ASSESSING THE ROBUSTNESS OF META-ANALYSIS FOR THE FIXED-EFFECT MODEL

ASSESSING THE ROBUSTNESS OF META-ANALYSIS FOR THE FIXED-EFFECT MODEL

By YUQING HUANG, B.Sc.

A Thesis Submitted to the School of Graduate Studies in Partial Fulfilment of the Requirements for the Degree Master of Science

McMaster University

MASTER OF SCIENCE (2025)

Hamilton, Ontario, Canada (Department of Mathematics and Statistics)

TITLE:	Assessing the Robustness of Meta-Analysis for the Fixed-Effect Model
AUTHOR:	Yuqing Huang B.Sc. (Statistics)
SUPERVISOR:	Dr. Stephen Walter
NUMBER OF PAGES:	xxiii, 165

Abstract

The current meta-analysis methods for the fixed-effect model with continuous outcome variables have been developed based on the assumption that the variation of the outcome variable between patients within treatment groups for each study follows a normal distribution. However, real-world data does not always follow a normal distribution, which may lead to unreliable meta-analysis results.

This study uses the Monte Carlo simulation to evaluate robustness by comparing the analysis results with the truth when the normal assumption is violated; performance measures include the relative bias of the estimated treatment effect, the coverage probability of the estimates, and the power and type I error rate of the test of the null hypothesis. We simulate various non-normal outcome data, including a mixture of normals, lognormal, gamma, and χ^2 distributions. We examine the impact of the sample size per study, the number of studies, the magnitude of skewness, and the effect sizes on the results.

The results show that small studies with highly skewed data provide non-robust meta-analysis results for a fixed-effect model. Moreover, increasing the number of studies without sufficient sample sizes worsens the relative bias, coverage probability, and power. Therefore, this simulation suggests that investigators must be cautious when applying the fixed-effect model to small studies, particularly with respect to the potential non-normality of the data.

This study recommends that investigators include large trials whenever possible. If large trials are not feasible, they should always assess the normality of the datasets and select an appropriate meta-analysis method to obtain robust results. This will help ensure that policies and guidelines are based on reliable evidence, thereby minimizing the risk of implementing ineffective and harmful policies and guidelines.

Acknowledgments

My heartfelt gratitude goes to my supervisor, Dr. Stephen Walter, for his continuous mentorship and encouragement during my master's studies. I would not have completed this thesis without his invaluable suggestions and guidance. I feel extremely fortunate to be his student. I also appreciate the opportunity he provided to attend the academic conference, which allowed me to connect with other extraordinary scholars in the field and gain insights from them that enriched my master's experience to the fullest.

Also, I would like to thank Dr. Lehana Thabane and Dr. Shui Feng for being on my examining committee.

Also, thanks to my lovely friend Jessa, who provides me emotional support throughout the whole master's journey. I could not have passed the thesis defence without her help and encouragement in preparing for it. I am truly grateful to have a friend like her.

Lastly, I want to express my deepest appreciation to my parents and my grandfather for their emotional and financial support throughout my academic journey. They always offered inspiring words when I felt discouraged. Although physical distance keeps us apart, their unwavering support has strengthened me, and I would not have come this far without them.

Table of Contents

ABSTRACT ..... **III**

ACKNOWLEDGMENTS ..... **V**

LIST OF FIGURES ..... **VIII**

LIST OF TABLES ..... **XXIII**

CHAPTER 1 INTRODUCTION **1**

CHAPTER 2 LITERATURE REVIEW **4**

2.1 MEAN DIFFERENCE AND STANDARDIZED MEAN DIFFERENCE 4

2.1.1 Mean Difference (MD) 4

2.1.2 Standardized Mean Difference (SMD) 5

2.2 META-ANALYSIS 7

2.2.1 Fixed-Effect Model 8

2.3 NON-NORMALITY 9

2.3.1 Causes of Non-normality 10

2.3.2 Problems of Non-normality 11

CHAPTER 3 METHOD **15**

3.1 SIMULATION 15

3.2 DISTRIBUTIONS 16

3.2.1 The Mixture of Normal Distributions 16

3.2.2 Lognormal Distribution 17

3.2.3 Gamma Distribution 20

3.2.4 $\chi 2$ Distribution 22

3.3 EVALUATION CRITERIA 23

3.3.1 Relative Bias 23

3.3.2 Coverage Probability 24

3.3.3 Power 24

3.3.4 Type I Error 25

CHAPTER 4 RESULTS 26

4.1 The Mixture of Normal Distributions 26

4.1.1 Mean Difference 27

4.1.2 Standardized Mean Difference 41

4.2 Lognormal Distribution 56

4.2.1 Mean Difference 56

4.2.2 Standardized Mean Difference 73

4.3 Gamma Distribution 89

4.3.1 Mean Difference 89

4.3.2 Standardized Mean Difference 105

4.4 $\chi 2$ Distribution 121

4.4.1 Mean Difference 121

4.4.2 Standardized Mean Difference 129

CHAPTER 5 CONCLUSION AND DISCUSSION 137

5.1 Summary of the Results 137

5.1.1 The Mixture of Normal Distributions 138

5.1.2 Lognormal Distribution 139

5.1.3 Gamma Distribution 140

5.1.4 $\chi 2$ Distribution 141

5.2 Limitations 142

5.3 Conclusion 143

REFERENCE 144

APPENDIX 149

List of Figures

Figure 4.1.1 The relative bias in the estimated treatment effect, as a function of p and n across different effect sizes, with a mixture of normals data using the mean difference for a fixed value of k at all levels of 2, 5, 10, and 20. 28
Figure 4.1.2 The relative bias in the estimated treatment effect, as a function of p and k across different effect sizes, with a mixture of normals data using the mean difference for a fixed value of n at all levels of 5, 10, 20, and 50. 29
Figure 4.1.3 The relative bias in the estimated treatment effect, as a function of p and the different combinations of k and n across different effect sizes, with a mixture of normals data using mean difference for a fixed total sample size of 200. 29
Figure 4.1.4 The coverage probability in the estimated treatment effect, as a function of p and n across different effect sizes, with a mixture of normals data using the mean difference for a fixed value of k at all levels of 2, 5, 10, and 20. 30
Figure 4.1.5 The coverage probability in the estimated treatment effect, as a function of p and k across different effect sizes, with a mixture of normals data using the mean difference for a fixed value of n at all levels of 5, 10, 20, and 50. 31
Figure 4.1.6 The coverage probability in the estimated treatment effect, as a function of p and the different combinations of k and n across different effect sizes, with a mixture of normals data using the mean difference for a fixed total sample size of 200. 31
Figure 4.1.7 The power in the test, as a function of p and n across different effect sizes, with a mixture of normals data using the mean difference for a fixed value of k at all levels of 2, 5, 10, and 20. 32
Figure 4.1.8 The power in the test, as a function of p and k across different effect sizes, with a mixture of normals data using the mean difference for a fixed value of n at all levels of 5, 10, 20, and 50. 33
Figure 4.1.9 The power in the test, as a function of p and different combinations of k and n across different effect sizes, with a mixture of normals data using the mean difference for a fixed total sample size of 200. 33

Figure 4.1.10 The relative bias in the estimated treatment effect, as a function of p and n across different effect sizes, with a mixture of normals data using the mean difference for a fixed value of k at all levels of 2, 5, 10, and 20. 35
Figure 4.1.11 The relative bias in the estimated treatment effect, as a function of p and k across different effect sizes, with a mixture of normals data using the mean difference for a fixed value of n at all levels of 5, 10, 20, and 50. 36
Figure 4.1.12 The relative bias in the estimated treatment effect, as a function of p and different combinations of k and n across different effect sizes, with a mixture of normals data using the mean difference for a fixed total sample size of 200. 36
Figure 4.1.13 The coverage probability in the estimated treatment effect, as a function of p and n across different effect sizes, with a mixture of normals data using the mean difference for a fixed value of k at all levels of 2, 5, 10, and 20. 37
Figure 4.1.14 The coverage probability in the estimated treatment effect, as a function of p and k across different effect sizes, with a mixture of normals data using the mean difference for a fixed value of n at all levels of 5, 10, 20, and 50. 38
Figure 4.1.15 The coverage probability in the estimated treatment effect, as a function of p and different combinations of k and n across different effect sizes, with a mixture of normals data using the mean difference for a fixed total sample size of 200. 38
Figure 4.1.16 The power in the test, as a function of p and n across different effect sizes, with a mixture of normals data using the mean difference for a fixed value of k at all levels of 2, 5, 10, and 20. 39
Figure 4.1.17 The power in the test, as a function of p and k across different effect sizes, with a mixture of normals data using the mean difference for a fixed value of n at all levels of 5, 10, 20, and 50. 40
Figure 4.1.18 The power in the test, as a function of p and different combinations of k and n across different effect sizes, with a mixture of normals data using the mean difference for a fixed total sample size of 200. 40

Figure 4.1.19 The relative bias in the estimated treatment effect, as a function of p and n across different effect sizes, with a mixture of normals data using the standardized mean difference for a fixed value of k at all levels of 2, 5, 10, and 20. 42
Figure 4.1.20 The relative bias in the estimated treatment effect, as a function of p and k across different effect sizes, with a mixture of normals data using the standardized mean difference for a fixed value of n at all levels of 5, 10, 20, and 50. 43
Figure 4.1.21 The relative bias in the estimated treatment effect, as a function of p and different combinations of k and n across different effect sizes, with a mixture of normals data using the standardized mean difference for a fixed total sample size of 200..... 43
Figure 4.1.22 The coverage probability in the estimated treatment effect, as a function of p and n across different effect sizes, with a mixture of normals data using the standardized mean difference for a fixed value of k at all levels of 2, 5, 10, and 20..... 45
Figure 4.1.23 The coverage probability in the estimated treatment effect, as a function of p and k across different effect sizes, with a mixture of normals data using the standardized mean difference for a fixed value of n at all levels of 5, 10, 20, and 50..... 46
Figure 4.1.24 The coverage probability in the estimated treatment effect, as a function of p and different combinations of k and n across different effect sizes, with a mixture of normals data using the standardized mean difference for a fixed total sample size of 200. 46
Figure 4.1.25 The power in the test, as a function of p and n across different effect sizes, with a mixture of normals data using the standardized mean difference for a fixed value of k at all levels of 2, 5, 10, and 20. 47
Figure 4.1.26 The power in the test, as a function of p and k across different effect sizes, with a mixture of normals data using the standardized mean difference for a fixed value of n at all levels of 5, 10, 20, and 50. 48
Figure 4.1.27 The power in the test as a function of p and different combinations of k and n across different effect sizes, with a mixture of normals data using the standardized mean difference for a fixed total sample size of 200. 48

Figure 4.1.28 The relative bias in the estimated treatment effect, as a function of p and n across different effect sizes, with a mixture of normals data using the standardized mean difference for a fixed value of k at all levels of 2, 5, 10, and 20. 50
Figure 4.1.29 The relative bias in the estimated treatment effect, as a function of p and k across different effect sizes, with a mixture of normals data using the standardized mean difference for a fixed value of n at all levels of 5, 10, 20, and 50. 51
Figure 4.1.30 The relative bias in the estimated treatment effect, as a function of p and different combinations of k and n across different effect sizes, with a mixture of normals data using the standardized mean difference for a fixed total sample size of 200..... 51
Figure 4.1.31 The coverage probability in the estimated treatment effect, as a function of p and n across different effect sizes, with a mixture of normals data using the standardized mean difference for a fixed value of k at all levels of 2, 5, 10, and 20..... 52
Figure 4.1.32 The coverage probability in the estimated treatment effect, as a function of p and k across different effect sizes, with a mixture of normals data using the standardized mean difference for a fixed value of n at all levels of 5, 10, 20, and 50..... 53
Figure 4.1.33 The coverage probability in the estimated treatment effect, as a function of p and different combinations of k and n across different effect sizes, with a mixture of normals data using the standardized mean difference for a fixed total sample size of 200. 53
Figure 4.1.34 The power in the test, as a function of p and n across different effect sizes, with a mixture of normals data using the standardized mean difference for a fixed value of k at all levels of 2, 5, 10, and 20. 54
Figure 4.1.35 The power in the test, as a function of p and k across different effect sizes, with a mixture of normals data using the standardized mean difference for a fixed value of n at all levels of 5, 10, 20, and 50. 55
Figure 4.1.36 The power in the test, as a function of p and different combinations of k and n across different effect sizes, with a mixture of normals data using the standardized mean difference for a fixed total sample size of 200. 55

Figure 4.2.1 The relative bias in the estimated treatment effect, as a function of σ and n across different effect sizes, with adding a constant to lognormal data using the mean difference for a fixed value of k at all levels of 2, 5, 10, and 20. 57
Figure 4.2.2 The relative bias in the estimated treatment effect, as a function of σ and k across different effect sizes, with adding a constant to lognormal data using mean difference for a fixed value of n at all levels of 5, 10, 20, and 50. 58
Figure 4.2.3 The relative bias in the estimated treatment effect, as a function of σ and different combinations of k and n across different effect sizes, with adding a constant to lognormal data using the mean difference for a fixed total sample size of 200. 58
Figure 4.2.4 The coverage probability in the estimated treatment effect, as a function of σ and n across different effect sizes, with adding a constant to lognormal data using the mean difference for a fixed value of k at all levels of 2, 5, 10, and 20. 59
Figure 4.2.5 The coverage probability in the estimated treatment effect, as a function of σ and k across different effect sizes, with adding a constant to lognormal data using the mean difference for a fixed value of n at all levels of 5, 10, 20, and 50. 60
Figure 4.2.6 The coverage probability in the estimated treatment effect, as a function of σ and different combinations of k and n across different effect sizes, with adding a constant to lognormal data using the mean difference for a fixed total sample size of 200. 60
Figure 4.2.7 The power in the test, as a function of σ and n across different effect sizes, with adding a constant to lognormal data using the mean difference for a fixed value of k at all levels of 2, 5, 10, and 20. 62
Figure 4.2.8 The power in the estimated treatment effect, as a function of σ and k across different effect sizes, with adding a constant to lognormal data using the mean difference for a fixed value of n at all levels of 5, 10, 20, and 50. 63
Figure 4.2.9 The power in the test, as a function of σ and different combinations of k and n across different effect sizes, with adding a constant to lognormal data using the mean difference for a fixed total sample size of 200. 63
Figure 4.2.10 The relative bias in the estimated treatment effect, as a function of σ and n across different effect sizes, with simulating two lognormal distributions with different

parameters using the mean difference for a fixed value of k at all levels of 2, 5, 10, and 20..... 65
Figure 4.2.11 The relative bias in the estimated treatment effect, as a function of σ and k across different effect sizes, with simulating two lognormal distributions with different parameters using the mean difference for a fixed value of n at all levels of 5, 10, 20, and 50..... 66
Figure 4.2.12 The relative bias in the estimated treatment effect, as a function of σ and different combinations of k and n across different effect sizes, with simulating two lognormal distributions with different parameters using the mean difference for a fixed total sample size of 200..... 66
Figure 4.2.13 The coverage probability in the estimated treatment effect, as a function of σ and n across different effect sizes, with simulating two lognormal distributions with different parameters using the mean difference for a fixed value of k at all levels of 2, 5, 10, and 20..... 68
Figure 4.2.14 The coverage probabilities in the estimated treatment effect, as a function of σ and k across different effect sizes, with simulating two lognormal distributions with different parameters using the mean difference for a fixed value of n at all levels of 5, 10, 20, and 50..... 69
Figure 4.2.15 The coverage probability in the estimated treatment effect, as a function of σ and different combinations of k and n across different effect sizes, with simulating two lognormal distributions with different parameters using the mean difference for a fixed total sample size of 200..... 69
Figure 4.2.16 The power in the test, as a function of σ and n across different effect sizes, with simulating two lognormal distributions with different parameters using the mean difference for a fixed value of k at all levels of 2, 5, 10, and 20. 71
Figure 4.2.17 The power in the test, as a function of σ and k across different effect sizes, with simulating two lognormal distributions with different parameters using the mean difference for a fixed value of n at all levels of 5, 10, 20, and 50. 72

