

Reproducibility Guidelines: Step 2

Noirrit Kiran Chandra

- Please double check whether **Step 1** has been executed.
- Set the location of this readme file as the working directory.
- Make sure to install the **CAPPMx** and **OpenMPController** R packages.

Load the R packages

```
packages=c("Rcpp", "RcppArmadillo", "RcppGSL", "RcppDist", "plyr", "reshape2", "ggplot2",  
           "caret", "parallel", "psrwe", "mvtnorm", "foreach", "doParallel", "MASS",  
           "inline", "devtools", "writexl", "ggpubr", "coda", "ggmcmc", "latex2exp",  
           "ape", "survival", "survminer", "data.table", "dplyr", "optmatch", "MCMCpack",  
           "ROCR", "dbarts", "randomForest", "mltools", "gtools", "ape", "OpenMPController", "CAPPMx")  
lapply(packages, require, character.only = TRUE)
```

- The simulations require a lot of space in the hard drive. Before proceeding please secure **1 terabyte** of disk space.

CAM and MIX Scenarios

Simulate Data

First set the location of this directory and run the following in R to generate the simulated data.

```
dir=paste0(getwd(), "/simData/")  
  
source("gen_data_codes/gen_data_Mixture.R")  
source("gen_data_codes/gen_data_Mixture2.R")
```

The simulated data will be stored in the **simData** folder.

Analyze the Simulated Data

CA-PPMx Method

Run the following in R to analyze the data using our proposed CA-PPMx method. On multicore processors it runs parallelly. However, depending on the configuration of the computer, it might still need days to finish.

```
input_dir=paste0(getwd(), "/simData/")  
output_dir=paste0(getwd(), "/propensity_score_sims/Outputs/")  
source("propensity_score_sims/analysis_CA-PPMx_propscore.R")  
gc()
```

The outputs will be stored in **propensity_score_sims/Outputs** folder. Then we estimate the treatment effects:

```
input_dir=paste0(getwd(), "/propensity_score_sims/Outputs/")  
output_dir=paste0(getwd(), "/propensity_score_sims/")
```

```
source("propensity_score_sims/gen_biases_CA-PPMx_propscore.R")
gc()
```

```
##          used (Mb) gc trigger      (Mb)    max used   (Mb)
## Ncells  3282645 175.4   6310714   337.1     6310714   337.1
## Vcells 30567133 233.3 1480229203 11293.3 2007315308 15314.7
```

The outputs are stored in the folder `propensity_score_sims` with the name `CA-PPMx.txt`.

Propensity-score (PS) Based Method

Then analyze the simulated data using the PS-based methods:

```
input_dir=paste0(getwd(), "/simData/")
output_dir=paste0(getwd(), "/propensity_score_sims/")
source("propensity_score_sims/analysis_propscore.R")
```

The outputs are stored in the folder `propensity_score_sims` with the name `prop_score.txt`.

Matching Based Method

The analyze the data using the Matching-based methods:

```
input_dir=paste0(getwd(), "/simData/")
output_dir=paste0(getwd(), "/propensity_score_sims/")
source("propensity_score_sims/analysis_optmatch.R")
```

The outputs are stored in the folder `propensity_score_sims` with the name `Matching.txt`.

Showing Population Equivalence

We generate the boxplots of the area under the receiver operating characteristic curve (AUC) under repeat simulations that we have used in **Figure 3(a)** of the main article to show empirical evidence towards dealing with the lack of randomization.

```
output_dir=paste0(getwd(), "/propensity_score_sims/")
simdir=paste0(getwd(), "/simData/")
loc_MCMC_outputs=paste0(getwd(), "/propensity_score_sims/Outputs/")
source("propensity_score_sims/get_auc.R")
source("propensity_score_sims/get_auc_random.R")
source("propensity_score_sims/auc_boxplots.R")
p1

ggsave(plot=p1, filename=paste0(output_dir, "AUC_boxplots.pdf"), width=9, height=5.5, units="in")
```

A pdf copy of this figure is saved in the folder `propensity_score_sims` with the name `AUC_boxplots.pdf`.

Compare Powers of Detecting Treatment Effects

Based on the above three analyzes we report the power of detecting treatment effects in the following figure which is also used in **Figure 4(a)** of the main article.

```
input_dir=output_dir=paste0(getwd(), "/propensity_score_sims/")
source("propensity_score_sims/gen_type_I_n_II_errors.R")
p_fig
```

