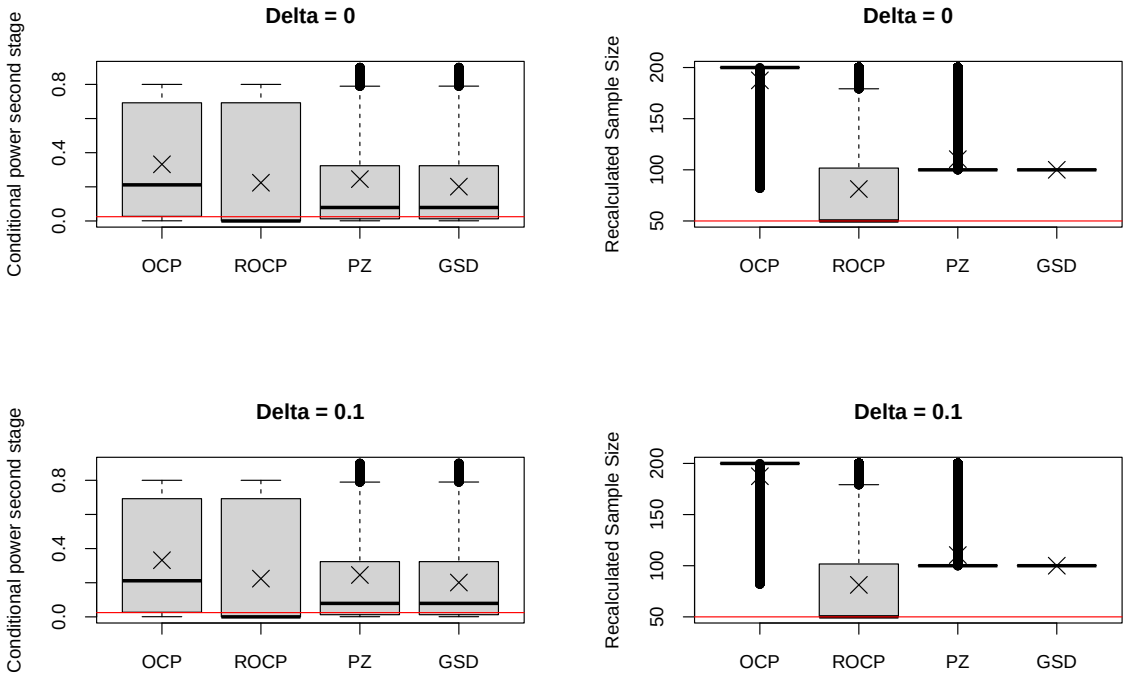


FIGURE S15. R1 Approach when both treatment and control generated from MEM

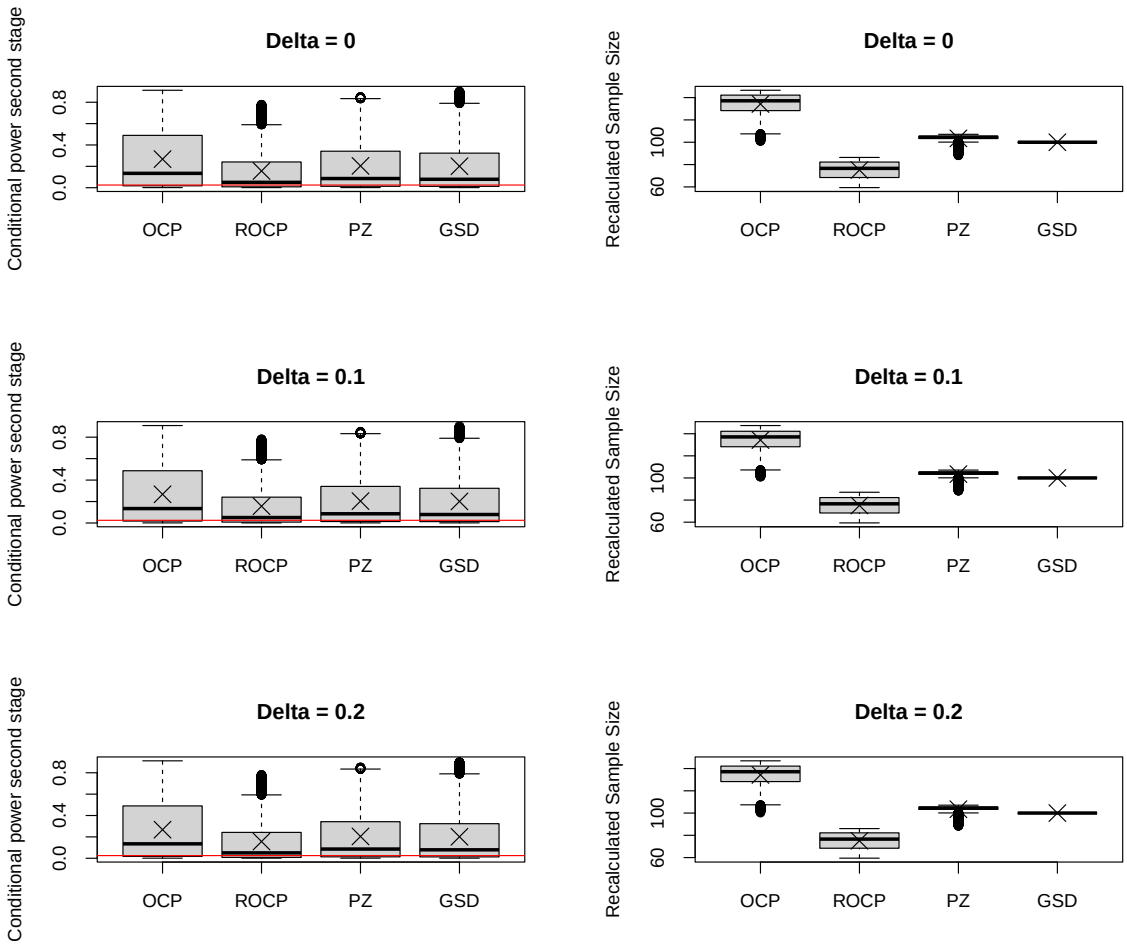


