

# Estimate the Treatment Effect of Simulated Data on an Outcome Variable Y

Jing Fu

December 2024

## Abstract

**Background:** Cluster randomized trials (CRTs) are commonly used to evaluate interventions applied at the group level, such as in schools or clinics. Designing CRTs involves balancing the number of clusters and the number of observations per cluster to optimize statistical efficiency while staying within budget constraints. These decisions are further complicated by factors such as intra-cluster correlation and differing costs for initial and additional samples in each cluster. Simulation studies provide a powerful tool to systematically evaluate design options under varying budget constraints and data-generating conditions, allowing researchers to identify optimal designs for estimating treatment effects accurately and efficiently.

**Methods:** In this study, we use a simulation-based approach to evaluate the design of cluster randomized trials (CRTs) under fixed budget constraints. The hierarchical model assumes that the outcome variable (Y) is influenced by fixed effects (treatment assignment) and random effects (cluster-level variability). We systematically vary key parameters, including the treatment effect size ( $\beta$ ), cluster-level variance ( $\gamma^2$ ), and within-cluster variance ( $\sigma^2$ ), while considering realistic cost structures where the first sample in a cluster (c1) is more expensive than subsequent samples (c2). Performance metrics, such as bias, standard error, and coverage probability, are calculated to assess the accuracy and precision of the treatment effect estimates. Additionally, the study explores how varying the cost ratio (c1/c2) and data generation parameters impact the optimal allocation of resources.

**Results:** Our simulation study evaluated the impact of underlying data generation parameters and cost ratios on the estimation of treatment effects. The results demonstrated that bias remained minimal across all parameter settings, while standard errors increased with larger  $\beta$  and  $\gamma^2$ , reflecting heightened variability in these conditions. Coverage rates were consistent across most settings but slightly reduced under higher  $\alpha^2$  values. Cost ratios significantly influenced the allocation of clusters and observations per cluster, with lower c1/c2 ratios favoring smaller clusters and more observations, resulting in reduced bias and variability.

**Conclusions:** Optimal study designs depend on balancing the trade-offs between increasing the number of clusters and observations per cluster. Lower c1/c2 ratios were most efficient in minimizing bias and maintaining high precision, particularly when within-cluster variance was low. These findings emphasize the importance of aligning cost allocation strategies with variance components to achieve accurate and reliable treatment effect estimates.

## Introduction

Cluster randomized trials (CRTs) are a commonly used design for evaluating interventions implemented at the group level, such as schools, hospitals, or communities. Compared to individually randomized trials, CRTs face unique challenges, including intra-cluster correlation, which reduces statistical efficiency, and the need to balance limited budgets with optimal design choices. A key question in CRT design is how to

allocate resources effectively between the number of clusters (G) and the number of observations per cluster (R), particularly when costs vary for the first sample (c1) and subsequent samples (c2). Optimizing these trade-offs is essential for ensuring that CRTs achieve sufficient statistical power while remaining cost-efficient (Donner & Klar, 2000).

This study addresses these challenges by using simulation-based methods to optimize CRT designs under fixed budget constraints. A hierarchical model is employed to simulate the outcome variable (Y), incorporating both fixed effects (treatment assignment) and random effects (cluster-level variability). By systematically varying key parameters, including treatment effect size ( $\beta$ ), cluster-level variance ( $\gamma^2$ ), and within-cluster variance ( $\sigma^2$ ), we evaluate how design decisions impact the estimation of the treatment effect. The inclusion of realistic cost structures, where  $c1 > c2$ , allows for an in-depth exploration of how cost ratios ( $c1/c2$ ) influence optimal design choices.

Our simulation study evaluates performance metrics such as bias, standard error, and coverage probability to assess the accuracy and precision of treatment effect estimates. Furthermore, we extend the analysis to Poisson-distributed outcomes to explore broader applicability in count data scenarios, which are common in health and social research. This approach provides actionable insights into efficient design strategies for CRTs, aligning statistical objectives with real-world constraints (Hemming et al., 2011).

## Methods

### Aims

The objective of this simulation study is to determine the optimal experimental design for estimating the treatment effect  $\beta$  under a fixed budget  $B$ . The specific aims are as follows:

**AIM 1:** Evaluate the impact of varying the number of clusters (G) and the number of observations per cluster (R) on the estimation of  $\beta$ .

**AIM 2:** Explore relationships between the underlying data generation mechanism parameters (e.g.,  $\alpha$ ,  $\beta$ ) and the relative costs  $c1/c2$  and how these impact the optimal study design.

**AIM 3:** Extend the simulation study to settings where the outcome variable  $Y$  follows a Poisson distribution and assess how this impacts the results compared to the normal distribution assumption.

### Data-generating mechanisms

For each observation  $j$  in cluster  $i$ , the outcome  $Y_{ij}$  was generated based on the following models:

#### 1. Normal Distribution Model:

$\mu_i = \alpha + \beta * X_i + \epsilon_i$ , where  $\epsilon_i \sim N(0, \gamma^2)$ ,  $Y_{ij} | \mu_i \sim N(\mu_i, \sigma^2)$

#### 2. Poisson Distribution Model: $\log(\mu_i) \sim N(\alpha + \beta * X_i, \gamma^2)$ , $Y_{ij} | \mu_i \sim \text{Poisson}(\mu_i)$

For both models,  $X_i$  is a binary indicator for the treatment group ( $X_i = 1$  for treatment,  $X_i = 0$  for control). Observations within the same cluster ( $j = 1, \dots, R$ ) are correlated due to shared cluster-specific random effects.