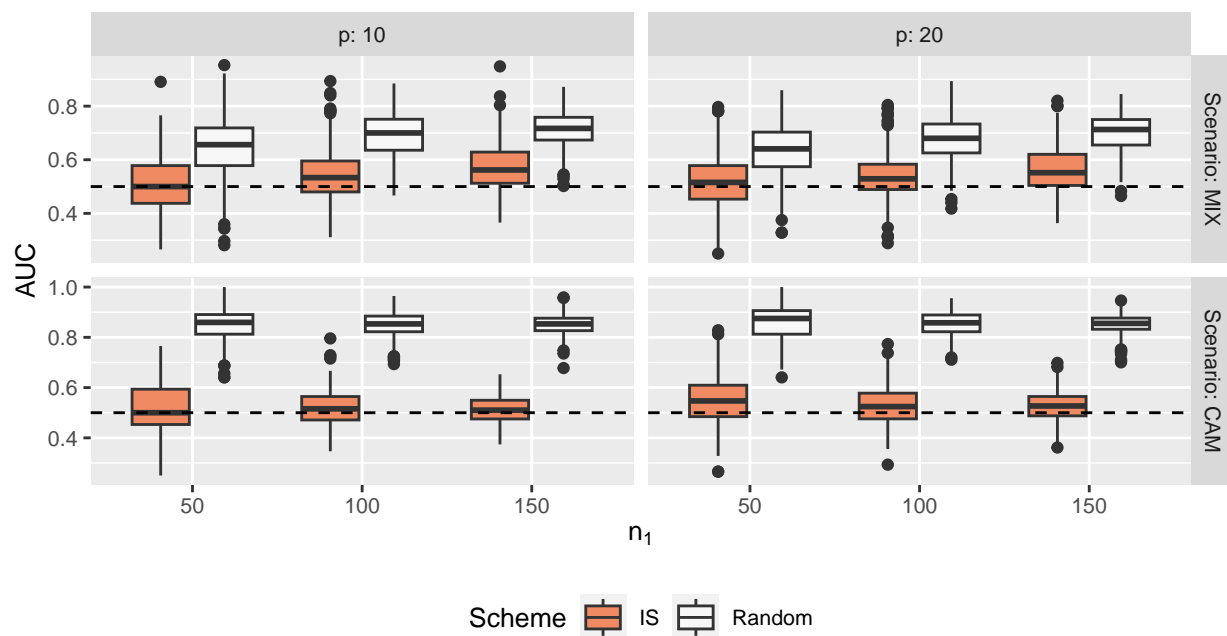
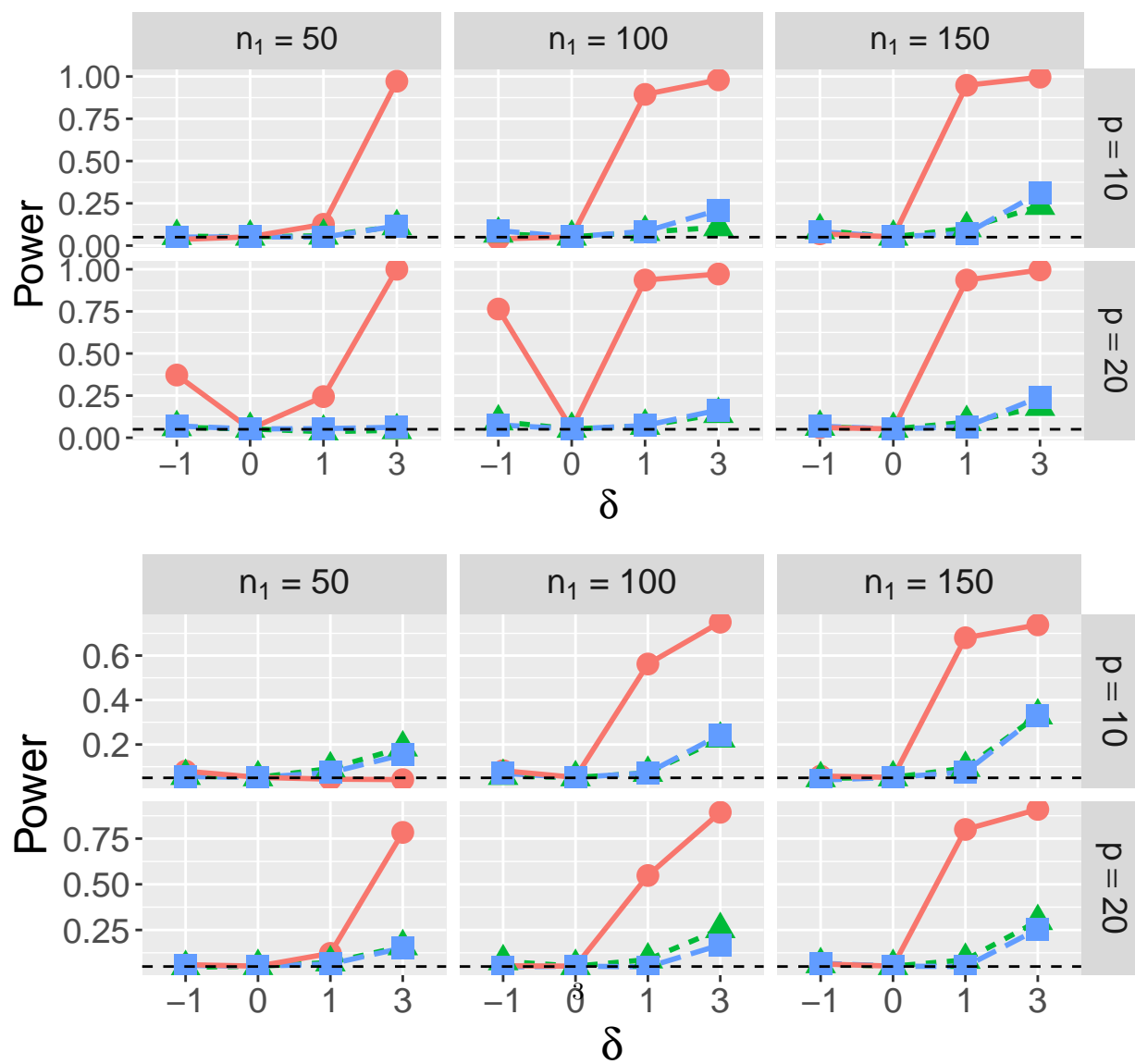