FIGURE S16. R2 Approach when both treatment and control generated from MEM

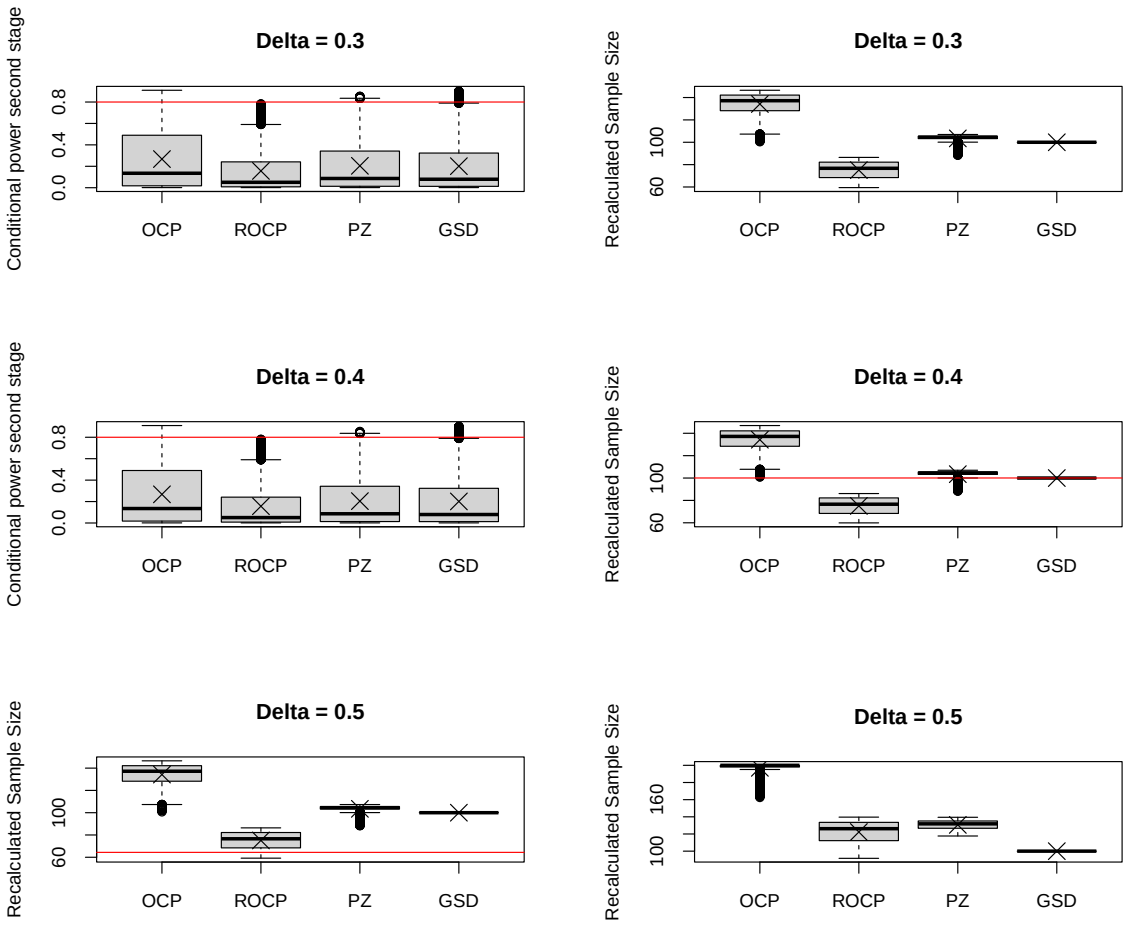


FIGURE S17. R2 Approach when both treatment and control generated from MEM