### Estimands

The primary estimand of interest is the average treatment effect  $\beta$ . The average treatment effect ( $\beta$ ) is estimated using hierarchical models that account for the clustered structure of the data. For datasets where the outcome is normally distributed, a linear mixed-effects model is used. This model includes a fixed effect for the treatment variable and a random effect to capture variability at the cluster level. For datasets with a

Poisson-distributed outcome, a generalized linear mixed-effects model is employed, using a log link function to model the count data.

In both cases, the fixed effect associated with the treatment variable provides the estimate of  $\beta$ , which represents the average difference in the outcome between the treatment and control groups. The standard error of this estimate is derived from the variance-covariance matrix of the model.

## Methods

The simulation study will systematically vary the number of clusters (G) and the number of observations per cluster (R) to evaluate their impact on performance metrics.

The following settings will be considered:

**Design Parameters:**  $G * c_1 + G * (R - 1) * c_2 = B$ , we set  $G > 10, R > 2$  for convenience.

**Budget Constraint:** Total budget  $B$ , with costs  $c_1$  for the first observation in a cluster and  $c_2 < c_1$  for additional observations.  $B = 1000$  and  $(c_1, c_2)$  pairs such as (20,5), (30, 10) will be used

**Simulation Scenarios:** Parameters:

$$\alpha \in \{1, 5\}, \beta \in \{0.1, 0.5\}, \gamma^2 \in \{0.1, 0.5\}, \sigma^2 \in \{1, 2\}$$

### Other Setting:

Each simulation scenario will be replicated 1000 times, and I set 202412 as our random seed.

## Performance measures

The following performance measures will be computed for each design scenario:

**Bias:**

$$\text{Bias} = \frac{1}{n_{sim}} \sum_{i=1}^n \hat{\beta}_i - \beta$$

**Standard Error:**

$$\text{SE} = \sqrt{\frac{1}{(n_{sim} - 1)} \sum_{i=1}^n (\hat{\beta}_i - \bar{\beta})^2}$$

**Coverage:**

$$\text{Coverage} = \frac{1}{n_{sim}} \sum_{i=1}^n \Pi(\beta \in \text{CI}_i)$$

# Results

## Normal Setting

Evaluate the impact of  $G$  and  $R$  on the estimation of treatment effect.

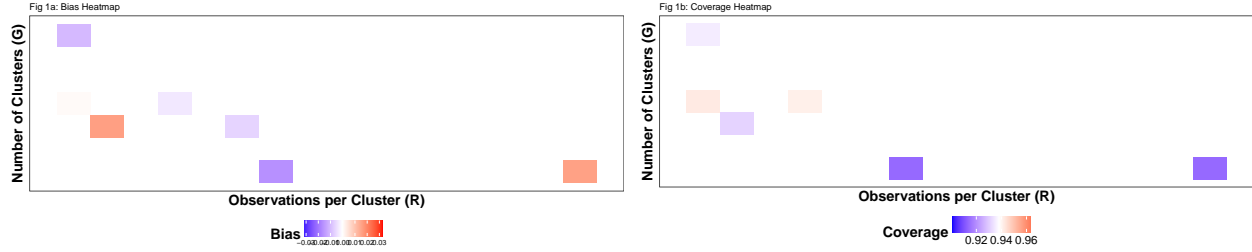


Figure 1a represents the effects of  $G$  and  $R$  on bias. The bias appears to increase slightly as the number of clusters decreases and the number of observations per cluster  $R$  increases. The bias values are closer to zero when there is a balanced combination of a moderate number of clusters and observations per cluster, suggesting that overly small or overly large  $R$  can increase bias.

Figure 1b represents the impact of  $G$  and  $R$  on coverage. Coverage is generally higher (closer to 0.96) for designs with larger  $G$  and smaller  $R$ , suggesting that spreading observations across more clusters helps maintain better interval coverage for the estimation of  $\beta$ . Designs with a small number of clusters ( $G$ ) and many observations per cluster ( $R$ ) show reduced coverage, indicating that such designs may fail to capture the true variability of the treatment effect.

Explore relationships between the underlying data generation mechanism parameters and the relative costs  $c_1/c_2$  and how these impact the optimal study design.