Figure 4.2.18 The power in the test, as a function of σ and different combinations of k and n across different effect sizes, with simulating two lognormal distributions with different parameters using the mean difference for a fixed total sample size of 200. 72
Figure 4.2.19 The relative bias in the estimated treatment effect, as a function of σ and n across different effect sizes, with adding a constant to lognormal data using the standardized mean difference for a fixed value of k at all levels of 2, 5, 10, and 20..... 74
Figure 4.2.20 The relative bias in the estimated treatment effect, as a function of σ and k across different effect sizes, with adding a constant to lognormal data using the standardized mean difference for a fixed value of n at all levels of 5, 10, 20, and 50..... 75
Figure 4.2.21 The relative bias in the estimated treatment effect, as a function of σ and different combinations of k and n across different effect sizes, with adding a constant to lognormal data using the standardized mean difference for a fixed total sample size of 200..... 75
Figure 4.2.22 The coverage probability in the estimated treatment effect, as a function of σ and n across different effect sizes, with adding a constant to lognormal data using the standardized mean difference for a fixed value of k at all levels of 2, 5, 10, and 20..... 77
Figure 4.2.23 The coverage probability in the estimated treatment effect, as a function of σ and k across different effect sizes, with adding a constant to lognormal data using the standardized mean difference for a fixed value of n at all levels of 5, 10, 20, and 50..... 78
Figure 4.2.24 The coverage probability in the estimated treatment effect, as a function of σ and different combinations of k and n across different effect sizes, with adding a constant to lognormal data using the standardized mean difference for a fixed total sample size of 200..... 78
Figure 4.2.25 The power in the test, as a function of σ and n across different effect sizes, with adding a constant to lognormal data using the standardized mean difference for a fixed value of k at all levels of 2, 5, 10, and 20. 79
Figure 4.2.26 The power in the test, as a function of σ and k across different effect sizes, with adding a constant to lognormal data using the standardized mean difference for a fixed value of n at all levels of 5, 10, 20, and 50. 80

Figure 4.2.27 The power in the test, as a function of σ and different combinations of k and n across different effect sizes, with adding a constant to lognormal data using the standardized mean difference for a fixed total sample size of 200. 80
Figure 4.2.28 The relative bias in the estimated treatment effect, as a function of σ and n across different effect sizes, with simulating two lognormal distributions with different parameters using the standardized mean difference for a fixed value of k at all levels of 2, 5, 10, and 20. 82
Figure 4.2.29 The relative bias in the estimated treatment effect, as a function of σ and k across different effect sizes, with simulating two lognormal distributions with different parameters using the standardized mean difference for a fixed value of n at all levels of 5, 10, 20, and 50. 83
Figure 4.2.30 The relative bias in the estimated treatment effect, as a function of σ and different combinations of k and n across different effect sizes, with simulating two lognormal distributions with different parameters using the standardized mean difference for a fixed value of n at all levels of 200. 83
Figure 4.2.31 The coverage probability in the estimated treatment effect, as a function of σ and n across different effect sizes, with simulating two lognormal distributions with different parameters using the standardized mean difference for a fixed value of k at all levels of 2, 5, 10, and 20. 85
Figure 4.2.32 The coverage probability in the estimated treatment effect, as a function of σ and k across different effect sizes, with simulating two lognormal distributions with different parameters using the standardized mean difference for a fixed value of n at all levels of 5, 10, 20, and 50. 86
Figure 4.2.33 The coverage probability in the estimated treatment effect, as a function of σ and different combinations of k and n across different effect sizes, with simulating two lognormal distributions with different parameters using the standardized mean difference for a fixed total sample size of 200. 86

Figure 4.2.34 The power in the test, as a function of σ and n across different effect sizes, with simulating two lognormal distributions with different parameters using the standardized mean difference for a fixed value of k at all levels. of 2, 5, 10, and 20..... 87
Figure 4.2.35 The power in the test, as a function of σ and k across different effect sizes, with simulating two lognormal distributions with different parameters using the standardized mean difference for a fixed value of n at all levels of 5, 10, 20, and 50..... 88
Figure 4.2.36 The power in the test, as a function of σ and different combinations of k and n across different effect sizes, with simulating two lognormal distributions with different parameters using the standardized mean difference for a fixed total sample size of 200. 88
Figure 4.3.1 The relative bias in the estimated treatment effect, as a function of α and n across different effect sizes, with adding a constant to gamma data using the mean difference for a fixed value of k at all levels of 2, 5, 10, and 20. 90
Figure 4.3.2 The relative bias in the estimated treatment effect, as a function of α and k across different effect sizes, with adding a constant to gamma data using the mean difference for a fixed value of n at all levels of 5, 10, 20, and 50. 91
Figure 4.3.3 The relative bias in the estimated treatment effect, as a function of α and different combinations of k and n across different effect sizes, with adding a constant to gamma data using the mean difference for a fixed total sample size of 200. 91
Figure 4.3.4 The coverage probability in the estimated treatment effect, as a function of α and n across different effect sizes, with adding a constant to gamma data using the mean difference for a fixed value of k at all levels of 2, 5, 10, and 20. 93
Figure 4.3.5 The coverage probability in the estimated treatment effect, as a function of α and k across different effect sizes, with adding a constant to gamma data using the mean difference for a fixed value of n at all levels of 5, 10, 20, and 50. 94
Figure 4.3.6 The coverage probability in the estimated treatment effect, as a function of α and different combinations of k and n across different effect sizes, with adding a constant to gamma data using the mean difference for a fixed value of k at all levels of 200. 94

Figure 4.3.7 The power in the test, as a function of α and n across different effect sizes, with adding a constant to gamma data using the mean difference for a fixed value of k at all levels of 2, 5, 10, and 20. 95
Figure 4.3.8 The power in the test, as a function of α and k across different effect sizes, with adding a constant to gamma data using the mean difference for a fixed value of n at all levels of 5, 10, 20, and 50. 96
Figure 4.3.9 The power in the test, as a function of α and different combinations of k and n across different effect sizes, with adding a constant to gamma data using the mean difference for a fixed total sample size of 200. 96
Figure 4.3.10 The relative bias in the estimated treatment effect, as a function of α and n across different effect sizes, with simulating two gamma distributions with different parameters using the mean difference for a fixed value of k at all levels of 2, 5, 10, and 20..... 98
Figure 4.3.11 The relative bias in the estimated treatment effect, as a function of α and k across different effect sizes, with simulating two gamma distributions with different parameters using the mean difference for a fixed value of n at all levels of 5, 10, 20, and 50..... 99
Figure 4.3.12 The relative bias in the estimated treatment effect, as a function of α and different combinations of k and n across different effect sizes, with simulating two gamma distributions with different parameters using the mean difference for a fixed total sample size of 200..... 99
Figure 4.3.13 The coverage probability in the estimated treatment effect, as a function of α and n across different effect sizes, with simulating two gamma distributions with different parameters using the mean difference for a fixed value of k at all levels of 2, 5, 10, and 20. 101
Figure 4.3.14 The coverage probability in the estimated treatment effect, as a function of α and k across different effect sizes, with simulating two gamma distributions with different parameters using the mean difference for a fixed value of n at all levels of 5, 10, 20, and 50. 102

Figure 4.3.15 The coverage probability in the estimated treatment effect, as a function of α and different combinations of k and n across different effect sizes, with simulating two gamma distributions with different parameters using the mean difference for a fixed total sample size of 200. 102
Figure 4.3.16 The power in the test, as a function of α and n across different effect sizes, with simulating two gamma distributions with different parameters using the mean difference for a fixed value of k at all levels of 2, 5, 10, and 20. 103
Figure 4.3.17 The power in the test, as a function of α and k across different effect sizes, with simulating two gamma distributions with different parameters using the mean difference for a fixed value of n at all levels of 5, 10, 20, and 50. 104
Figure 4.3.18 The power in the test, as a function of α and different combinations of k and n across different effect sizes, with simulating two gamma distributions with different parameters using the mean difference for a fixed total sample size of 200. 104
Figure 4.3.19 The relative bias in the estimated treatment effect, as a function of α and n across different effect sizes, with adding a constant to gamma data using the standardized mean difference for a fixed value of k at all levels of 2, 5, 10, and 20. 106
Figure 4.3.20 The relative bias in the estimated treatment effect, as a function of α and k across different effect sizes, with adding a constant to gamma data using the standardized mean difference for a fixed value of n at all levels of 5, 10, 20, and 50. 107
Figure 4.3.21 The relative bias in the estimated treatment effect, as a function of α and different combinations of k and n across different effect sizes, with adding a constant to gamma data using the standardized mean difference for a fixed total sample size of 200. 107
Figure 4.3.22 The coverage probability in the estimated treatment effect, as a function of α and n across different effect sizes, with adding a constant to gamma data using the standardized mean difference for a fixed value of k at all levels of 2, 5, 10, and 20..... 109
Figure 4.3.23 The coverage probability in the estimated treatment effect, as a function of α and k across different effect sizes, with adding a constant to gamma data using the standardized mean difference for a fixed value of n at all levels of 5, 10, 20, and 50.... 110

Figure 4.3.24 The coverage probability in the estimated treatment effect, as a function of α and different combinations of k and n across different effect sizes, with adding a constant to gamma data using the standardized mean difference for a fixed total sample size of 200. 110
Figure 4.3.25 The power in the test, as a function of α and n across different effect sizes, with adding a constant to gamma data using the standardized mean difference for a fixed value of k at all levels of 2, 5, 10, and 20. 111
Figure 4.3.26 The power in the test, as a function of α and k across different effect sizes, with adding a constant to gamma data using the standardized mean difference for a fixed value of n at all levels of 5, 10, 20, and 50. 112
Figure 4.3.27 The power in the test, as a function of α and different combinations of k and n across different effect sizes, with adding a constant to gamma data using the standardized mean difference for a fixed total sample size of 200. 112
Figure 4.3.28 The relative bias in the estimated treatment effect, as a function of α and n across different effect sizes, with simulating two gamma distributions with different parameters using the standardized mean difference for a fixed value of k at all levels of 2, 5, 10, and 20. 114
Figure 4.3.29 The relative bias in the estimated treatment effect, as a function of α and k across different effect sizes, with simulating two gamma distributions with different parameters using the standardized mean difference for a fixed value of n at all levels of 5, 10, 20, and 50. 115
Figure 4.3.30 The relative bias in the estimated treatment effect, as a function of α and different combinations of k and n across different effect sizes, with simulating two gamma distributions with different parameters using the standardized mean difference for a fixed total sample size of 200. 115
Figure 4.3.31 The coverage probability in the estimated treatment effect, as a function of α and n across different effect sizes, with two gamma distributions with different parameters using the standardized mean difference for a fixed value of k at all levels of 2, 5, 10, and 20. 117

Figure 4.3.32 The coverage probability in the estimated treatment effect, as a function of α and k across different effect sizes, with simulating two gamma distributions with different parameters using the standardized mean difference for a fixed value of n at all levels of 5, 10, 20, and 50. 118
Figure 4.3.33 The coverage probability in the estimated treatment effect, as a function of α and different combinations of k and n across different effect sizes, with simulating two gamma distributions with different parameters using the standardized mean difference for a fixed total sample size of 200..... 118
Figure 4.3.34 The power in the test, as a function of α and n across different effect sizes, with simulating two gamma distributions with different parameters using the standardized mean difference for a fixed value of k at all levels of 2, 5, 10, and 20..... 119
Figure 4.3.35 The power in the test, as a function of α and k across different effect sizes, with simulating two gamma distributions with different parameters using the standardized mean difference for a fixed value of n at all levels of 5, 10, 20, and 50..... 120
Figure 4.3.36 The power in the test, as a function of α and different combinations of k and n across different effect sizes, with simulating two gamma distributions with different parameters using the standardized mean difference for a fixed total sample size of 200. 120
Figure 4.4.1 The relative bias in the estimated treatment effect, as a function of $d.f.$ and n across different effect sizes, with adding a constant to $\chi 2$ data using the mean difference for a fixed value of k at all levels of 2, 5, 10, and 20. 122
Figure 4.4.2 The relative bias in the estimated treatment effect, as a function of $d.f.$ and k across different effect sizes, with adding a constant to $\chi 2$ data using the mean difference for a fixed value of n at all levels. 123
Figure 4.4.3 The relative bias in the estimated treatment effect, as a function of $d.f.$ and different combinations of k and n across different effect sizes, with adding a constant to $\chi 2$ data using the mean difference for a fixed total sample size of 200. 123

Figure 4.4.4 The coverage probability in the estimated treatment effect, as a function of $d.f.$ and n across different effect sizes, with adding a constant to $\chi 2$ data using the mean difference for a fixed value of k at all levels of 2, 5, 10, and 20. 125
Figure 4.4.5 The coverage probability in the estimated treatment effect, as a function of $d.f.$ and k across different effect sizes, with adding a constant to $\chi 2$ data using the mean difference for a fixed value of n at all levels of 5, 10, 20, and 50. 126
Figure 4.4.6 The coverage probability in the estimated treatment effect, as a function of $d.f.$ and different combinations of k and n across different effect sizes, with adding a constant to $\chi 2$ data using the mean difference for a fixed total sample size of 200. 126
Figure 4.4.7 The power in the test, as a function of $d.f.$ and n across different effect sizes, with adding a constant to $\chi 2$ data using the mean difference for a fixed value of k at all levels of 2, 5, 10, and 20. 127
Figure 4.4.8 The power in the test, as a function of $d.f.$ and k across different effect sizes, with adding a constant to $\chi 2$ data using the mean difference for a fixed value of n at all levels of 5, 10, 20, and 50. 128
Figure 4.4.9 The power in the test, as a function of $d.f.$ and different combinations of k and n across different effect sizes, with adding a constant to $\chi 2$ data using the mean difference for a fixed total sample size of 200. 128
Figure 4.4.10 The relative bias in the estimated treatment effect, as a function of $d.f.$ and n across different effect sizes, with adding a constant to $\chi 2$ data using the standardized mean difference for a fixed value of k at all levels of 2, 5, 10, and 20. 130
Figure 4.4.11 The relative bias in the estimated treatment effect, as a function of $d.f.$ and k across different effect sizes, with adding a constant to $\chi 2$ data using the standardized mean difference for a fixed value of n at all levels of 5, 10, 20, and 50. 131
Figure 4.4.12 The relative in the estimated treatment effect, as a function of $d.f.$ and different combinations of k and n across different effect sizes, with adding a constant to $\chi 2$ data using the standardized mean difference for a fixed total sample size of 200. ... 131