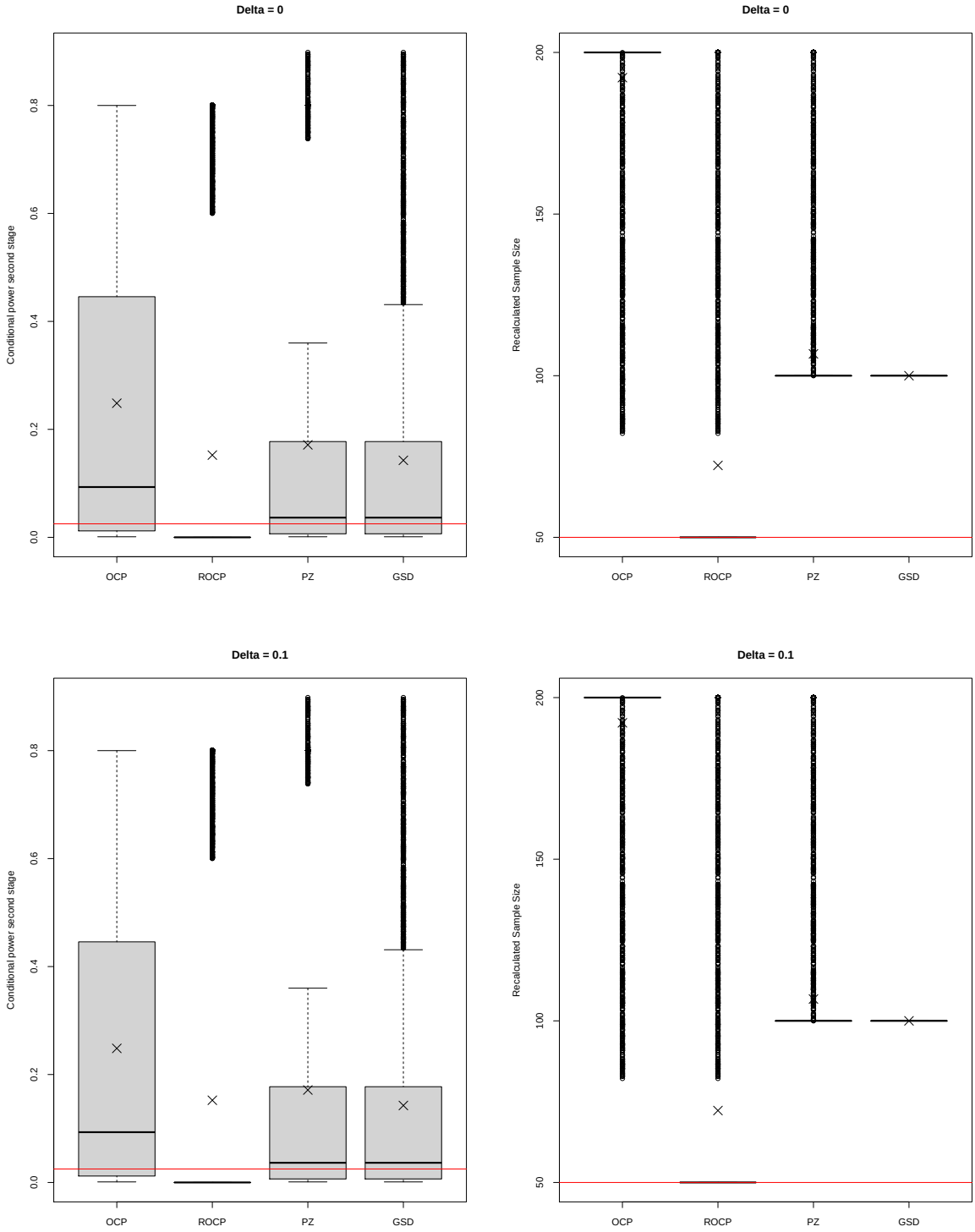


FIGURE S18. Standard simulation where only treatment is generated with MEM and control from a normal distribution