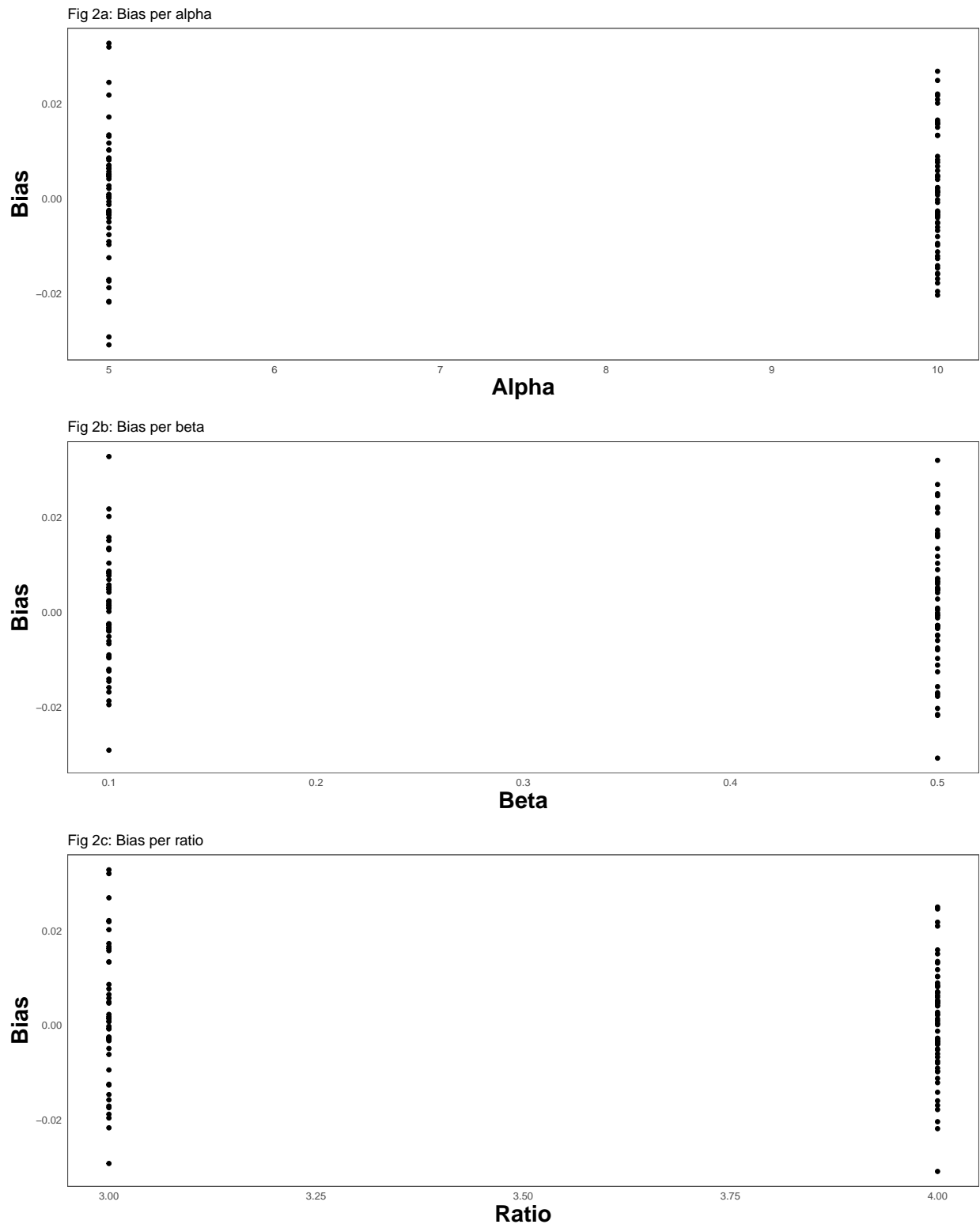


Figure 2a isolates the relationship between the baseline mean  $\alpha$  and bias, disregarding other parameters

such as  $\beta$ ,  $G$ ,  $R$ . The results shows the relationship between the baseline mean  $\alpha$  and the bias in treatment effect estimation. Across different values of  $\alpha$ , the bias remains consistently close to zero, indicating that the baseline mean does not significantly influence the accuracy of the treatment effect estimation. This stability highlights the robustness of the estimation process concerning changes in the baseline mean. However, it is important to note that this conclusion holds only when other parameters remain constant. Variations in other factors might introduce interactions that were not captured in this analysis.

Figure 2b examines the effect of the true treatment effect ( $\beta$ ) on bias while assuming other parameters are constant. The bias is consistently small for both low (0.1) and high (0.5) values of  $\beta$ . The lack of substantial variation suggests that the treatment effect size does not substantially affect bias in estimation. This implies that study designs do not need to be heavily tailored based on expectations of treatment effect magnitude.

Figure 2c illustrates the relationship between the cost ratio ( $c1/c2$ ) and bias, independent of other factors like  $\alpha$ ,  $\beta$ ,  $G$ ,  $R$ . Bias values remain close to zero across different cost ratios, indicating that varying the relative costs of cluster versus within-cluster observations does not significantly impact bias. This result suggests that decisions regarding the allocation of budget between increasing cluster size ( $G$ ) versus within-cluster observations ( $R$ ) can be guided by practical considerations, such as feasibility and resource availability, without compromising the accuracy of treatment effect estimation.

						Table 1: Sensitivity Analysis Summary		
alpha	beta	gamma2	sigma2	c1	c2	mean_bias	mean_se	mean_coverage
5	0.1	0.1	1	20	5	0.0017676	0.2447802	0.9335000
5	0.1	0.1	1	30	10	0.0013768	0.3110922	0.9410000
5	0.1	0.1	2	20	5	-0.0033722	0.3060074	0.9432500
5	0.1	0.1	2	30	10	-0.0033030	0.4049642	0.9476667
5	0.1	0.5	1	20	5	0.0039114	0.3848542	0.9315000
5	0.1	0.5	1	30	10	-0.0170419	0.4524274	0.9306667
5	0.1	0.5	2	20	5	0.0084099	0.4362161	0.9282500
5	0.1	0.5	2	30	10	0.0092262	0.5367051	0.9326667
5	0.5	0.1	1	20	5	0.0028506	0.2360768	0.9400000
5	0.5	0.1	1	30	10	0.0069673	0.3202258	0.9323333
5	0.5	0.1	2	20	5	-0.0062681	0.3008551	0.9487500
5	0.5	0.1	2	30	10	0.0025239	0.4102378	0.9446667
5	0.5	0.5	1	20	5	-0.0032104	0.3841242	0.9317500
5	0.5	0.5	1	30	10	0.0054793	0.4481004	0.9313333
5	0.5	0.5	2	20	5	0.0118621	0.4329209	0.9285000
5	0.5	0.5	2	30	10	-0.0098933	0.5164046	0.9373333
10	0.1	0.1	1	20	5	0.0051355	0.2404416	0.9382500
10	0.1	0.1	1	30	10	0.0077042	0.3093508	0.9433333
10	0.1	0.1	2	20	5	0.0013264	0.3129707	0.9387500
10	0.1	0.1	2	30	10	-0.0090400	0.4063678	0.9553333
10	0.1	0.5	1	20	5	-0.0063357	0.3834283	0.9350000
10	0.1	0.5	1	30	10	0.0043838	0.4649597	0.9256667
10	0.1	0.5	2	20	5	-0.0053195	0.4319728	0.9252500
10	0.1	0.5	2	30	10	0.0029643	0.5201911	0.9403333
10	0.5	0.1	1	20	5	-0.0039151	0.2433780	0.9350000
10	0.5	0.1	1	30	10	0.0041514	0.3069675	0.9493333
10	0.5	0.1	2	20	5	-0.0002462	0.3060225	0.9420000
10	0.5	0.1	2	30	10	0.0103204	0.4142842	0.9423333
10	0.5	0.5	1	20	5	0.0051649	0.3819082	0.9337500
10	0.5	0.5	1	30	10	0.0080792	0.4561284	0.9250000
10	0.5	0.5	2	20	5	-0.0007708	0.4266267	0.9320000
10	0.5	0.5	2	30	10	0.0004500	0.5380379	0.9290000