Figure 4.4.13 The coverage probability in the estimated treatment effect, as a function of $d.f.$ and n across different effect sizes, with adding a constant to $\chi 2$ data using the standardized mean difference for a fixed value of k at all levels of 2, 5, 10, and 20..... 133
Figure 4.4.14 The coverage probability in the estimated treatment effect, as a function of $d.f.$ and k across different effect sizes, with adding a constant to $\chi 2$ data using the standardized mean difference for a fixed value of n at all levels of 5, 10, 20, and 50.... 134
Figure 4.4.15 The coverage probability in the estimated treatment effect, as a function of $d.f.$ and different combinations of k and n across different effect sizes, with adding a constant to $\chi 2$ data using the standardized mean difference for a fixed total sample size of 200..... 134
Figure 4.4.16 The power in the test, as a function of $d.f.$ and n across different effect sizes, with adding a constant to $\chi 2$ data using the standardized mean difference for a fixed value of k at all levels of 2, 5, 10, and 20. 135
Figure 4.4.17 The power in the test, as a function of $d.f.$ and k across different effect sizes, with adding a constant to $\chi 2$ data using the standardized mean difference for a fixed value of n at all levels of 5, 10, 20, and 50. 136
Figure 4.4.18 The power in the test, as a function of $d.f.$ and different combinations of k and n across different effect sizes, with adding a constant to $\chi 2$ data using the standardized mean difference for a fixed total sample size of 200. 136

List of Tables

Table 3.1 value of skewness and kurtosis with give σ for lognormal distribution 18
Table 3.2 values of μ for different σ, d values 19
Table 3.3 the value for α with given skewness 21
Table 3.4 values for βt with different effect sizes and skewness 22
Table A.1 a summary of all the algebra for different cases within the same distribution of the outcome variables in terms of the skewness, parameters, and corresponding true effect for mean difference and the standardized mean difference 157
Table A.2 the type I error rate for the mixture of normal distributions of the outcome variable with the different number of studies, and the number of individuals per group when using the mean difference and the standardized mean difference as the effect sizes 159
Table A.3 the type I error rate for the lognormal, gamma, and $\chi 2$ distributions of the outcome variable with different skewness, the number of studies, and the number of individuals per group when using the mean difference as the effect sizes 162
Table A.4 the type I error rate for the lognormal, gamma, and $\chi 2$distributions of the outcome variable with different skewness, the number of studies, and the number of individuals per group when using the standardized mean difference as the effect sizes 165

Chapter 1 Introduction

Meta-analysis is a statistical method for synthesizing data from various independent studies (Higgins, et al., 2024). According to Borenstein, Hedges, Higgins, and Rothstein (2009), meta-analysis aims to identify the consistency of effect sizes, accurately estimate effect sizes, and measure the extent of variance across studies included in the synthesis. They also noted that, compared to other synthesis methods, such as narrative reviews, meta-analysis enhances transparency by systematically weighting studies and can manage a large number of studies. Meta-analysis has been employed in many fields, including medicine, psychology, and ecology (Weeks, George, Maclure, & Stewart, 2016; Roberts et al., 2019; Plieninger, Hui, Gaertner, & Huntsinger, 2014). In the past 10 years, there have been 174843 results found from searching Ovid MEDLINE(R) with the search terms (“meta-analys*.mp.” or “meta analys*.mp.”) and “systematic review/”, indicating a great interest in conducting meta-analyses.

The fixed- and the random-effects models are the most widely used meta-analysis methods. The fixed-effect model assumes all studies share the same true effect size and that the observed effect sizes across individual studies differ solely due to within-study variation, and the random-effects model accounts for both within- and between-study variation (Borenstein, Hedges, Higgins, & Rothstein, 2009). These meta-analysis methods with continuous outcome variables have been developed based on the assumption that the variation of the outcome variable between patients within treatment groups for each study follows a normal distribution (Jackson & White, 2018; Higgins,

White, & Anzures-Cabrera, 2008). However, there are many situations where the outcome data are non-normal, especially in the medical field, for example, blood biomarker levels (C-reactive protein), scale questionnaires, and BMI (body mass index). According to Higgins, White, and Anzures-Cabrera (2008), the Central Limit Theorem ensures that the meta-analytical result remains valid with a sufficiently large sample size per study with a skewed outcome variable. Unfortunately, many meta-analyses often include small studies or a small number of studies. For example, a meta-analysis about therapeutic options for rare rheumatic diseases included a study with only 22 patients (Bender, et al., 2020). Moreover, one study shows that the median number of studies included in a meta-analysis among 22435 eligible meta-analyses is 3 (Davey, Turner, Clarke, & Higgins, 2011). Therefore, identifying violations of the normality assumption and their impact on the meta-analysis is crucial for investigators to obtain robust results.

Two previous simulation studies have examined the impact of non-normality in primary studies on the performance of meta-analysis, each with a distinct focus (Kontopantelis & Reeves, 2010; Sun & Cheung, 2020). Kontopantelis and Reeves concentrated on comparing the performance of different random-effects models, including DerSimonian & Laird, Biggerstaff & Tweedie, Sidik and Jonkman, Q-based, maximum-likelihood, profile-likelihood, and permutation methods. On the other hand, Sun and Cheung explored the influence of the standardized mean difference as an estimate of effect size using DerSimonian & Laird random-effects models. However, neither study paid much attention to the fixed-effect model when the distribution of the study effect deviated from normality. The details of these two studies will be presented in Chapter 2. Therefore, this

thesis offers insights into the robustness of meta-analysis methods for fixed-effect models with continuous outcome variables.

The study's primary objective is to evaluate the robustness of the meta-analysis for a fixed-effect model when the normality assumption is violated. I explore how various non-normal distributions of the outcome variable influence the estimation and inference of the meta-analysis through a simulation study. Additionally, I investigate the impacts of the magnitude of the effect size within each outcome variable distribution, the number of studies, and the number of patients per group per study on the meta-analysis results. By examining different conditions, we aim to inform investigators about when they should take extra caution while interpreting the meta-analysis results so that they can make policies and guidelines based on reliable evidence and reduce the risk of implementing ineffective and harmful policies and guidelines.

The paper consists of five chapters. Chapter 1 explores the general concept of meta-analysis and the issue of the normality assumption. Chapter 2 offers a concise introduction to calculating a mean difference and a standardized mean difference, along with the steps of meta-analysis within a fixed-effect model. Furthermore, it presents two relevant previous simulation studies and discusses their limitations. Chapter 3 details the simulation design, the selection of distributions for the outcome variable, and the evaluation criteria. Chapter 4 showcases the simulation results based on different distributions of the outcome variables. Finally, Chapter 5 summarizes the findings, addresses the limitations of the simulation, and considers future directions for examining the robustness of the meta-analysis.

Chapter 2 Literature Review

In this chapter, I first introduce two different effect size measures that researchers use for continuous outcomes. Then, I present the standard procedure of meta-analysis for a fixed-effect model. Lastly, I discuss the issues raised by the non-normality of the data and describe two previous simulation studies in detail and how they considered non-normality, particularly the differences in the deviations from normality from this thesis.

2.1 Mean Difference and Standardized Mean Difference

Borenstein (2009) introduces the effect size as "a number that reflects the magnitude of the relationship between two variables." In hypothesis testing, the effect size can be treated as a parameter that takes 0 when the null hypothesis is true and any other values to imply the degree of departure from the null hypothesis when false (Cohen, 1988). Also, Cohen (1988, p11) stated that the effect size needs to be quantified and assessed using a specific unit suitable for the statistical analysis. The mean difference (MD) and the standardized mean difference (SMD) are two common effect sizes that measure continuous outcomes in the meta-analysis. The details of these two effect size measures are provided in the following sections.

2.1.1 Mean Difference (MD)

The MD is used when all studies report the continuous outcome using the same measurement scales and units, and these units are meaningful in practice (Takeshima, et

al., 2014). Moreover, MD is preferred because of its interpretability if the absolute magnitude of the difference between groups is the main interest.

According to Borenstein (2009), for each study, the population mean difference is defined as:

$$\Delta = \mu_1 - \mu_{2,} \tag{1}$$

where μ_1 and μ_2 are the population means of the two independent groups. From statistical algorithms in Review Manager (2022), to estimate Δ, we have:

$$MD = m_1 - m_2, \tag{2}$$

with the standard error:

$$SE(MD) = \sqrt{\frac{sd_1^2}{n_1} + \frac{sd_2^2}{n_2}}. \tag{3}$$

m_1 and m_2 is the sample mean of the two groups, sd_1 and sd_2 are the sample standard deviations, and n_1 and n_2 are the sample size of the two groups.

2.1.2 Standardized Mean Difference (SMD)

In clinical trials, it is common that different instruments can measure essentially the same outcome; for example, the most widely used assessments for the quality of life are MOS SF-36 (Medical Outcomes Study Short-Form 36), Euro EQ-5D, SF-12 (12-Item Short-Form Health Survey), and Visual Analogue Scale EQ-VAS (Pequeno, De Araújo Cabral, Marchioni, Lima, & De Oliveira Lyra, 2020). Therefore, the SMD is introduced in

situations where different studies use different measurement instruments. The difference can be referred to as a different scale and direction. The SMD can also indicate the magnitude of the effect when the scale is unfamiliar (Borenstein, 2009), since Cohen (1988) suggests an SMD of 0.2 as a small effect, an SMD of 0.5 as a moderate effect, and an SMD of 0.8 as a large effect.

2.1.2.1 Cohen's d

According to Borenstein (2009), the population standardized mean difference is defined as:

$$\delta = \frac{\mu_1 - \mu_2}{\sigma}, \tag{4}$$

where $\sigma_1 = \sigma_2 = \sigma$, and σ_1 and σ_2 are the population standard deviations of the two groups. Therefore, Cohen's d is estimated by:

$$d = \frac{m_1 - m_2}{S_{pool}}, \tag{5}$$

where S_{pool} is the sample pooled standard deviation,

$$S_{pool} = \sqrt{\frac{(n_1 - 1)sd_1^2 + (n_2 - 1)sd_2^2}{n_1 + n_2 - 2}}. \tag{6}$$

Also, the standard error of d is given by:

$$SE(d) = \sqrt{\frac{n_1 + n_2}{n_1 n_2} + \frac{d^2}{2(n_1 + n_2)}}. \tag{7}$$

2.1.2.2 Hedge's g

Hedge's g is introduced because the estimate of Cohen's d gives a slight bias when dealing with small samples, leading to an overestimation. From statistical algorithms in Review Manager (2022), Hedge's g is given by:

$$g = \frac{m_1 - m_2}{S_{pool}}\left(1 - \frac{3}{4(n_1 + n_2) - 9}\right), \tag{8}$$

with the standard error:

$$SE(g) = \sqrt{\frac{n_1 + n_2}{n_1 n_2} + \frac{g^2}{2(n_1 + n_2 - 3.94)}}. \tag{9}$$

For both Cohen's d and Hedge's g, they assume the underlying population standard deviations of the two groups are the same. Moreover, this thesis uses Hedge's g to estimate the SMD for the later simulation.

However, Hopkins and Rowlands (2024), noted that different SDs leads to different estimation of the effect size. They introduced various pooled SDs (standard deviation) including, post-only SDs (the outcome is measured only once after the intervention), pre SDs and pre-post SDs (the outcome is measured before and after the intervention).

2.2 Meta-Analysis

As mentioned in Chapter 1, meta-analysis is a statistical synthesis method that pools results from various independent studies (Higgins, et al., 2024). Two major meta-analysis models are introduced: the fixed-effect model and the random-effects model. These

models are based on different assumptions. The fixed-effect model assumes that all studies share the same true effect size, considering only the within-study variation. In contrast, the random-effects model assumes that the true effect size varies across all studies, incorporating both within-study and between-study variation (Borenstein, Hedges, Higgins, & Rothstein, 2009). Since this thesis exclusively considers the fixed-effect model, I will only present the mathematics of the fixed-effect model in the following section.

2.2.1 Fixed-Effect Model

In the fixed-effect model, let $\hat{Y}_i$ be the observed effect for the i^{th} study, then we have:

$$\hat{Y}_i = \theta + \epsilon_i, \tag{10}$$

where θ is the true effect, and ϵ_i is the sampling error for the i^{th} study. ϵ_i is assumed to follow a normal distribution, that $\epsilon_i \sim N(0, \sigma_i^2)$, where σ_i^2 is the within-study variance. Therefore, we have $\hat{Y}_i|\theta \sim N(\theta, \sigma_i^2)$ (Jackson & White, 2018; Hedges & Vevea, 1998).

According to statistical algorithms in Review Manager (2022), the estimated overall effect is given by:

$$\hat{Y} = \frac{\sum_{i=1}^{k} w_i \hat{Y}_i}{\sum_{i=1}^{k} w_i}, \tag{11}$$

where w_i is the estimated weight for the i^{th} study, giving:

$$w_i = \frac{1}{\left(SE\{\hat{Y}_i\}\right)^2}, \tag{12}$$

with

$$SE(\hat{Y}) = \frac{1}{\sqrt{\sum_{i=1}^{k} w_i}}. \tag{13}$$

Then, we have the $100(1-\alpha)\%$ confidence interval for $\hat{Y}$ as follows:

$$\hat{Y} \pm z_{\alpha/2} * SE(\hat{Y}) \tag{14}$$

where α is the significance level and is usually set to 0.05. $z_{\alpha/2}$ is the $(1-\alpha/2)^{th}$ quantile of the standard normal distribution. Moreover, to test whether there is an overall effect under the null hypothesis that there is no overall effect of the intervention, we have the test statistic given by:

$$Z = \frac{\hat{Y}}{SE(\hat{Y})}. \tag{15}$$

2.3 Non-normality

As mentioned, the fixed-effect model assumes that the between-patient study effects are normally distributed around a common true effect with only within-study variance and no systematic variation in the treatment effect among studies. However, real medical, social science, and education data are not always normal. For instance, Micceri (1989) evaluated different outcome measurements in psychology and found all of them

statistically significantly non-normal, with a 0.01 significance level with the Kolmogorov-Smirnov test of normality. Moreover, according to Blanca, Arnau, López-Montiel, Bono, & Bendayan (2013), among 693 measures used in psychological and educational studies, 74.4% of them are slightly or moderately non-normal, with both skewness and kurtosis ranging from 0.26 to 1.25. In addition, from a systematic review by Bono, Blanca, Arnau, & Gómez-Benito (2017), the most common non-normal distributions after screening 262 studies in the fields of health, education and social science are gamma, negative binomial, multinomial, binomial, lognormal, and exponential. In the following subsections, I introduce the possible sources of non-normality of the data and the possible problems caused by the violation of non-normality in meta-analysis provided with two simulation studies that tackle the similar issue.