Figure 1: Figure 3(a) of the main article



```
ggsave(plot=p_fig,filename=paste0(output_dir,"Comb_powerplots.pdf"),width=12,height=7.5,units="in")
```

A pdf copy of this figure is saved in the folder `propensity_score_sims` with the name `Comb_powerplots.pdf`.

Making the Table of Powers

The `data.frame` named `power_df` will be loaded in the R environment after successful compilation of the preceding chunk. We save it in the Excel file `power_table.xlsx` inside the `propensity_score_sims` folder.

```
power_df$Power=round(as.numeric(power_df$Power),3)
power_df$Scenario[power_df$Scenario=="Mixture"]="MIX"
power_df$Scenario[power_df$Scenario=="Mixture2"]="CAM"
power_df_save=power_df[order(power_df$Method,power_df$Scenario, power_df$delta,
power_df$p, power_df$nsamp1),c("Method","Scenario","delta", "nsamp1", "p", "Power")]
write_xlsx(power_df_save,paste0(output_dir,"power_table.xlsx"))
```

A slightly differently organized version of this table is shown in Tables S.1 and S.3 (upper) in the supplementary materials.

Interaction and Oracle Scenarios

Simulate Data

First set the location of this directory and run the following in R to generate the simulated data.

```
input_dir=paste0(getwd(),"/gen_data_codes/")
output_dir=paste0(getwd(),"/simData/")

source("gen_data_codes/Interaction.R")
source("gen_data_codes/Oracle.R")
```

The simulated data will be stored in the `simData` folder.

Analyze the Simulated Data

CA-PPMx Method

Run the following in R to analyze the data using our proposed CA-PPMx method. On multicore processors it runs parallelly. However, depending on the configuration of the computer, it might still need days to finish.

```
input_dir=paste0(getwd(),"/simData/")
output_dir=paste0(getwd(),"/GBM_sims/Outputs/")
source("GBM_sims/GBM_analysis_linear_CAPPMx.R")
gc()
```

```
##           used   (Mb) gc trigger   (Mb)    max used   (Mb)
## Ncells  3579227 191.2   6310714  337.1    6310714   337.1
## Vcells 447762540 3416.2 1373669010 10480.3 2146357707 16375.5
```

The outputs will be stored in `GBM_sims/Outputs` folder. Then we estimate the treatment effects:

```
input_dir=paste0(getwd(),"/GBM_sims/Outputs/")
output_dir=paste0(getwd(),"/GBM_sims/")
source("GBM_sims/GBM_gen_biases_CA-PPMx.R")
gc()
```

The outputs are stored in the folder `GBM_sims` with the name `CA-PPMx_GBM.txt`.

Propensity-score (PS) Based Method

The analyze the data using the PS-based methods:

```
input_dir=paste0(getwd(), "/simData/")
output_dir=paste0(getwd(), "/GBM_sims/")
source("GBM_sims/GBM_analysis_propscore.R")
```

The outputs are stored in the folder `GBM_sims` with the name `prop_score_GBM.txt`.

Matching Based Method

The analyze the data using the PS-based methods:

```
input_dir=paste0(getwd(), "/simData/")
output_dir=paste0(getwd(), "/GBM_sims/")
source("GBM_sims/GBM_analysis_optmatch.R")
```

The outputs are stored in the folder `GBM_sims` with the name `Matching_GBM.txt`.