Table 1 shows the impact of the parameters of the data generation and  $c1/c2$  on the optimal study design under normal setting. For  $c1/c2 = 4$  (e.g.,  $c1 = 20$ ,  $c2 = 5$ ), the mean standard errors are generally lower across all parameter combinations compared to  $c1/c2 = 3$  (e.g.,  $c1 = 30$ ,  $c2 = 10$ ). Coverage probabilities are slightly better aligned with the 95% level for  $c1/c2 = 4$ , suggesting more precise and reliable estimates. Higher  $\alpha$  values tend to result in marginally higher standard errors compared to lower  $\alpha$ , particularly

for larger  $\sigma^2$  or smaller clusters. The mean bias remains small and consistent across different  $\alpha$  values, indicating robustness to changes in the baseline mean. Larger  $\beta$  values (e.g.,  $\beta = 0.5$ ) are associated with higher standard errors and slightly reduced coverage probabilities compared to smaller  $\beta$ . Bias remains negligible across all combinations, indicating accurate estimation of  $\beta$ . Higher within-cluster variance ( $\sigma^2 = 2$ ) leads to increased standard errors and reduced coverage probabilities, especially when the cost ratio is less favorable ( $c1/c2 = 3$ ). Cluster variance ( $\gamma^2$ ) has a smaller impact compared to  $\sigma^2$ , though higher values of  $\gamma^2$  slightly increase the variability of estimates.

## Poisson Setting

Evaluate the impact of G and R on the estimation of treatment effect.

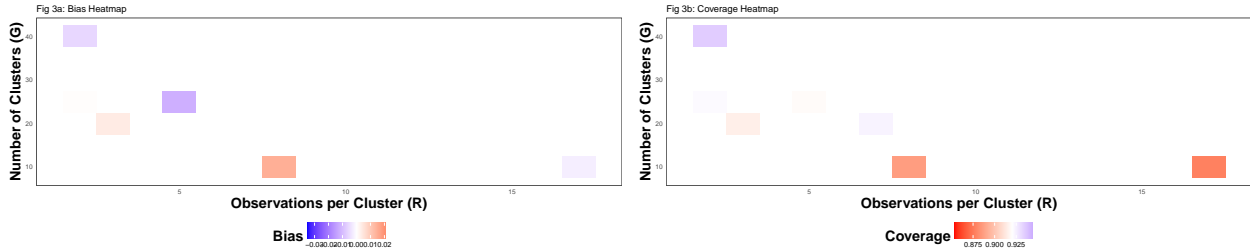


Fig 3a represents the bias in the estimation of  $\beta$  demonstrates clear variation across different combinations of G and R. Lower numbers of clusters combined with higher numbers of observations per cluster R show greater bias, as seen in the red regions of the heatmap. Conversely, increasing the number of clusters G while keeping the number of observations per cluster R moderate to low reduces bias, as shown in the blue regions. This suggests that increasing the number of clusters contributes more to reducing bias than increasing the number of observations within each cluster.

Fig 3b represents the coverage of the confidence intervals also varies with G and R. Lower coverage (red regions) is observed when G is small and R is large. This pattern indicates that the confidence intervals may not fully capture the true parameter value under these conditions. Higher coverage (purple regions) is achieved with a greater number of clusters G, particularly when the number of observations per cluster R is moderate or low. This finding reinforces the importance of increasing the number of clusters to improve the reliability of the parameter estimates.

It is important to note that in this analysis other variables (such as G, R) are held constant or are neglect when we evaluate the relationships between key data generation parameters and the bias in the estimation of  $\beta$ .

Explore relationships between the underlying data generation mechanism parameters and the relative costs  $c_1/c_2$  and how these impact the optimal study design.