2.3.1 Causes of Non-normality

The first possible reason for data to be non-normal is the nature of the outcome measurements, including the design of the outcome measures and the ceiling or floor effect. For example, a 7-point Global Overall Symptom scale results in discrete data collection following a multinomial distribution (Micceri, 1989). Moreover, outcomes like C-reactive protein (CRP) levels, blood pressure, and cholesterol levels have a lower bound (greater than 0) but can also have extremely high values, leading to a right-skewed distribution. A study examined the CRP levels of 2275 males and 3832 females and found that the distribution of the CRP levels is highly skewed to the right, given that most CRP values are less than 2mg/liter (Yamada, et al., 2001). Another reason might be that a subgroup within a treatment group responds differently to the same intervention, leading

to a bimodal distribution. For example, different genetic features might influence the response of antihypertensive treatment (Mellen & Herrington, 2005).

2.3.2 Problems of Non-normality

First, Wilcox (2005) mentioned that the standard error estimation is not robust to non-normality, followed by an example of heavy-tailed distributions giving an overestimation of the standard error, leading to a wider confidence interval of the mean difference. Also, existing studies have shown that the estimation of the SMD and its standard error is not robust to non-normality in a single study (Kelley, 2005; Wilcox & Keselman, 2003). Furthermore, they found that the coverage probability departs from the nominal level. Therefore, when Sun and Cheung (2020) examined the effect of non-normality in primary studies using the SMD in the meta-analysis, they summarized that the non-robustness of the pooled effect size could result from the accumulation of the non-robustness of the SMD estimator with non-normal data in the primary studies. Moreover, the non-robustness of the pooled effect size may arise from incorrectly estimated pooled sample standard errors, which in turn leads to biased study weights. In addition to non-normality, Hopkins and Rowlands (2024) observed that most meta-analyses used either inappropriate SDs or applied inconsistent SDs across studies, which introduces bias in estimating the effect size even for normally distributed data.

A previous simulation by Sun and Cheung (2020) assessed the meta-analysis's robustness using SMD with the DerSiMonian-Laird (DL) random-effects models when the data from the primary studies deviate from normal. They evaluated six combinations of

distributions of two groups chosen from normal, exponential, and reversed exponential distributions. In addition, they chose 6 effect sizes, 5 standard deviation ratios, 4 sample sizes per study, 5 study sizes, and 7 heterogeneity variances, resulting in 4200 conditions for each distribution combination, and each condition had 5000 replicates. They included the bias, MSE, and coverage probability for outcome measures. They found that non-normality increases the bias and MSE, especially when the distributions of the two groups are oppositely skewed. In addition, the coverage probability deviated more from the nominal levels with non-normal data, particularly with oppositely skewed data. However, in this simulation, the authors did not cover other common distributions that might occur in a study, such as gamma and lognormal distributions. Moreover, the cases where the two groups are oppositely skewed are rare in the parallel intervention study (e.g. randomized control trials). Another limitation of this study is that it did not investigate the influence of the magnitude of the deviation from normality on the result.

Another simulation by Kontopantelis and Reeves (2010) assessed the performance of the meta-analysis against the normality assumption and simulated the study's effect size using various distributions instead of the primary data. Also, this study has a different focus from Sun and Cheung's study (2020), where Kontopantelis and Reeves (2010) compared the performance of the fixed-effect model and seven other random-effects models through the coverage probability, the power probability, and overall effect estimation. They generated 25 different unimodal distributions through different combinations of skewness and kurtosis, three different bimodal distributions of equal probability ($p = 0.5$), and three extreme distributions such as uniform, 'U shaped ' beta,

and 'double spike'. The number of studies ranges from 2 to 35, each with 10000 replications. They did not include the results of cases where the effect size distributions are non-normal with no between-study variation. However, with a slight between-study variation of H^2 being 1.18, the coverage probability for the fixed-effect model is below 0.92 for effect sizes of all distributions, even when the number of studies exceeds 26, where H^2 is defined as the heterogeneity measure least affected by the number of studies; a value of 1 indicates homogeneity and a value greater than 1 indicates heterogeneity. The major difference between Kontopantelis and Reeves' simulation and this simulation is that they focus more on violating the normality assumption on the between-study variation by randomly simulating the effect size for each study by a non-normal distribution. In addition, the simulation design cannot assess the impact of the different degrees of effect size on the performance of meta-analysis.

Furthermore, neither simulation study investigated the robustness of the fixed-effect models when the data from primary studies were non-normal. Also, these two studies did not discuss the individual effect of the number of studies and sample size per study, nor did they discuss more sample sizes or more studies when the total sample size is fixed. Therefore, this thesis aims to address these questions and evaluate how the extent of non-normality of each distribution of outcome variables impacts the bias, coverage probability, power and type I error. In this simulation, I will include the distributions commonly seen in health and social science studies, such as lognormal, gamma, and χ^2 distributions. The mixture of normal distributions is also included to mimic the situation where two groups follow different distributions and a subgroup of non-responders is

present in the intervention group. The next chapter illustrates the details of the simulation design.

Chapter 3 Method

3.1 Simulation

A Monte Carlo simulation is conducted by R to assess the robustness of the meta-analysis against non-normality of the outcome variable. The simulation study considers the impact of four factors for each distribution of the outcome variable: the number of studies, the number of patients per group per study, the skewness of the outcome variable and the effect sizes. We will vary the number of patients in each study group (assumed to be equal between the treatment group and control group), $n_{ic} = n_{it} = 5, 10, 20, 50$, where n_{ic} is the number of patients in the control group from the i^{th} study, and n_{it} is the number of patients in the intervention group from the i^{th} study. We will also set up different numbers of studies k = 2, 5, 10, 20 to be included in the meta-analysis. For each distribution, we will change the parameters to obtain the effect of skewness on the estimated effect. Also, we will compare the impact of different effect sizes on the results while keeping other parameters constant. The effect size will be defined to be 0.2 (small effect), 0.5 (moderate effect), and 0.8 (large effect) by Cohen's d, which is given by Equation 5.

Moreover, we will calculate the overall effect size in a meta-analysis by the mean difference and standardized mean difference using the generic inverse variance method for the fixed-effect model. We include the standardized mean difference to represent the

situation where the studies assess the same outcome using different measurement scales. Each meta-analysis is replicated 10000 times.

3.2 Distributions

We will assess the following distributions for the outcome variable a mixture of normals, log-normal, gamma, and χ^2.

3.2.1 The Mixture of Normal Distributions

The mixture of normal distributions represents the situation when a subgroup of patients responds differently to the treatment in the intervention group. The patient outcomes of the control group and the intervention group for the i^{th} study are generated via $N(\mu_1, \sigma)$ and $p * N(\mu_1, \sigma) + (1 - p) * N(\mu_2, \sigma)$ respectively, where μ_1 is the mean of the outcome for the patients in the control group and the subgroup of patients who do not respond to the treatment in the intervention group, μ_2 is the mean of the outcome for the patients in the intervention group who respond to the treatment. p is the proportion of the subgroup of patients which does not respond to the treatment in the intervention group and σ is the between-patient variance within those subgroups. Within a mixture of normal distribution, we will set up the following: first, different values of p will be investigated to show how the various magnitude of deviation from normality influences the results of the analysis, and p will be set as 0, 0.1, 0.2, 0.3, 0.4, 0.5, 0.6, 0.7, 0.8, 0.9, and 1, while other parameters will be kept constant. Second, the true overall effect will be considered as $\mu_2 - \mu_1$ and $(1 - p) * (\mu_2 - \mu_1)$ separately for mean difference and d and $(1 - p) * d$ for standardized mean difference, where $(1 - p) * (\mu_2 - \mu_1)$ is the overall difference of

the two groups. Meanwhile, $\mu_2 - \mu_1$ covers the difference between the two groups, but also focuses on the responders of the intervention group.

3.2.2 Lognormal Distribution

We compare the deviation from the normal distribution through skewness and kurtosis. From Johnson, Kotz, and Balakrishnan (1994), if X is followed by a lognormal (μ, σ^2), it has a PDF of

$$P_X(x) = \frac{1}{x\sigma\sqrt{(2\pi)}} exp\left(-\frac{(ln\, x - \mu)^2}{2\sigma^2}\right). \tag{16}$$

The expected value of X is

$$exp\left(\mu + \frac{\sigma^2}{2}\right). \tag{17}$$

and the variance of X is

$$(exp(\sigma^2) - 1)exp(2\mu + \sigma^2). \tag{18}$$

For a log-normal distribution, the skewness and kurtosis change according to the σ.

Skewness is

$$(exp(\sigma^2) + 2)\sqrt{exp(\sigma^2) - 1}, \tag{19}$$

and kurtosis is

$$exp(4\sigma^2) + 2exp(3\sigma^2) + 3exp(2\sigma^2) - 3. \tag{20}$$

Table 3.1 shows the corresponding skewness and kurtosis values for a given value of σ.

σ	skewness	kurtosis
0.1	0.30	3.16
0.2	0.61	3.68
0.3	0.95	4.64
0.4	1.32	6.26
0.5	1.75	8.90
0.6	2.26	13.27
0.7	2.89	20.79
0.8	3.69	34.37
0.9	4.75	60.41
1	6.18	113.94

Table 3.1 value of skewness and kurtosis with give σ for lognormal distribution

From Table 3.1, the shape is nearly symmetric when σ approaches 0, and skewness and kurtosis increase rapidly when $\sigma > 0.5$; thus, we will focus on the cases with $\sigma < 0.5$ when evaluating the results. In the simulation, we choose 0.1, 0.2, 0.3, 0.4, and 0.5 as the values for σ to indicate different magnitudes of deviation from a skewness of 0. The simulation for lognormal distribution will be set up in two ways. The first method simulates lognormal $(0, \sigma^2)$ for the control and intervention groups, then adds a constant c for the intervention group. The constant c is calculated below to match the corresponding Cohen's d values, and the true treatment effect is c, where c is

$$d * \sqrt{\{(\exp(\sigma^2) - 1)\exp(2\mu + \sigma^2)\}} \quad (21)$$

for mean difference and the true treatment effect is d for standardized mean difference.

The second method simulates lognormal $(0, \sigma^2)$ for the control and lognormal (μ, σ^2) for the intervention groups. μ is calculated as

$$\ln\left(\frac{1+\sqrt{1-(1-a)^2}}{1-a}\right), \tag{22}$$

where

$$a = \frac{d^2(exp(\sigma^2)-1)}{2}. \tag{23}$$

The values of μ for the corresponding values of σ and d can be found in Table 3.2.

Therefore, the true treatment effect for the second method is

$$exp\left(\frac{\sigma^2}{2}\right)(\exp(\mu)-1) \tag{24}$$

for mean difference and d for standardized mean difference.

σ	μ ($d = 0.2$)	μ ($d = 0.5$)	μ ($d = 0.8$)
0.1	0.020	0.050	0.080
0.2	0.040	0.101	0.163
0.3	0.061	0.154	0.249
0.4	0.083	0.210	0.341
0.5	0.107	0.271	0.444

Table 3.2 values of μ for different σ, d values

3.2.3 Gamma Distribution

Like lognormal distribution, we compare the deviation from the normal distribution through skewness and kurtosis. From Johnson, Kotz, and Balakrishnan (1994), if X is followed by a gamma (α, β), it has a PDF of

$$P_X(x) = \frac{1}{\Gamma(\alpha)\beta^{\alpha}} x^{\alpha-1} e^{\frac{-x}{\beta}}. \tag{25}$$

For the Gamma distribution, the skewness and kurtosis solely depend on the shape parameter α. Skewness and kurtosis are shown below:

$$Skewness = 2/\sqrt{\alpha}, \tag{26}$$

$$Kurtosis = 3 + 6/\sqrt{\alpha}, \tag{27}$$

which means if $\alpha \to \infty$, the skewness of gamma distribution is equivalent to the zero skewness of the normal distribution. α is calculated to have the same skewness as a lognormal distribution, to indicate different magnitudes of deviation from a normal distribution. The values of skewness and corresponding α are listed in Table 3.3. Like the simulation for the lognormal distribution, we will construct the simulation using the same

Skewness	α
0.30	43.93
0.61	10.60
0.95	4.44
1.32	2.29

1.75	1.31

Table 3.3 the value for α with given skewness

two methods: simulating gamma distribution with the same parameters for two groups, then adding a constant c to the intervention group, and simulating two gamma distributions with different parameters. The constant c is calculated as

$$d * \sqrt{(\alpha\beta^2)} \tag{28}$$

to match the required effect size values and c is also the true treatment effect when using the mean difference. β is the scale parameter for the gamma distribution and is set to 1 in both study groups for the first method. Also, the true treatment effect when using standardized mean difference is d. For the second method, the outcome variable for the control group will be simulated from $gamma(\alpha, \beta_c)$, where β_c is 1. The outcome variable for the intervention group will be simulated from $gamma(\alpha, \beta_t)$, where

$$\beta_t = \frac{\left(\alpha\beta_c + \beta_c\sqrt{\alpha d^2 - \frac{d^4}{4}}\right)}{\alpha - \frac{d^2}{2}}. \tag{29}$$

Thus, the true treatment effect for the second method is $\alpha(\beta_t - \beta_c)$, with $\alpha(\beta_t - \beta_c) > 0$ for mean difference and d for standardized mean difference. The corresponding values of β_t with different effect sizes and different skewness are shown in Table 3.4.

skewness	$\beta_t(d = 0.2)$	$\beta_t(d = 0.5)$	$\beta_t(d = 0.8)$

0.30	1.031	1.078	1.129
0.61	1.063	1.167	1.283
0.95	1.100	1.272	1.480
1.32	1.142	1.403	1,755
1.75	1.193	1.578	2.193

Table 3.4 values for β_t with different effect sizes and skewness

3.2.4 χ^2 Distribution

The χ^2 distribution is a special case of the gamma distribution. If a random variable follows a χ^2 distribution with n degrees of freedom, then it is equivalent to following a gamma ($\frac{d.f.}{2}$, 2). The skewness depends on the degrees of freedom for the χ^2 distribution. For convenience, we will simulate two $\chi^2 \sim (d.f.)$, and add a constant c to the intervention group to assess different effect sizes, where $\text{d.f.} = 2\alpha$. Also, $d.f.$ is rounded to the nearest integer. The values for α are the same as the ones we use for the gamma distribution, and $c = d * \sqrt{(2d.f.)}$, where c is also the true effect for the mean difference. The true effect for the standardized mean difference is *d*.

A summary table of all the algebra for different cases within the same distribution of the outcome variables in terms of the skewness, parameters, and corresponding true effect for mean difference and the standardized mean difference is provided in the Appendix.

3.3 Evaluation Criteria

To assess robustness, we will look at the relative bias of the estimated treatment effect from the true treatment effect, the coverage probability of the confidence interval for the estimated treatment effect, the power, and the type I error of the hypothesis test of the treatment effect.