Showing Population Equivalence

We generate the boxplots of the area under the receiver operating characteristic curve (AUC) under repeat simulations that we have used in **Figure 3(b)** of the main article to show empirical evidence towards dealing with the lack of randomization.

```
output_dir=paste0(getwd(), "/GBM_sims/")
simdir=paste0(getwd(), "/simData/")
loc_MCMC_outputs=paste0(getwd(), "/GBM_sims/Outputs/")
source("GBM_sims/GBM_get_auc.R")
source("GBM_sims/GBM_get_auc_random.R")
source("GBM_sims/GBM_auc_boxplots.R")
p1
```

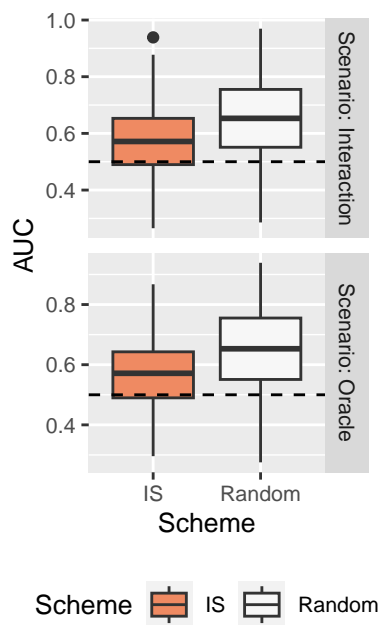


Figure 3: Figure 3(b) of the main article

```
ggsave(plot=p1,filename=paste0(output_dir,"GBM_AUC_boxplots.pdf"),width=3,height=5.5,units="in")
```

A pdf copy of this figure is saved in the folder GBM_sims with the name GBM_AUC_boxplots.pdf.

Compare Powers of Detecting Treatment Effects

Based on the above three analyzes we report the power of detecting treatment effects in the following figure which is also used in **Figure 4(b)** of the main article.

```
input_dir=output_dir=paste0(getwd(),"/GBM_sims/")
source("GBM_sims/GBM_gen_type_I_n_II_errors.R")
p_fig
```

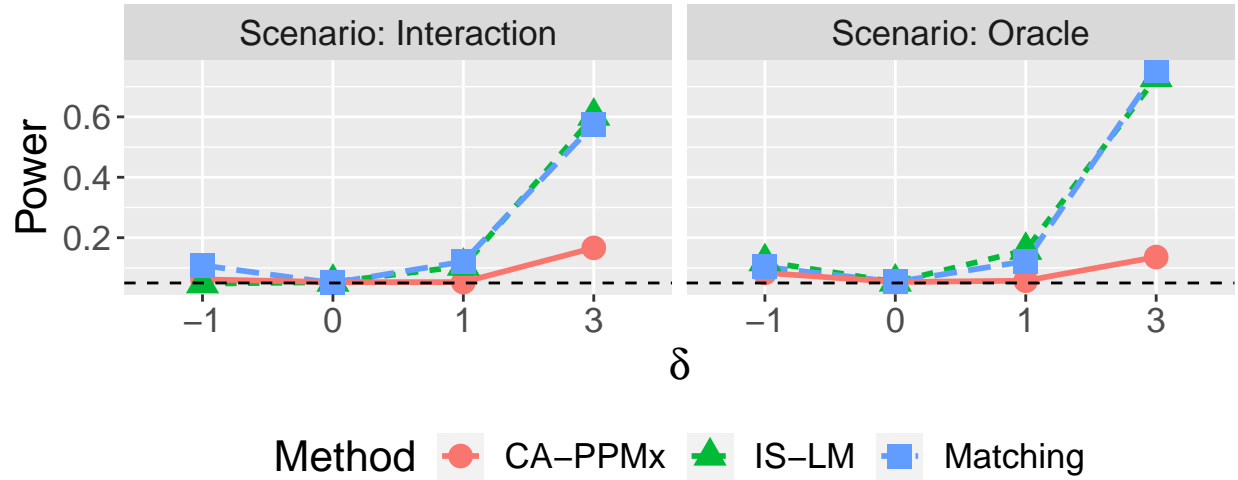


Figure 4: Figure 4(b) of the main article

```
ggsave(plot=p_fig,filename=paste0(output_dir,"GBM_power.pdf"),width=9,height=2.25,units="in")
```

A pdf copy of this figure is saved in the folder GBM_sims with the name GBM_power.pdf.

Making the Table of Powers

The data.frame named power1 will be loaded in the R environment after successful compilation of the preceding chunk. We save it in the Excel file GBM_power_table.xlsx inside the GBM_sims folder.

```
power1$Power=round(as.numeric(power1$Power),3)
dat_save=power1[order(power1$Method,power1$Scenario, power1$delta),
  c("Method","Scenario","delta","Power")]
write_xlsx(dat_save,paste0(output_dir,"GBM_power_table.xlsx"))
rm(dat_save)
```

A slight differently organized version of this table is shown in Tables S.2 and S.3 (lower) in the supplementary materials.