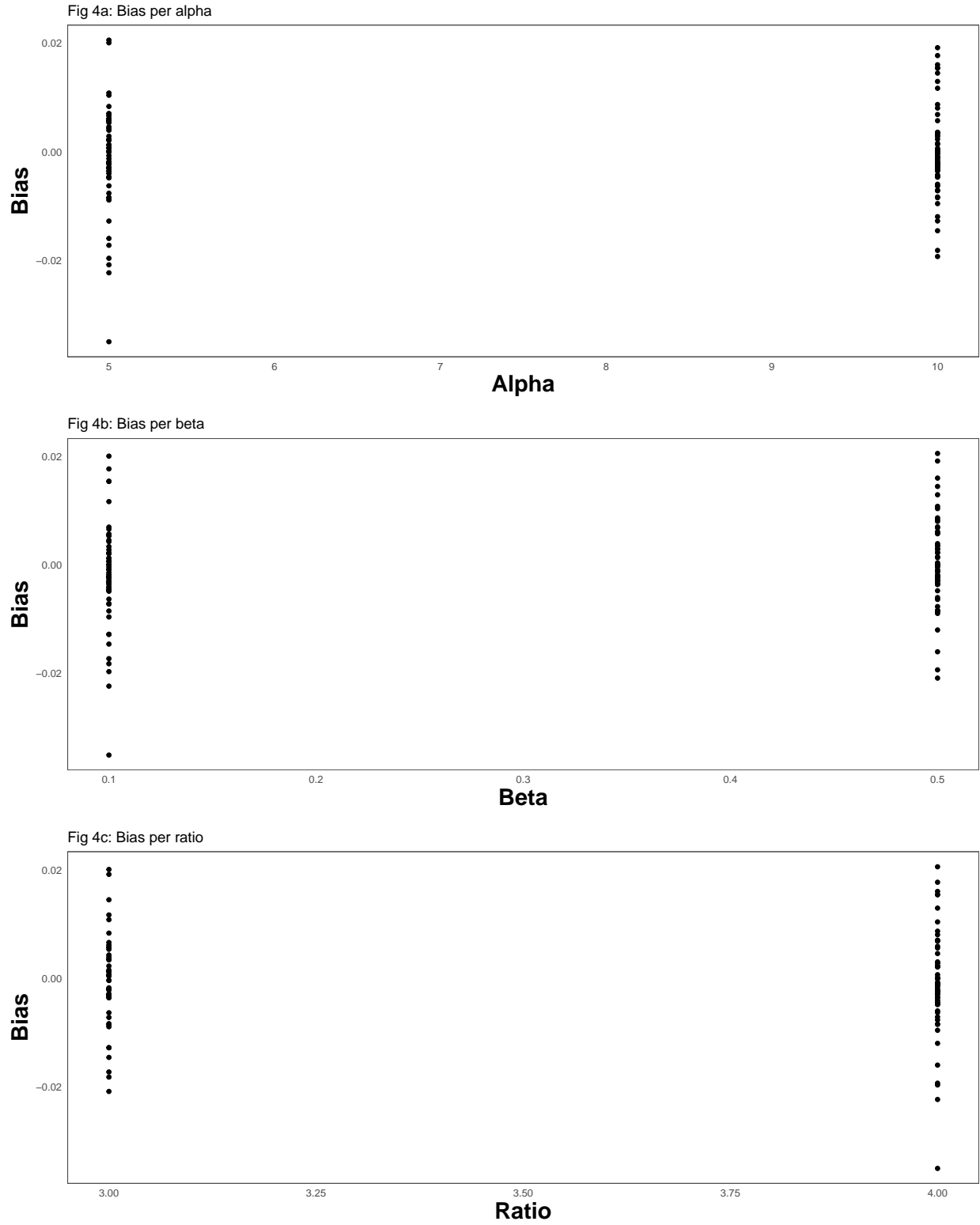


Fig 4a demonstrates the variation in bias with changes in the baseline mean  $\alpha$ . Across the tested range,



the bias remains relatively small and centered around zero, indicating that  $\alpha$  has a minimal effect on bias under the current simulation settings. This suggests that the baseline mean does not strongly influence the reliability of the  $\beta$  estimate, which supports flexibility in selecting  $\alpha$  values for study design.

Fig 4b shows the relationship between  $\beta$  and bias reveals that the bias is consistently close to zero regardless of the  $\beta$  value. This indicates that the estimator is robust to varying magnitudes of the treatment effect.

Fig 4c explores the impact of the cost ratio on bias. The bias remains negligible across all tested ratios of  $c1/c2$ . This finding highlights that the cost structure of the study—whether the number of clusters or the number of observations per cluster dominates the budget—does not introduce systematic bias into the estimation. Therefore, study designers can prioritize cost considerations without compromising estimation accuracy.