3.3.1 Relative Bias

Relative bias is chosen to evaluate the accuracy of the estimated effect because it better compares the estimated effect across difference effect sizes. The estimated relative bias is the proportion of absolute bias to the true effect and is calculated as

$$relative\ bias = \frac{1}{N}\sum_{\mathrm{i}=1}^{\mathrm{N}}\left(\frac{\widehat{\delta_{\mathrm{i}}} - \delta}{\delta}\right) \tag{30}$$

in the simulation, where N is the number of replications, $\widehat{\delta_i}$ is the estimated effect of each replication and δ is the true effect. The further the relative bias is away from zero, the worse the estimate is, and vice versa. A positive relative bias indicates an overestimation of the true effect, and a negative relative bias indicates an underestimation of the true effect. Negative relative bias would probably be preferred by investigators because it gives more conservative treatment effect estimates and minimizes type I error rates.

3.3.2 Coverage Probability

The estimated coverage probability is the proportion of the confidence intervals that include the true effect in the 10000 replications, and it represents how well the confidence interval captures the true effect. It is measured as follows in the simulation:

$$coverage\ probability = \frac{\sum_{i=1}^{N} I(L_i \leq \delta \leq U_i)}{N}. \tag{31}$$

L_i and U_i are the lower and upper bound of the 95% confidence interval of the estimated effect in the i^{th} replication respectively, and $I(\cdot)$ is the indicator function, which equals 1 if $\widehat{\delta_i}$ falls between the confidence interval and equals 0 otherwise. The 95% coverage probability is desired and any deviation from 95% indicates a potential issue. However, an acceptable range of coverage probability is 92.5% to 97.5%, which is obtained from Bradley (1978), and the coverage probability outside of this range represents insufficient estimation of true effect.

3.3.3 Power

The power is the ability of a test to detect the true effect, also known as the probability of rejecting the null hypothesis ($\delta = 0$) when the alternative hypothesis ($\delta \neq 0$) is true. In our simulation, the power is estimated as

$$power = \frac{\sum_{i=1}^{N} I(p_i < \alpha)}{N}, \tag{32}$$

where p_i is the p-value of the i^{th} iteration, and α is the significance level, set to 0.05. The acceptable range for power is 80% and above, which is usually the desired level when calculating power for a clinical trial. The higher the power, the more sensitive the test is, and vice versa.

3.3.4 Type I Error

Type I error rate is the proportion of times of rejecting the null hypothesis when it is true. In the simulation, we calculate the type I error rate when the true effect is zero, as follows:

$$type\ I\ error\ = \frac{\sum_{i=1}^{N} I(p_i < \alpha)}{N}. \tag{33}$$

Ideally, the type I error rate should be close to 0.05. However, in this study, we will use a more liberal criterion suggested by Bradley (1978) where $0.5\alpha \leq type\ I\ error\ \leq 1.5\alpha$, and is 0.25 to 0.75. A type I error rate greater than this range suggests an increased false positive rate, and a type I error rate less than this range suggests that the test is underpowered.

Chapter 4 Results

This chapter contains the simulation results of different distributions of the outcome variable. The results are presented by different distributions of the outcome variable. Within each distribution, the results are separated into two parts, using the mean difference and the standardized mean difference to estimate the overall treatment effect, respectively. Furthermore, the relative bias and the coverage probability of the estimated overall treatment effect, as well as the power and the type I error rate of the meta-analysis, are presented in each section. As mentioned in Chapter 3, the coverage probability is taken to be acceptable when it falls between 0.925 and 0.975; the power is acceptable when greater than 0.8; and the acceptable range for the type I error rate is between 0.025 to 0.075. In addition, the table of the results for the type I error rate of the meta-analysis will be provided in the Appendix.

4.1 The Mixture of Normal Distributions

The relative bias is undefined when p=1, with $(1-p)*(\mu_t-\mu_c)$ being the true treatment effect, given $\mu_t-\mu_c \neq 0$. Therefore, we will only investigate the cases when p equals 0 to 0.9.

4.1.1 Mean Difference

4.1.1.1 $(1-p)*(\mu_t-\mu_c)$ as the True Effect

From Figure 4.1.1 and Figure 4.1.2, the relative bias slightly decreases as the number of studies and patients increases. A fluctuation of the relative bias with changing p is observed for a small effect size, especially when p is greater than 0.5 and small sample sizes (k=2, 5 or n=5, 10). The fluctuation disappears as the effect size increases. In most cases where the relative bias is beyond 5%, the patients per group per study are 5 with large p (0.8 and 0.9). As shown in Figure 4.1.3, with a fixed total sample size of 200, different combinations of k and n do not noticeably impact the relative bias.

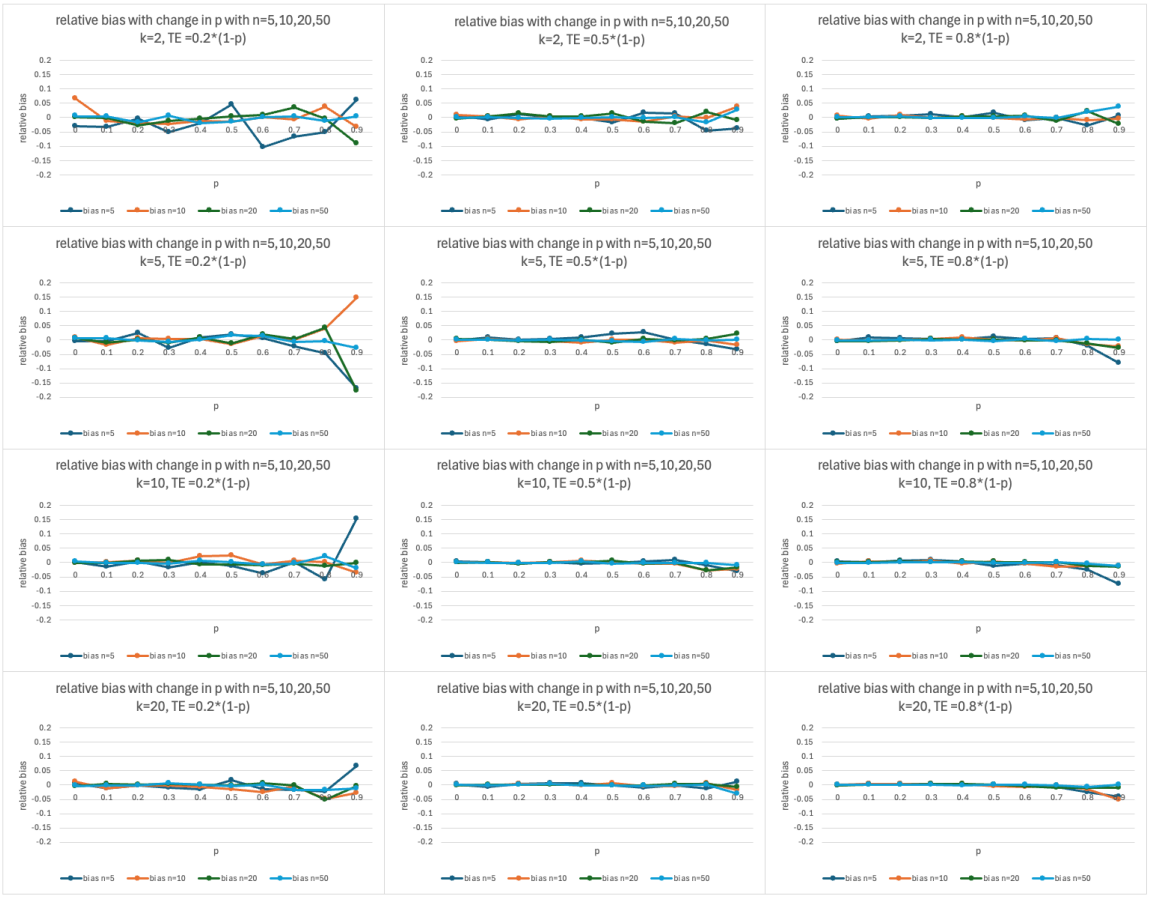

Figure 4.1.1The relative bias in the estimated treatment effect, as a function of p and n across different effect sizes, with a mixture of normals data using the mean difference for a fixed value of k at all levels of 2, 5, 10, and 20.

Figure 4.1.2 The relative bias in the estimated treatment effect, as a function of p and k across different effect sizes, with a mixture of normals data using the mean difference for a fixed value of n at all levels of 5, 10, 20, and 50.

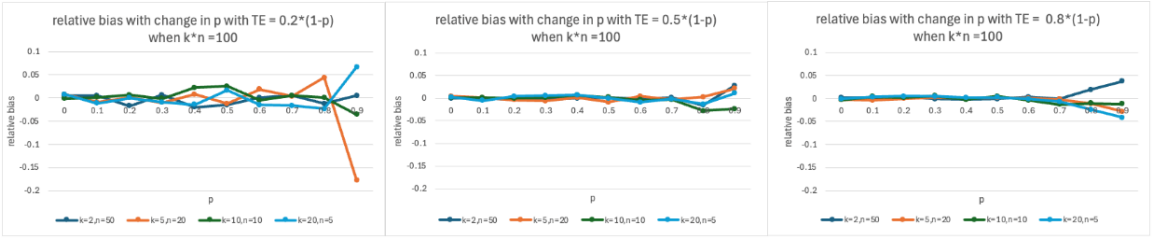

Figure 4.1.3 The relative bias in the estimated treatment effect, as a function of p and the different combinations of k and n across different effect sizes, with a mixture of normals data using mean difference for a fixed total sample size of 200.

From Figure 4.1.4, with a fixed value of k at all levels, the coverage probability is the closest to 0.95 when n increases. With a fixed value of n at 5, a smaller k gives better

coverage probability, as shown in Figure 4.1.5. However, the impact of k on the coverage probability gradually disappears as n increases. The effect size and p have a negligible effect on the coverage probability. Whenever the coverage probability is below 0.925, n is always equal to or less than 10. The coverage probability is closest to 0.95 when including the fewest studies and most patients per study with a fixed total sample size of 200, which is shown in Figure 4.1.6.

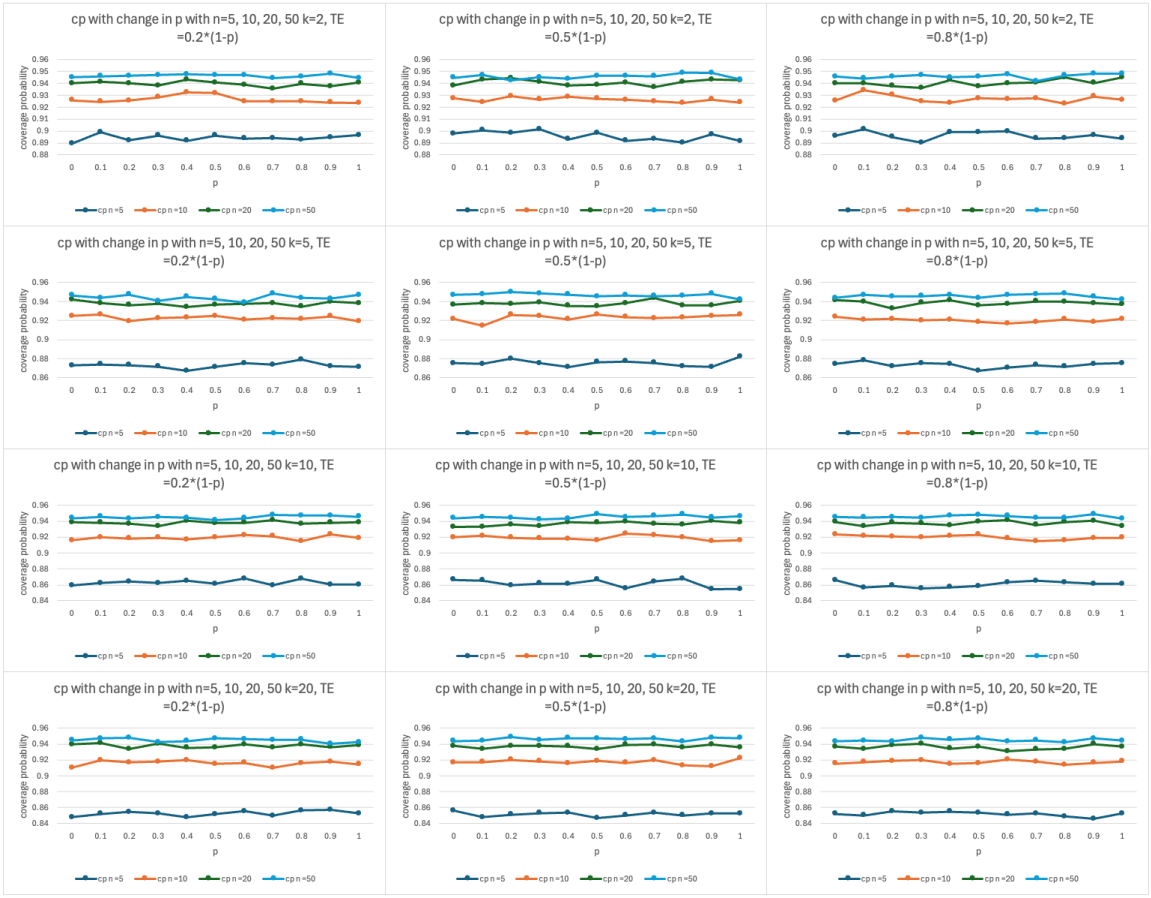

Figure 4.1.4 The coverage probability in the estimated treatment effect, as a function of p and n across different effect sizes, with a mixture of normals data using the mean difference for a fixed value of k at all levels of 2, 5, 10, and 20.