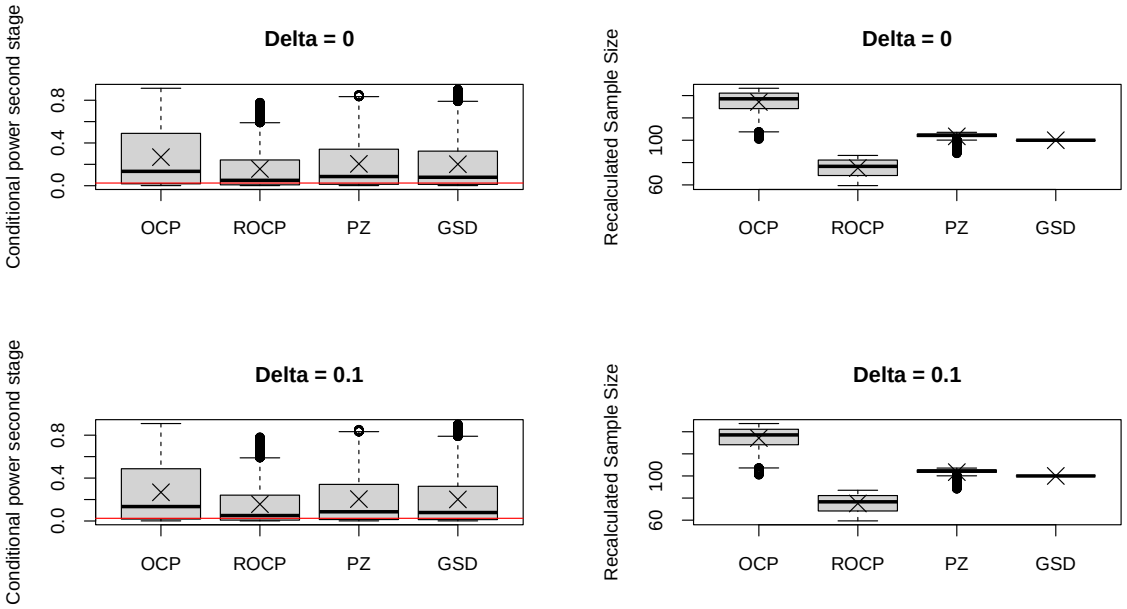


FIGURE S19. R2 approach where only treatment is generated with MEM and control from a normal distribution