alpha	beta	gamma2	sigma2	c1	c2	Table 2: Sensitivity Analysis Summary		
						mean_bias	mean_se	mean_coverage
5	0.1	0.1	1	20	5	0.0013530	0.1488384	0.9170000
5	0.1	0.1	1	30	10	-0.0000728	0.1620887	0.9160000
5	0.1	0.1	2	20	5	-0.0020052	0.1478945	0.9200000
5	0.1	0.1	2	30	10	-0.0003532	0.1680354	0.9060000
5	0.1	0.5	1	20	5	-0.0021932	0.3305766	0.9097500
5	0.1	0.5	1	30	10	-0.0031404	0.3607195	0.9200000
5	0.1	0.5	2	20	5	-0.0153774	0.3361816	0.9115000
5	0.1	0.5	2	30	10	0.0038769	0.3736540	0.8993333
5	0.5	0.1	1	20	5	-0.0002655	0.1476912	0.9147500
5	0.5	0.1	1	30	10	-0.0037737	0.1679295	0.9043333
5	0.5	0.1	2	20	5	-0.0020861	0.1484720	0.9162500
5	0.5	0.1	2	30	10	0.0002677	0.1628062	0.9110000
5	0.5	0.5	1	20	5	0.0064759	0.3270643	0.9167500
5	0.5	0.5	1	30	10	0.0031530	0.3664666	0.9036667
5	0.5	0.5	2	20	5	-0.0043902	0.3342630	0.9162500
5	0.5	0.5	2	30	10	-0.0025730	0.3651610	0.9056667
10	0.1	0.1	1	20	5	-0.0020938	0.1492101	0.9142500
10	0.1	0.1	1	30	10	0.0002546	0.1625565	0.9096667
10	0.1	0.1	2	20	5	-0.0037179	0.1471368	0.9182500
10	0.1	0.1	2	30	10	0.0010886	0.1668634	0.9126667
10	0.1	0.5	1	20	5	0.0015954	0.3301336	0.9135000
10	0.1	0.5	1	30	10	-0.0015320	0.3571111	0.9190000
10	0.1	0.5	2	20	5	0.0049812	0.3336157	0.9100000
10	0.1	0.5	2	30	10	-0.0127155	0.3642886	0.9090000
10	0.5	0.1	1	20	5	-0.0058537	0.1496804	0.9115000
10	0.5	0.1	1	30	10	-0.0034190	0.1653515	0.9070000
10	0.5	0.1	2	20	5	0.0072148	0.1462739	0.9175000
10	0.5	0.1	2	30	10	0.0003269	0.1650990	0.9033333
10	0.5	0.5	1	20	5	0.0062100	0.3268612	0.9222500
10	0.5	0.5	1	30	10	0.0047981	0.3652501	0.9126667
10	0.5	0.5	2	20	5	-0.0051193	0.3358005	0.9140000
10	0.5	0.5	2	30	10	0.0061751	0.3656530	0.9080000

Table 2 shows the impact of the parameters of the data generation and  $c1/c2$  on the optimal study design under poisson setting. Across different  $\alpha$  levels (e.g., 5 vs. 10), the mean bias remains small and relatively stable, suggesting that the baseline mean does not strongly influence the accuracy of  $\beta$ 's estimation. However, slight differences are seen in coverage rates, with higher  $\alpha$  values tending to exhibit marginally better coverage. Increasing  $\beta$  (e.g., 0.1 to 0.5) generally leads to increased standard errors, which could indicate higher variability in treatment effect estimation as the true effect becomes larger. Bias remains minimal across all  $\beta$  values, suggesting the model is robust in estimating the treatment effect across different effect sizes. Higher  $\gamma^2$  values (0.1 vs. 0.5) correspond to increased variability in mean bias and standard errors, highlighting the influence of cluster-level random effects. Coverage rates tend to remain fairly consistent across different  $\gamma^2$  values, suggesting the model's robustness in accounting for between-cluster variance. Variations in  $\sigma^2$  (1 vs. 2) slightly affect both mean bias and standard errors, but the effects are less pronounced compared

to  $\gamma^2$ . Coverage rates are slightly reduced for higher  $\sigma^2$  values, indicating that increased within-cluster variability may pose challenges for maintaining nominal confidence interval coverage. The relative cost ratio  $c1/c2$  impacts the sample size allocation, influencing the number of clusters  $G$  and observations per cluster  $R$ . For example: Lower  $c1/c2$  ratios (e.g., 20/5) favor smaller clusters with more observations per cluster, resulting in reduced bias and slightly lower standard errors. Higher  $c1/c2$  ratios (e.g., 30/10) allocate more resources to increasing the number of clusters, which can stabilize coverage but may slightly increase standard errors.

## Discussion

Our simulation study explored the relationship between data generation parameters and cost ratios in determining the optimal design for estimating treatment effects. The results showed that bias was consistently close to zero across parameter settings, indicating robust estimation of  $\beta$ . However, standard errors increased with larger values of  $\gamma$  and  $\beta$ , reflecting higher variability in scenarios with greater cluster-level heterogeneity and stronger treatment effects. Coverage rates generally remained high, though slight reductions were observed in settings with higher within-cluster variance, suggesting potential challenges in maintaining confidence interval reliability under these conditions. Additionally, cost ratios had a significant influence on the distribution of clusters and observations per cluster, with lower ratios favoring designs with more observations per cluster and reduced bias.

### Strengths and Limitations

While the proposed simulation framework offers a comprehensive exploration of parameter impacts on study design, several limitations must be acknowledged. First, the assumption of normally or Poisson-distributed outcomes may not fully capture the complexities of real-world data, where outcome distributions could be more complex or skewed. Second, the algorithm ignores potential interactions between parameters (e.g.,  $\alpha$ ,  $\sigma$ ), which may lead to oversimplified interpretations of how these parameters jointly affect bias and precision. Additionally, the reliance on pre-specified cost ratios may limit the generalizability of findings to settings with highly variable cost structures or resource constraints.