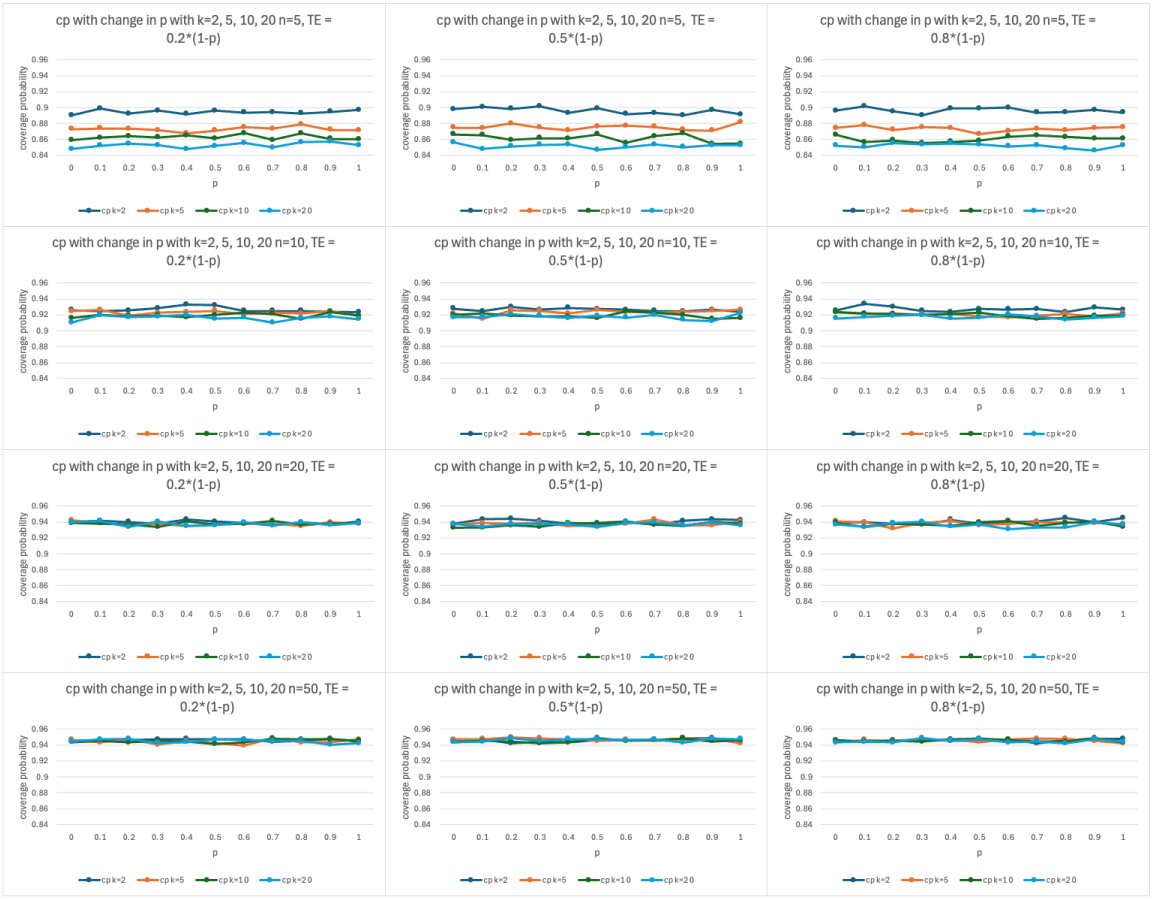

Figure 4.1.5 The coverage probability in the estimated treatment effect, as a function of p and k across different effect sizes, with a mixture of normals data using the mean difference for a fixed value of n at all levels of 5, 10, 20, and 50.

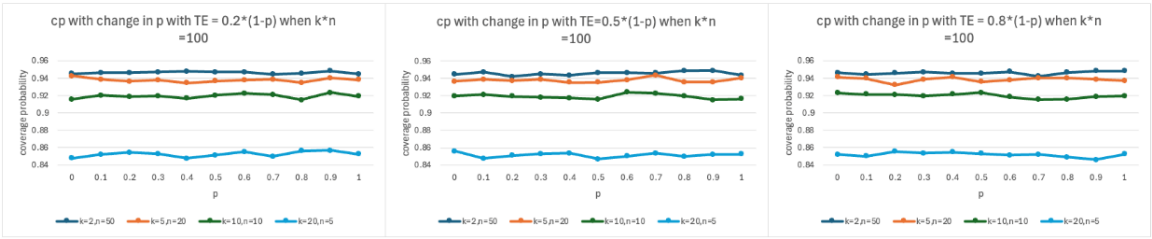

Figure 4.1.6 The coverage probability in the estimated treatment effect, as a function of p and the different combinations of k and n across different effect sizes, with a mixture of normals data using the mean difference for a fixed total sample size of 200.

As shown in Figure 4.1.7 and Figure 4.1.8, the power increases as the number of studies, patients per study, and the effect size increases. The power declines as p increases and

drops sharply from 1 when p > 0.5, especially with a larger sample size and effect size. The power is greater than 0.8 when the sample size exceeds 800 for a small effect size. With a moderate effect size, the power exceeds 0.8 when the total sample size exceeds 200. With a large effect size, the power exceeds 0.8 when the total sample size exceeds 50. In addition, when the power exceeds 0.8, the corresponding range of p increases as sample sizes and effect sizes increase. Figure 4.1.9 shows that with a fixed total sample size of 200, there is no detectable impact from varied combinations of k and n on power.

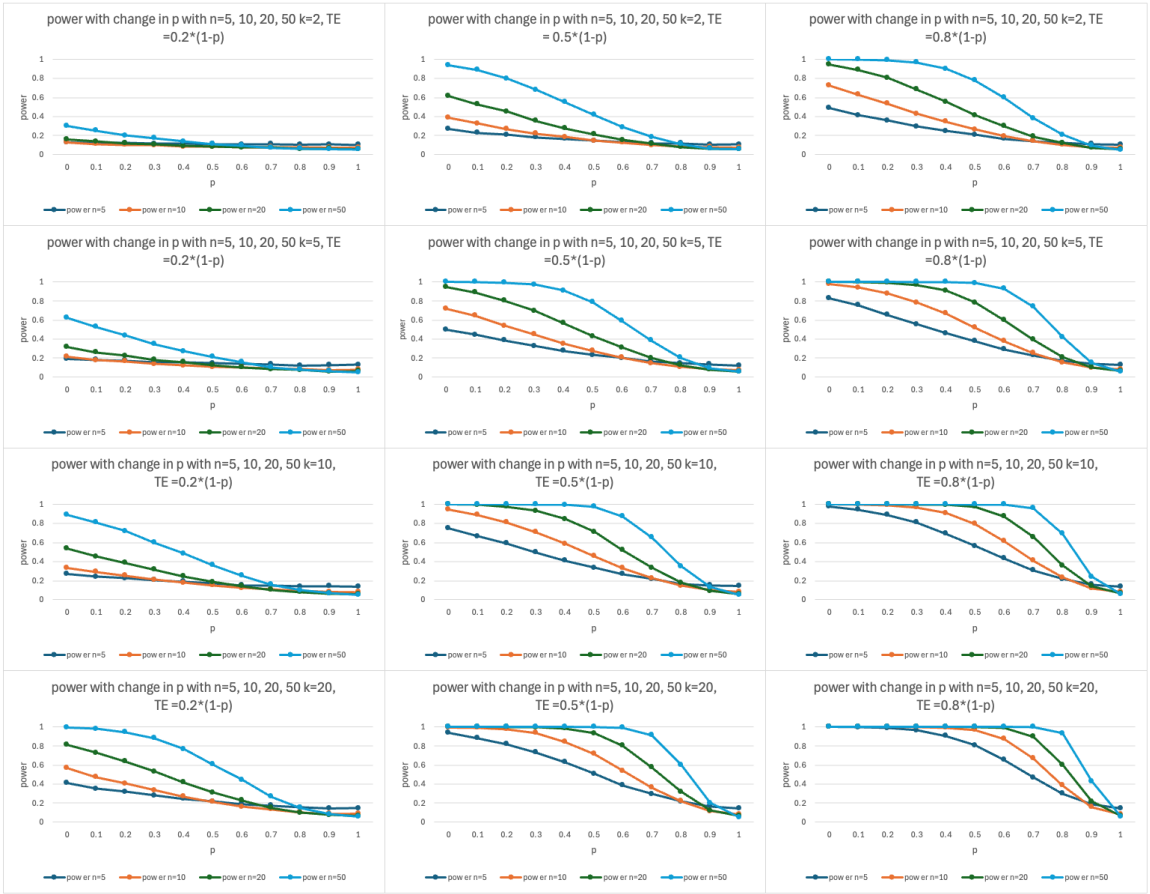

Figure 4.1.7 The power in the test, as a function of p and n across different effect sizes, with a mixture of normals data using the mean difference for a fixed value of k at all levels of 2, 5, 10, and 20.

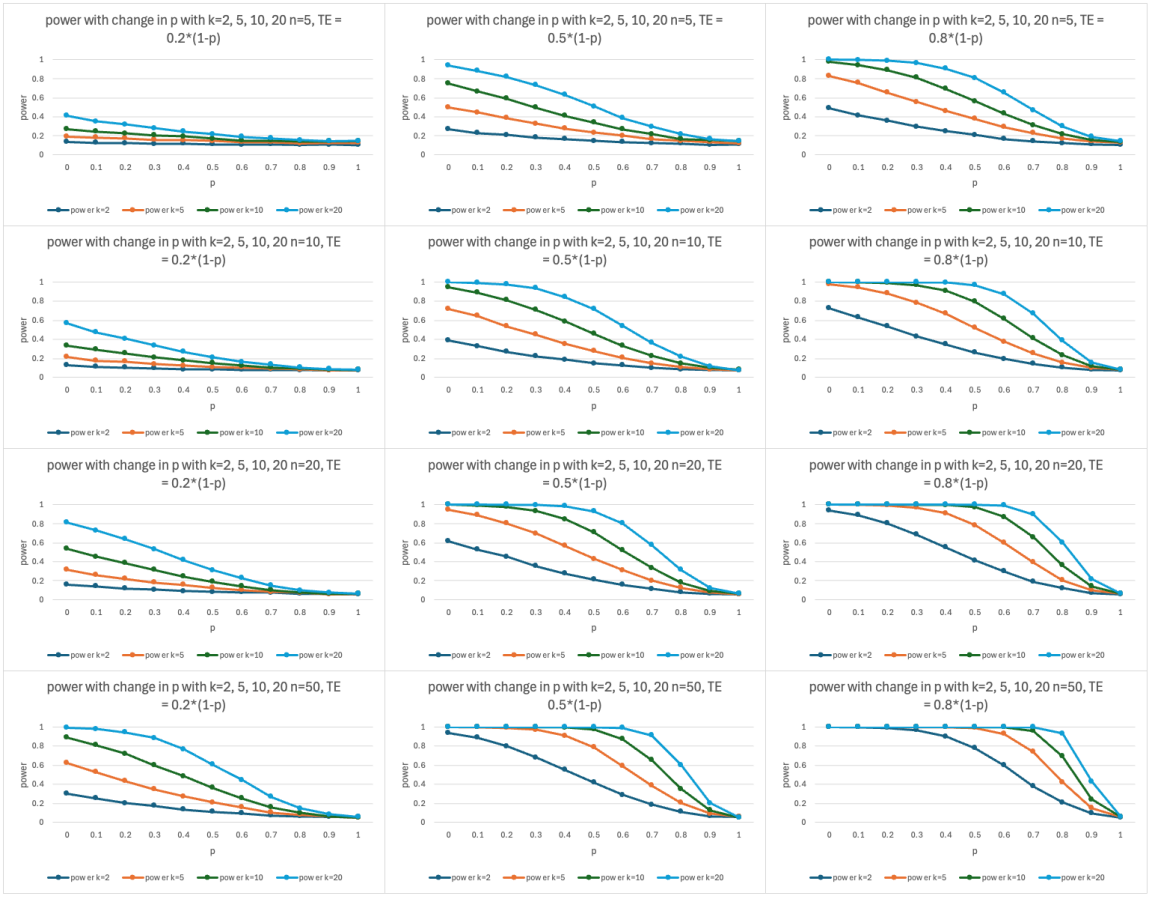

Figure 4.1.8 The power in the test, as a function of p and k across different effect sizes, with a mixture of normals data using the mean difference for a fixed value of n at all levels of 5, 10, 20, and 50.

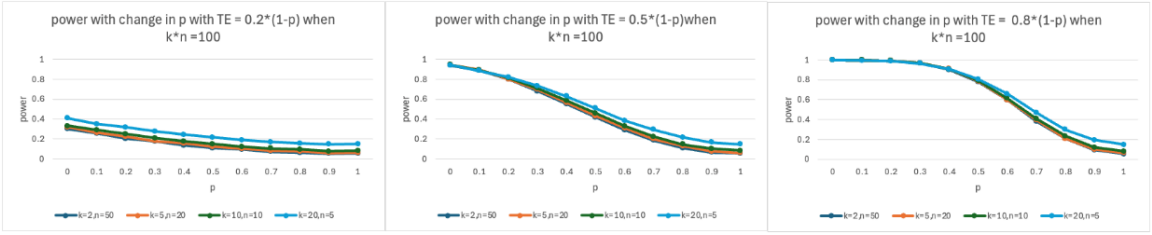

Figure 4.1.9 The power in the test, as a function of p and different combinations of k and n across different effect sizes, with a mixture of normals data using the mean difference for a fixed total sample size of 200.

The type I error rate falls outside the 0.025 and 0.075 range when n is 5 or 10 with varying values of k. Moreover, it is above 0.05 in most cases.

4.1.1.2 $\mu_t - \mu_c$ as the True Effect

Figure 4.1.10 and Figure 4.1.11 show a slight fluctuation in the relative bias with small k and n for a small effect size (d=0.2). The relative bias moves farther from zero to 100% as p increases. Moreover, the relative bias is mostly negative when $\mu_t - \mu c$ is the true effect. The absolute value of relative bias is below 5% in the negative direction only when p=0 for all values of n, k, and effect sizes. From Figure 4.1.12, there is no detectable effect of different combinations of k and n on the relative bias when the total effect size is fixed at 200.

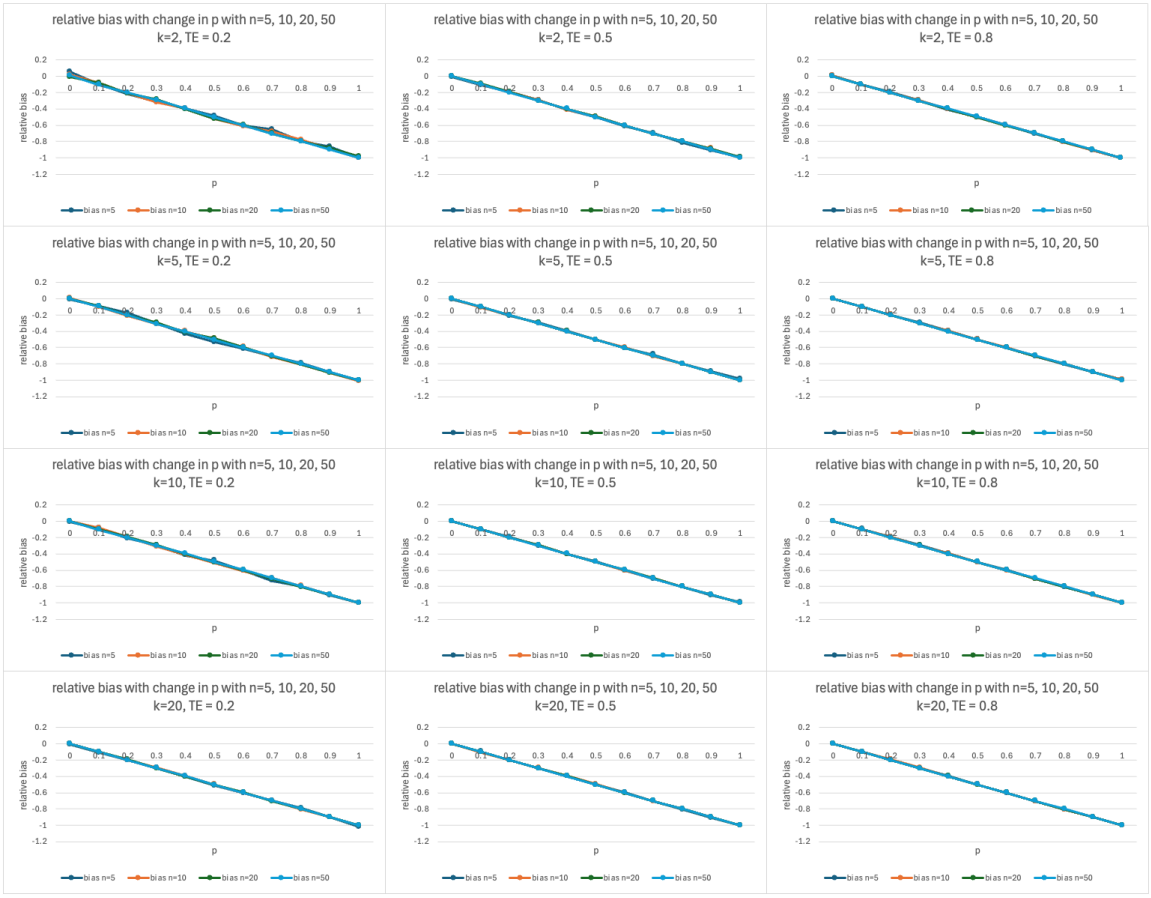

Figure 4.1.10 The relative bias in the estimated treatment effect, as a function of p and n across different effect sizes, with a mixture of normals data using the mean difference for a fixed value of k at all levels of 2, 5, 10, and 20.