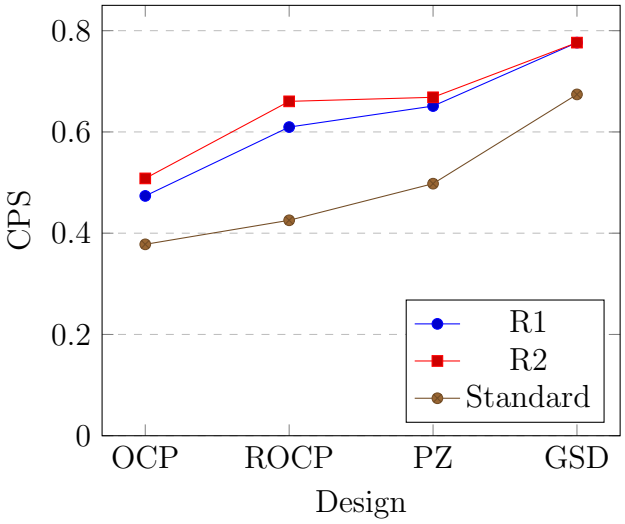


FIGURE S20. CPS comparison at $\Delta = 0.0$ using $\gamma_{\text{loc}} = 0.5$ and $\pi = 0.5$ for the MEM

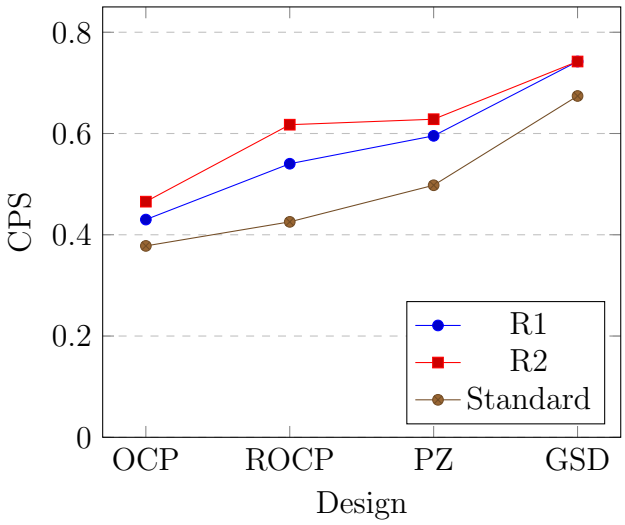


FIGURE S21. CPS comparison at $\Delta = 0.1$ using $\gamma_{\text{loc}} = 0.5$ and $\pi = 0.5$ for the MEM

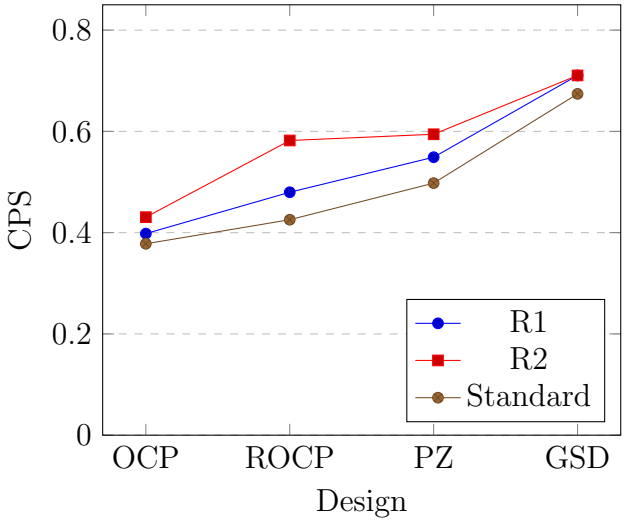


FIGURE S22. CPS comparison at $\Delta = 0.2$ using $\gamma_{\text{loc}} = 0.5$ and $\pi = 0.5$ for the MEM