## References

- [1] Donner, A., & Klar, N. (2000). Design and analysis of cluster randomization trials in health research. Arnold.
- [2] Hemming, K., Haines, T. P., Chilton, P. J., Girling, A. J., & Lilford, R. J. (2011). The stepped wedge cluster randomised trial: rationale, design, analysis, and reporting. *BMJ*, 343, d6024. <https://doi.org/10.1136/bmj.d6024>

## Code Appendix:

```
knitr::opts_chunk$set(echo = FALSE)
library(ggplot2)
library(lme4)
library(dplyr)
library(knitr)
library(kableExtra)
library(gtsummary)
library(gridExtra)
## load simulated data
### normal setting
normal_results <- read.csv("/Users/fusei/Desktop/24FALL/PHP2550/Project3/normal_simulation.csv")
### poisson setting
poisson_results <- read.csv("/Users/fusei/Desktop/24FALL/PHP2550/Project3/poi_simulation.csv")
# bias
p1 <- ggplot(normal_results, aes(x = R, y = G, fill = bias)) +
  geom_tile() +
  scale_fill_gradient2(low = "blue", high = "red") +
  labs(title = "Fig 1a: Bias Heatmap", x = "Observations per Cluster (R)", y = "Number of Clusters (G)")
  theme(plot.margin = margin(0.3, 0.3, 0.3, 0.3, "cm"),
        legend.position="bottom",
        panel.background = element_blank(),
        plot.background = element_blank(),
        panel.border = element_rect(colour = "grey30", fill=NA, size=0.5),
        axis.text.y = element_blank(),
        axis.text.x = element_blank(),
        axis.title = element_text(size = 20,face="bold"),
        legend.title=element_text(size = 20,face="bold"),
        legend.text=element_text(size = 10),
        axis.ticks = element_blank(),
        #axis.text.x = element_blank(),
        legend.key = element_rect(colour = "transparent", fill = "white"),
        legend.key.size = unit(1.5, 'lines'),
        strip.text = element_text(size = 16,face = "bold"))

# coverage
p2 <- ggplot(normal_results, aes(x = R, y = G, fill = coverage)) +
  geom_tile() +
  scale_fill_gradient2(low = "blue", mid = "white", high = "red", midpoint = median(normal_results$coverage))
  labs(title = "Fig 1b: Coverage Heatmap", x = "Observations per Cluster (R)", y = "Number of Clusters (G)")
  theme(plot.margin = margin(0.3, 0.3, 0.3, 0.3, "cm"),
        legend.position="bottom",
        panel.background = element_blank(),
        plot.background = element_blank(),
        panel.border = element_rect(colour = "grey30", fill=NA, size=0.5),
        axis.text.y = element_blank(),
        axis.text.x = element_blank(),
        axis.title = element_text(size = 20,face="bold"),
        legend.title=element_text(size = 20,face="bold"),
        legend.text=element_text(size = 16),
        axis.ticks = element_blank(),
        #axis.text.x = element_blank(),
```

```

        legend.key = element_rect(colour = "transparent", fill = "white"),
        legend.key.size = unit(1.5, 'lines'),
        strip.text = element_text(size = 16, face = "bold"))

print(p1)
print(p2)
### bias by alpha
p3 <- ggplot(normal_results, aes(x = alpha, y = bias)) +
  geom_point() +
  labs(title = "Fig 2a: Bias per alpha", x = "Alpha", y = "Bias") +
  theme(plot.margin = margin(0.3, 0.3, 0.3, 0.3, "cm"),
        legend.position="bottom",
        panel.background = element_blank(),
        plot.background = element_blank(),
        panel.border = element_rect(colour = "grey30", fill=NA, size=0.5),
        axis.title = element_text(size = 20, face="bold"),
        legend.title=element_text(size = 20, face="bold"),
        legend.text=element_text(size = 16),
        axis.ticks = element_blank(),
        #axis.text.x = element_blank(),
        legend.key = element_rect(colour = "transparent", fill = "white"),
        legend.key.size = unit(1.5, 'lines'),
        strip.text = element_text(size = 16, face = "bold"))

# bias by beta
p4 <- ggplot(normal_results, aes(x = beta, y = bias)) +
  geom_point() +
  labs(title = "Fig 2b: Bias per beta", x = "Beta", y = "Bias") +
  theme(plot.margin = margin(0.3, 0.3, 0.3, 0.3, "cm"),
        legend.position="bottom",
        panel.background = element_blank(),
        plot.background = element_blank(),
        panel.border = element_rect(colour = "grey30", fill=NA, size=0.5),
        axis.title = element_text(size = 20, face="bold"),
        legend.title=element_text(size = 20, face="bold"),
        legend.text=element_text(size = 16),
        axis.ticks = element_blank(),
        #axis.text.x = element_blank(),
        legend.key = element_rect(colour = "transparent", fill = "white"),
        legend.key.size = unit(1.5, 'lines'),
        strip.text = element_text(size = 16, face = "bold"))

# bias by c1/c2
p5 <- ggplot(normal_results, aes(x = c1/c2, y = bias)) +
  geom_point() +
  labs(title = "Fig 2c: Bias per ratio", x = "Ratio", y = "Bias") +
  theme(plot.margin = margin(0.3, 0.3, 0.3, 0.3, "cm"),
        legend.position="bottom",
        panel.background = element_blank(),
        plot.background = element_blank(),
        panel.border = element_rect(colour = "grey30", fill=NA, size=0.5),
        axis.title = element_text(size = 20, face="bold"),
        legend.title=element_text(size = 20, face="bold"),