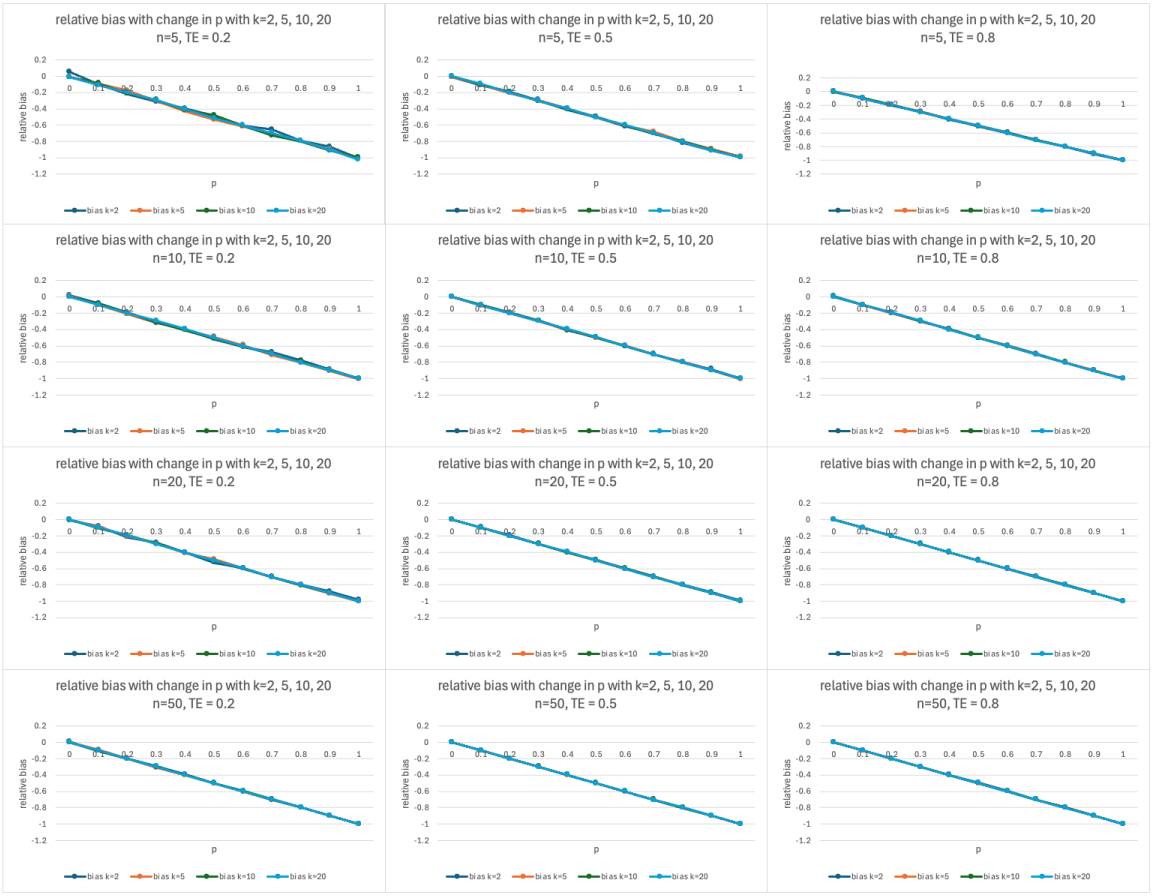

Figure 4.1.11 The relative bias in the estimated treatment effect, as a function of p and k across different effect sizes, with a mixture of normals data using the mean difference for a fixed value of n at all levels of 5, 10, 20, and 50.

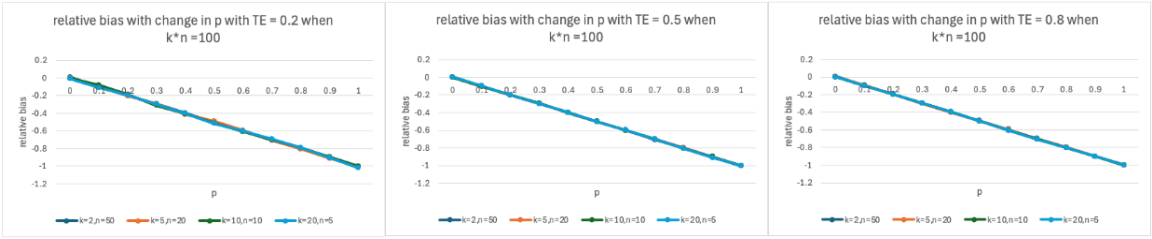

Figure 4.1.12 The relative bias in the estimated treatment effect, as a function of p and different combinations of k and n across different effect sizes, with a mixture of normals data using the mean difference for a fixed total sample size of 200.

Figure 4.1.13 and Figure 4.1.14 show that the coverage probability decreases as p increases and decreases more rapidly as n, k, and the effect size increase. Most cases give

a coverage probability below 0.925. Whenever the coverage probability is within the acceptable range, p is equal to and less than 0.3, as long as the sample size is sufficiently large. The coverage probability is below 0.95 in all situations. From Figure 4.1.15, with a fixed total sample size of 200, a higher sample size per study led to coverage probability closer to 0.95 for small p and smaller effect sizes. The coverage probability dropped sharply from 0% to 100% with larger effect sizes as p changes.

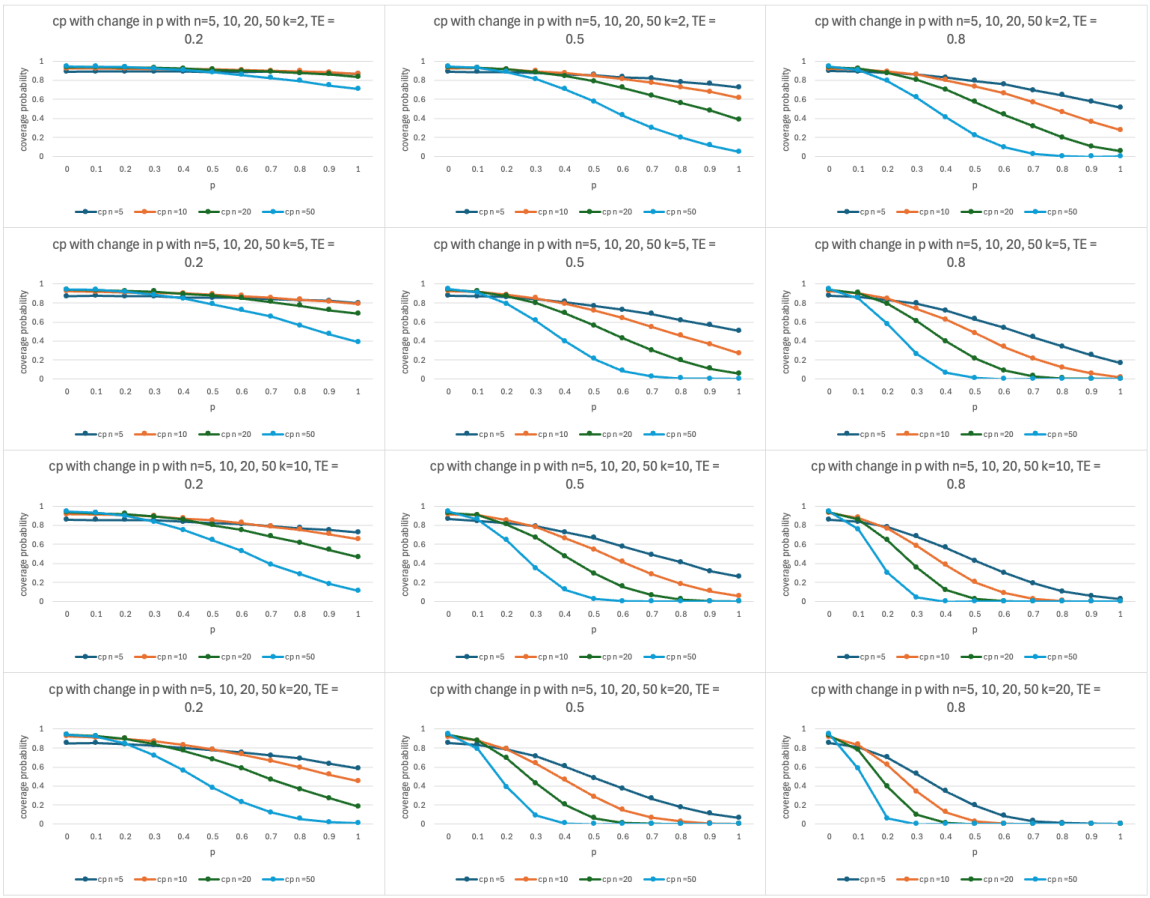

Figure 4.1.13 The coverage probability in the estimated treatment effect, as a function of p and n across different effect sizes, with a mixture of normals data using the mean difference for a fixed value of k at all levels of 2, 5, 10, and 20.

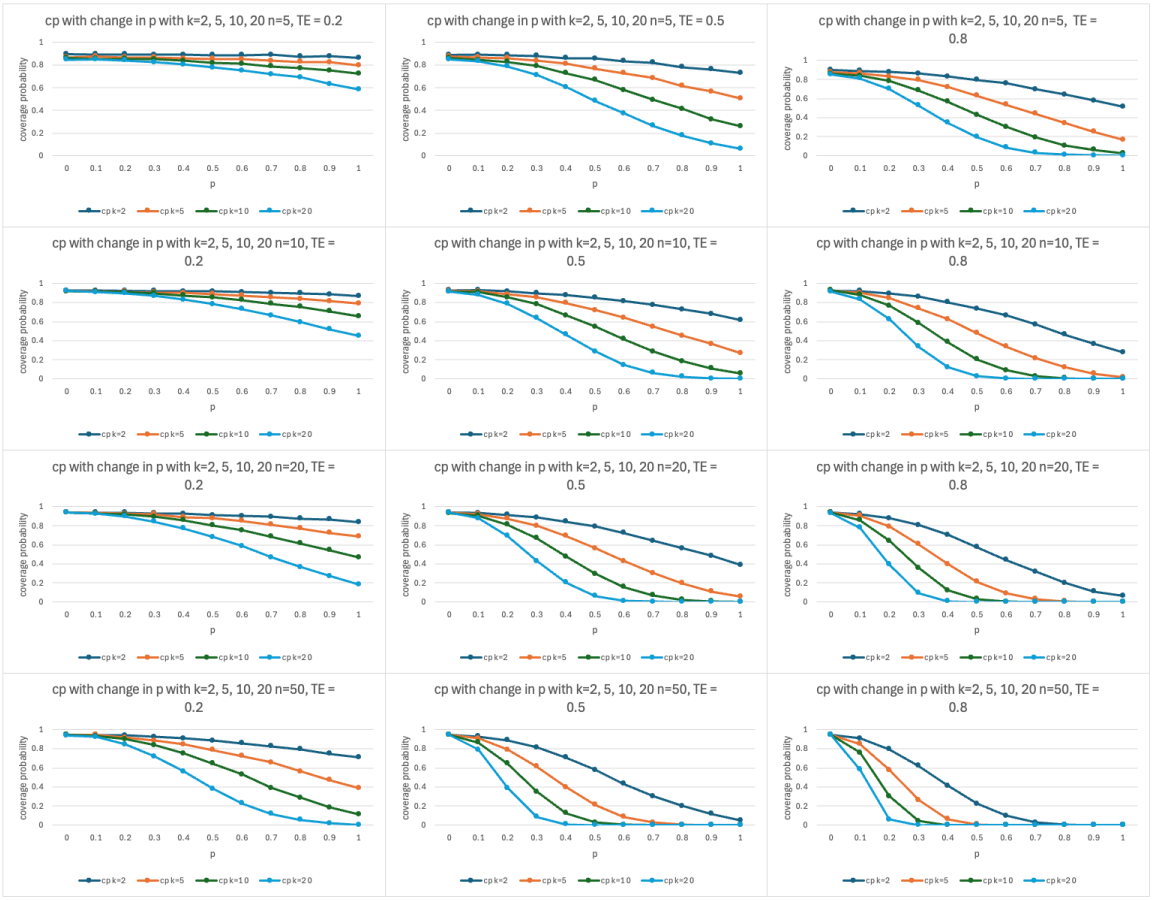

Figure 4.1.14 The coverage probability in the estimated treatment effect, as a function of p and k across different effect sizes, with a mixture of normals data using the mean difference for a fixed value of n at all levels of 5, 10, 20, and 50.

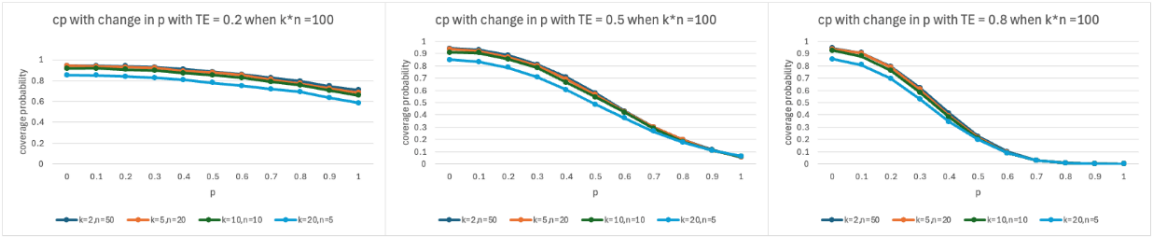

Figure 4.1.15 The coverage probability in the estimated treatment effect, as a function of p and different combinations of k and n across different effect sizes, with a mixture of normals data using the mean difference for a fixed total sample size of 200.