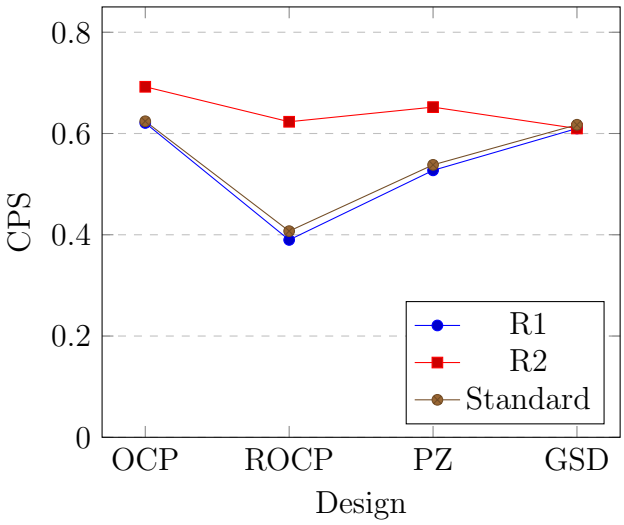


FIGURE S23. CPS comparison at $\Delta = 0.3$ using $\gamma_{\text{loc}} = 0.5$ and $\pi = 0.5$ for the MEM

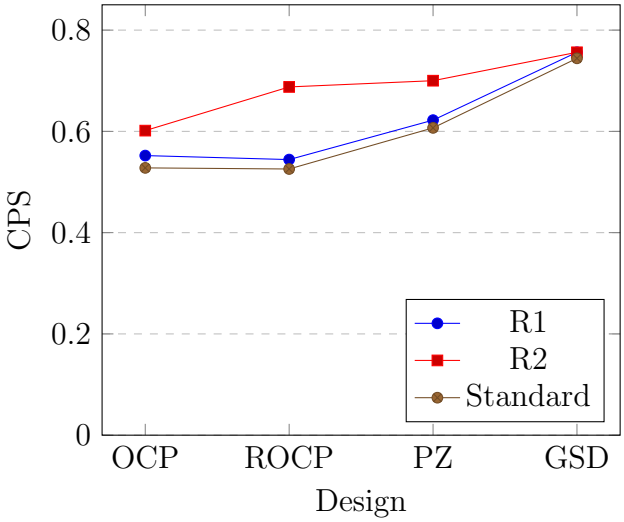


FIGURE S24. CPS comparison at $\Delta = 0.4$ using $\gamma_{\text{loc}} = 0.5$ and $\pi = 0.5$ for the MEM

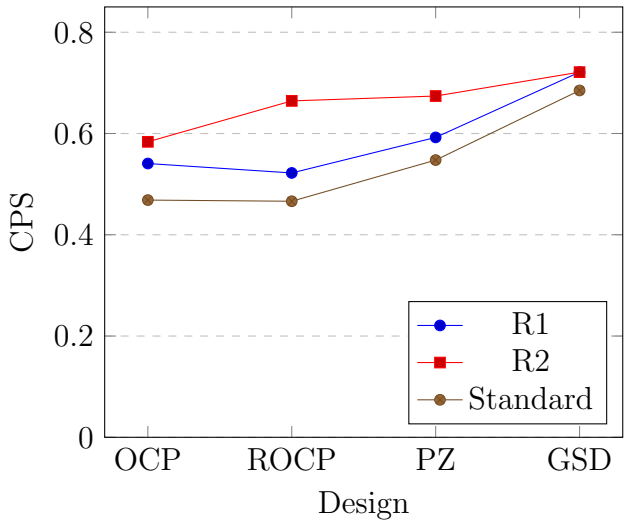


FIGURE S25. CPS comparison at $\Delta = 0.5$ using $\gamma_{\text{loc}} = 0.5$ and $\pi = 0.5$ for the MEM