```

```

        legend.text=element_text(size = 16),
        axis.ticks = element_blank(),
        #axis.text.x = element_blank(),
        legend.key = element_rect(colour = "transparent", fill = "white"),
        legend.key.size = unit(1.5, 'lines'),
        strip.text = element_text(size = 16,face ="bold"))

p3
p4
p5
sensitivity_summary <- normal_results %>%
  group_by(alpha, beta, gamma2, sigma2, c1, c2) %>%
  summarize(
    mean_bias = mean(bias),
    mean_se = mean(se),
    mean_coverage = mean(coverage)
  )

kable(sensitivity_summary, format = "latex") %>%
  kableExtra::add_header_above(c(" " = 6, "Table 1: Sensitivity Analysis Summary" = 3)) %>%
  kableExtra::kable_styling(font_size = 8,
                             latex_options = c("repeat_header", "HOLD_position"))

# bias
p1 <- ggplot(poisson_results, aes(x = R, y = G, fill = bias)) +
  geom_tile() +
  scale_fill_gradient2(low = "blue", high = "red") +
  labs(title = "Fig 3a: Bias Heatmap", x = "Observations per Cluster (R)", y = "Number of Clusters (G)")
  theme(plot.margin = margin(0.3, 0.3, 0.3, 0.3, "cm"),
        legend.position="bottom",
        panel.background = element_blank(),
        plot.background = element_blank(),
        panel.border = element_rect(colour = "grey30", fill=NA, size=0.5),
        axis.title = element_text(size = 20,face="bold"),
        legend.title=element_text(size = 20,face="bold"),
        legend.text=element_text(size = 10),
        axis.ticks = element_blank(),
        #axis.text.x = element_blank(),
        legend.key = element_rect(colour = "transparent", fill = "white"),
        legend.key.size = unit(1.5, 'lines'),
        strip.text = element_text(size = 16,face ="bold"))

# coverage
p2 <- ggplot(poisson_results, aes(x = R, y = G, fill = coverage)) +
  geom_tile() +
  scale_fill_gradient2(low = "red", mid = "white", high = "blue", midpoint = median(poisson_results$coverage))
  labs(title = "Fig 3b: Coverage Heatmap", x = "Observations per Cluster (R)", y = "Number of Clusters (G)")
  theme(plot.margin = margin(0.3, 0.3, 0.3, 0.3, "cm"),
        legend.position="bottom",
        panel.background = element_blank(),
        plot.background = element_blank(),
        panel.border = element_rect(colour = "grey30", fill=NA, size=0.5),
        axis.title = element_text(size = 20,face="bold"),
        legend.title=element_text(size = 20,face="bold"),
        legend.text=element_text(size = 10),

```

```

axis.ticks = element_blank(),
#axis.text.x = element_blank(),
legend.key = element_rect(colour = "transparent", fill = "white"),
legend.key.size = unit(1.5, 'lines'),
strip.text = element_text(size = 16,face ="bold"))

p1
p2
### bias by alpha
p3 <- ggplot(poisson_results, aes(x = alpha, y = bias)) +
  geom_point() +
  labs(title = "Fig 4a: Bias per alpha", x = "Alpha", y = "Bias") +
  theme(plot.margin = margin(0.3, 0.3, 0.3, 0.3, "cm"),
    legend.position="bottom",
    panel.background = element_blank(),
    plot.background = element_blank(),
    panel.border = element_rect(colour = "grey30", fill=NA, size=0.5),
    axis.title = element_text(size = 20,face="bold"),
    legend.title=element_text(size = 20,face="bold"),
    legend.text=element_text(size = 16),
    axis.ticks = element_blank(),
    #axis.text.x = element_blank(),
    legend.key = element_rect(colour = "transparent", fill = "white"),
    legend.key.size = unit(1.5, 'lines'),
    strip.text = element_text(size = 16,face ="bold"))

# bias by beta
p4 <- ggplot(poisson_results, aes(x = beta, y = bias)) +
  geom_point() +
  labs(title = "Fig 4b: Bias per beta", x = "Beta", y = "Bias") +
  theme(plot.margin = margin(0.3, 0.3, 0.3, 0.3, "cm"),
    legend.position="bottom",
    panel.background = element_blank(),
    plot.background = element_blank(),
    panel.border = element_rect(colour = "grey30", fill=NA, size=0.5),
    axis.title = element_text(size = 20,face="bold"),
    legend.title=element_text(size = 20,face="bold"),
    legend.text=element_text(size = 16),
    axis.ticks = element_blank(),
    #axis.text.x = element_blank(),
    legend.key = element_rect(colour = "transparent", fill = "white"),
    legend.key.size = unit(1.5, 'lines'),
    strip.text = element_text(size = 16,face ="bold"))

# bias by c1/c2
p5 <- ggplot(poisson_results, aes(x = c1/c2, y = bias)) +
  geom_point() +
  labs(title = "Fig 4c: Bias per ratio", x = "Ratio", y = "Bias") +
  theme(plot.margin = margin(0.3, 0.3, 0.3, 0.3, "cm"),
    legend.position="bottom",
    panel.background = element_blank(),
    plot.background = element_blank(),
    panel.border = element_rect(colour = "grey30", fill=NA, size=0.5),

```

```

axis.title = element_text(size = 20,face="bold"),
legend.title=element_text(size = 20,face="bold"),
legend.text=element_text(size = 16),
axis.ticks = element_blank(),
#axis.text.x = element_blank(),
legend.key = element_rect(colour = "transparent", fill = "white"),
legend.key.size = unit(1.5, 'lines'),
strip.text = element_text(size = 16,face ="bold"))

p3
p4
p5
sensitivity_summary <- poisson_results %>%
  group_by(alpha, beta, gamma2, sigma2, c1, c2) %>%
  summarize(
    mean_bias = mean(bias),
    mean_se = mean(se),
    mean_coverage = mean(coverage)
  )

table2 <- kable(sensitivity_summary, format = "latex")
table2 %>%
  kableExtra::add_header_above(c(" " = 6, "Table 2: Sensitivity Analysis Summary" = 3)) %>%
  kableExtra::kable_styling(font_size = 8,
    latex_options = c("repeat_header", "HOLD_position"))

```