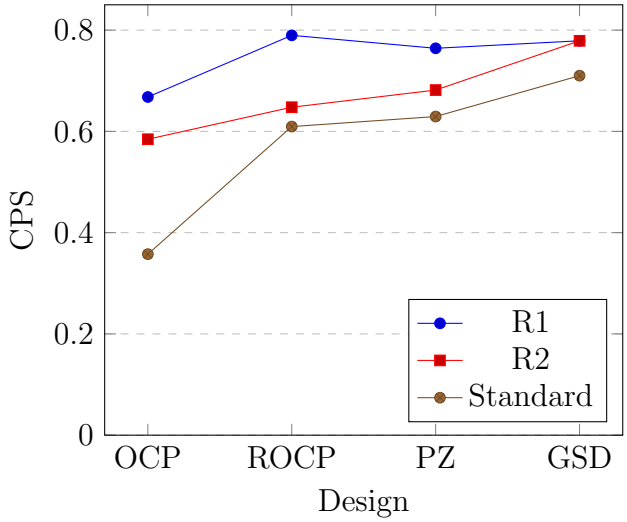


FIGURE S26. CPS comparison at $\Delta = 0.0$ using $\gamma_{\text{loc}} = 0.2$ and $\pi = 0.5$ for the MEM

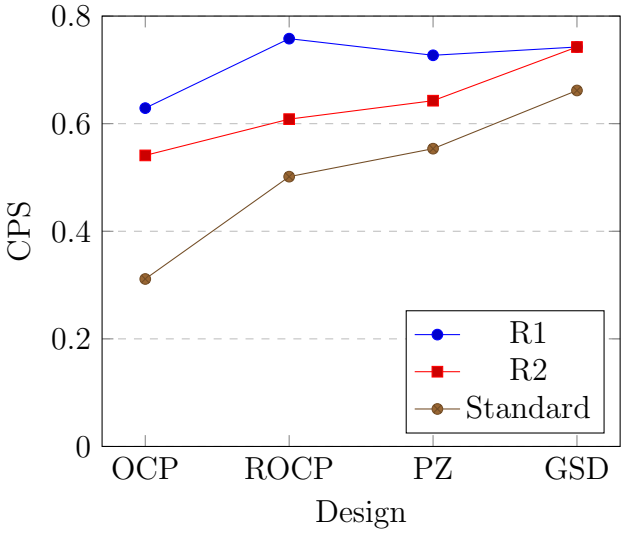


FIGURE S27. CPS comparison at $\Delta = 0.1$ using $\gamma_{\text{loc}} = 0.2$ and $\pi = 0.5$ for the MEM

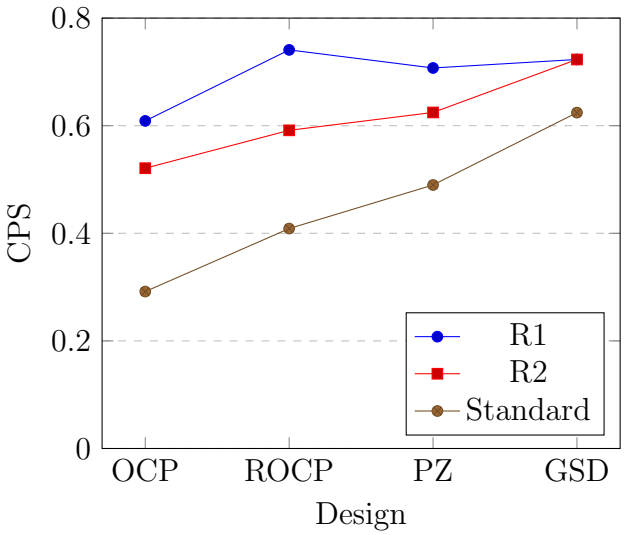


FIGURE S28. CPS comparison at $\Delta = 0.2$ using $\gamma_{\text{loc}} = 0.2$ and $\pi = 0.5$ for the MEM

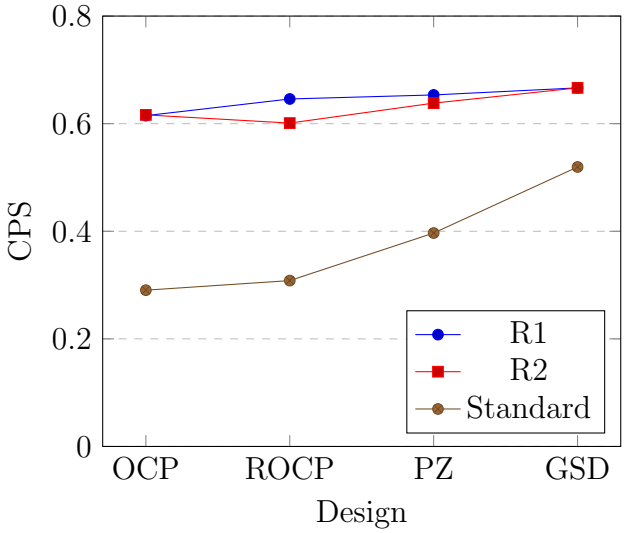


FIGURE S29. CPS comparison at $\Delta = 0.3$ using $\gamma_{\text{loc}} = 0.2$ and $\pi = 0.5$ for the MEM

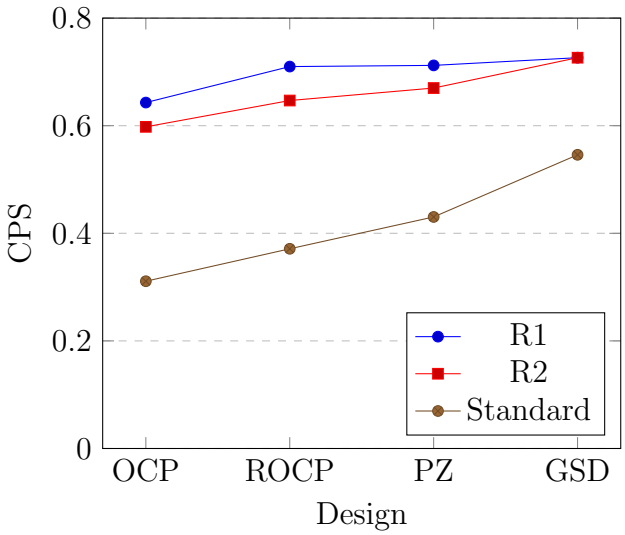


FIGURE S30. CPS comparison at $\Delta = 0.4$ using $\gamma_{\text{loc}} = 0.2$ and $\pi = 0.5$ for the MEM

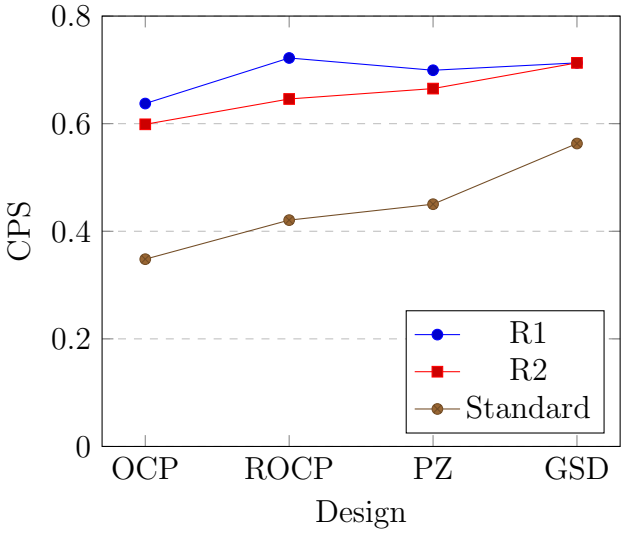


FIGURE S31. CPS comparison at $\Delta = 0.5$ using $\gamma_{\text{loc}} = 0.2$ and $\pi = 0.5$ for the MEM