As shown in Figure 4.1.16, Figure 4.1.17, and Figure 4.1.18, the trend of power and type

I error rate is the same as when the true treatment effect is $(1-p)*(\mu_t-\mu_c)$.

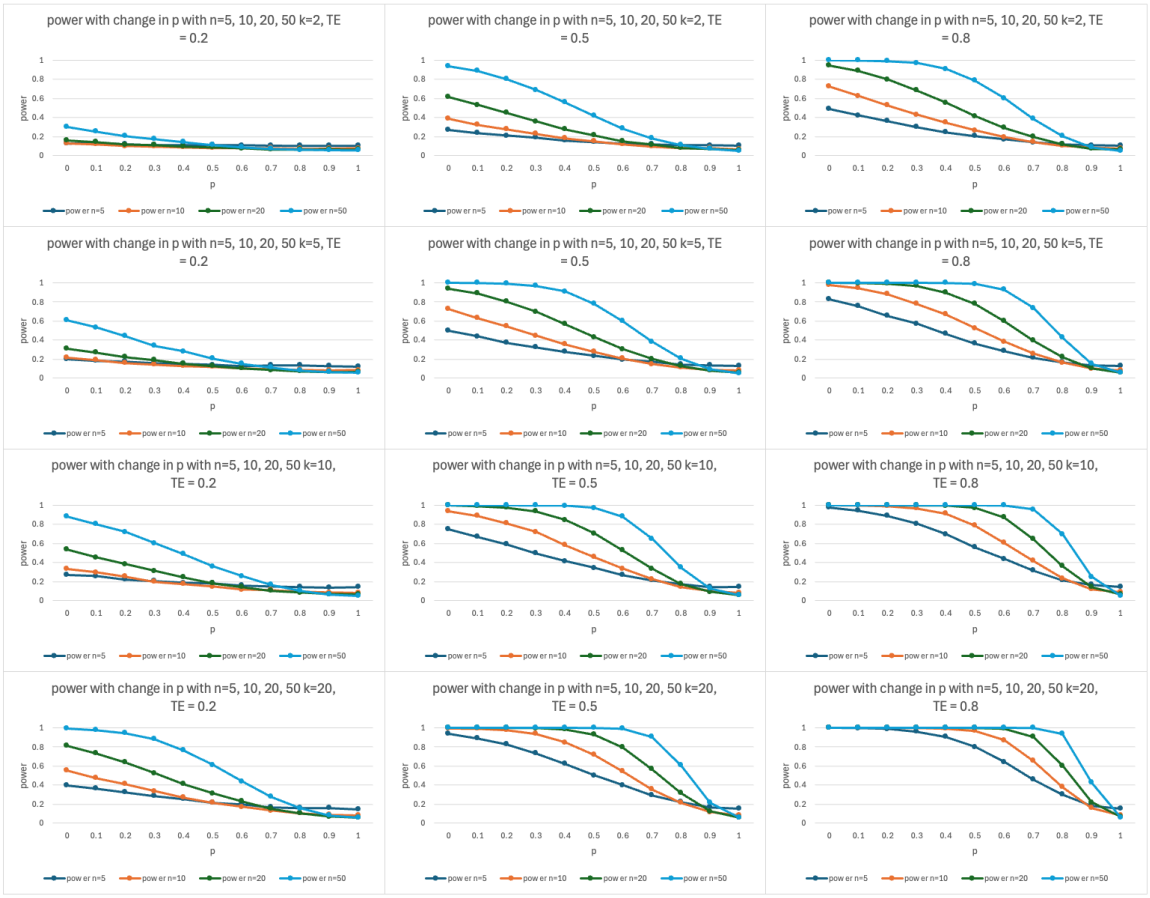

Figure 4.1.16 The power in the test, as a function of p and n across different effect sizes, with a mixture of normals data using the mean difference for a fixed value of k at all levels of 2, 5, 10, and 20.

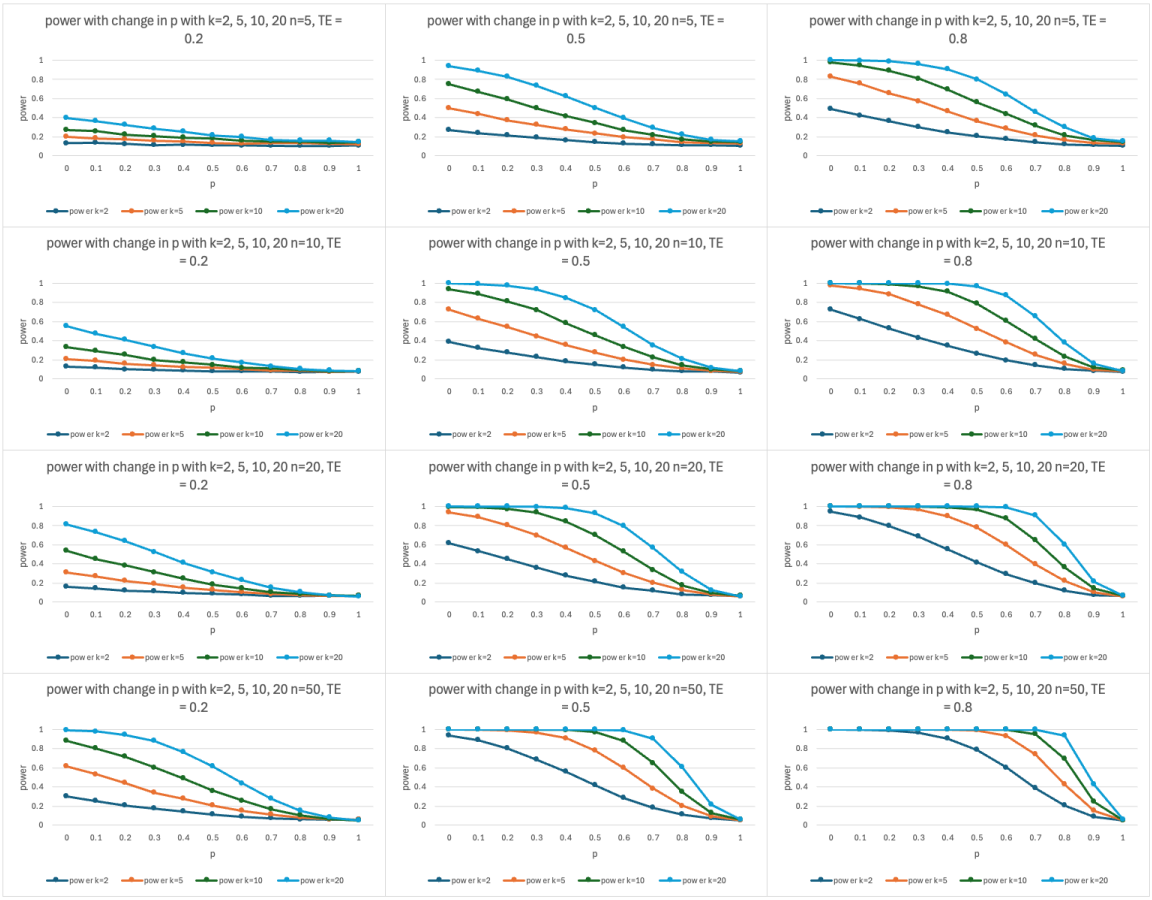

Figure 4.1.17 The power in the test, as a function of p and k across different effect sizes, with a mixture of normals data using the mean difference for a fixed value of n at all levels of 5, 10, 20, and 50.

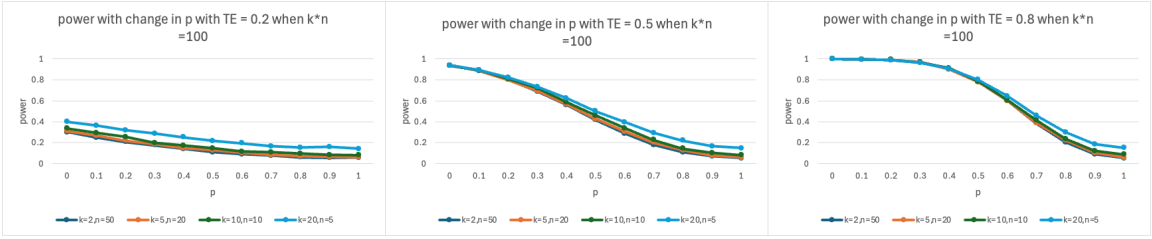

Figure 4.1.18 The power in the test, as a function of p and different combinations of k and n across different effect sizes, with a mixture of normals data using the mean difference for a fixed total sample size of 200.

4.1.2 Standardized Mean Difference

4.1.2.1 $(1-p)*d$ as the True Effect

Figure 4.1.19 shows that the relative bias approaches zero as n increases with a fixed value of k at all levels. When n=5, the relative bias moves farther from zero as p increases to 0.9 for a small effect size. There is a negligible effect of p on the relative bias for moderate and large effect sizes. As shown in Figure 4.1.20, with a fixed value of n at 5, a smaller k gives a better relative bias as it is the closest to zero. However, with a larger value of fixed n, the effect of k on the relative bias is minor. When n is 20 and 50, the relative bias is the worst at p = 0.5 and is the best when p = 0 or 1. Whenever the relative bias is greater than 10%, n is 5 for most cases. Moreover, most situations have a negative relative bias. With a fixed total sample size of 200, fewer studies and more patients per study give a better relative bias, as shown in Figure 4.1.21.

Figure 4.1.19 The relative bias in the estimated treatment effect, as a function of p and n across different effect sizes, with a mixture of normals data using the standardized mean difference for a fixed value of k at all levels of 2, 5, 10, and 20.

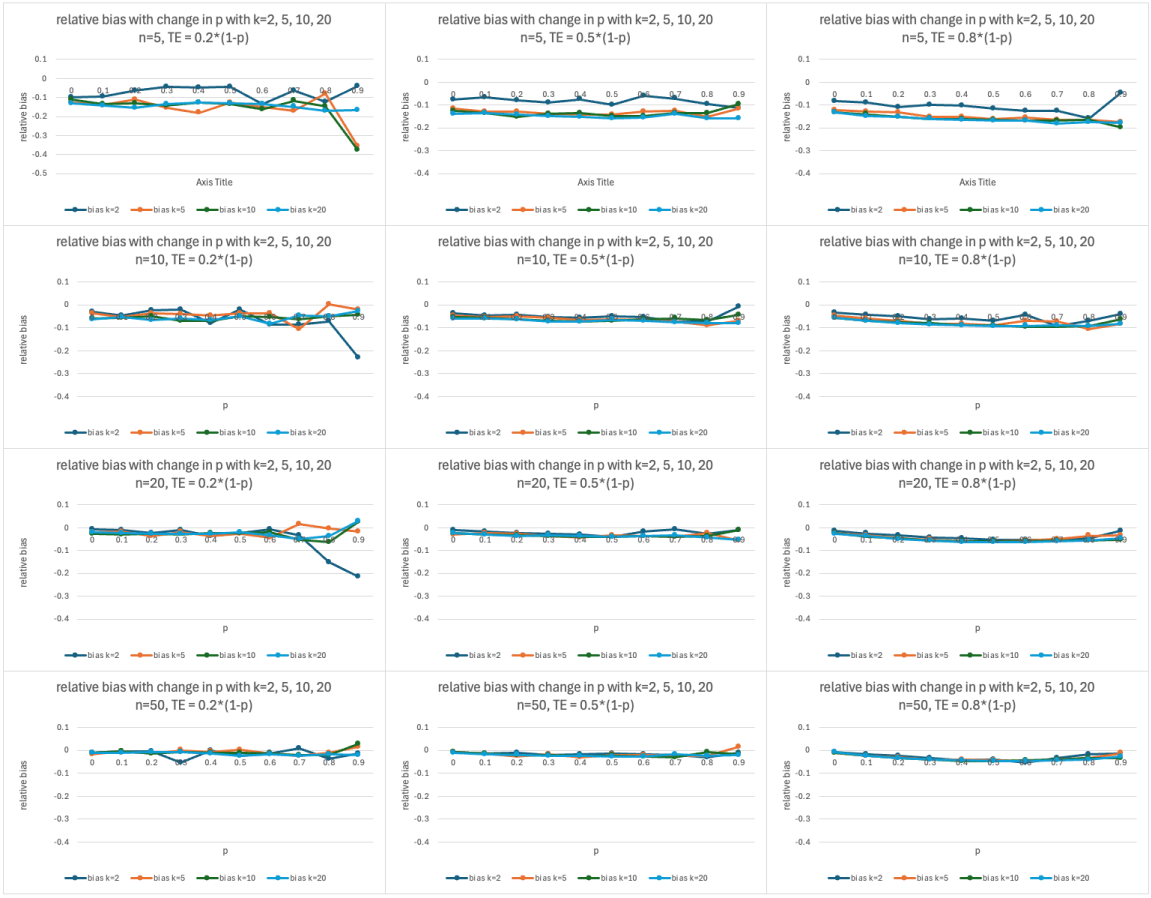

Figure 4.1.20 The relative bias in the estimated treatment effect, as a function of p and k across different effect sizes, with a mixture of normals data using the standardized mean difference for a fixed value of n at all levels of 5, 10, 20, and 50.

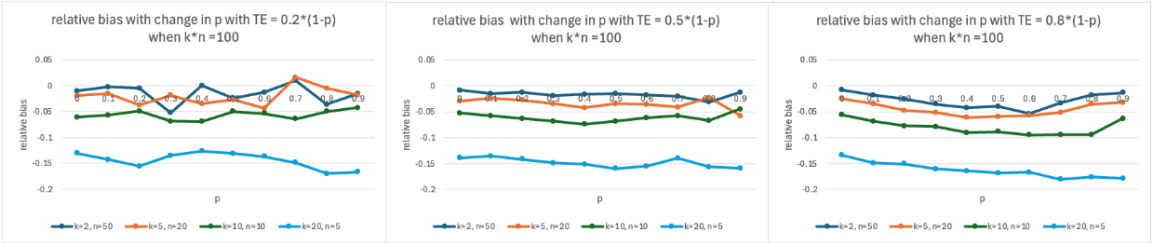

Figure 4.1.21 The relative bias in the estimated treatment effect, as a function of p and different combinations of k and n across different effect sizes, with a mixture of normals data using the standardized mean difference for a fixed total sample size of 200.

As illustrated in Figure 4.1.22, with a fixed value of smaller k (k= 2, 5), a larger n gives better coverage probability, and p does not impact the coverage probability. When with a

fixed larger k, the trend of the coverage probability is similar to the smaller k with a small effect size (d=0.2). However, the coverage probability increases from 0.95 as p increases with a large effect size (d=0.8). From Figure 4.1.23, with a fixed value of n at all levels, the coverage probability does not change much when changing k or p for small and moderate effect sizes. When n is 5, the coverage probability falls outside the acceptable range and is above 0.975 for most cases. The coverage probability is the closest to 0.95 and is the most stable with increasing p when the fewest studies and most patients are included in the meta-analysis for a fixed total sample size of 200 as shown in Figure 4.1.24.

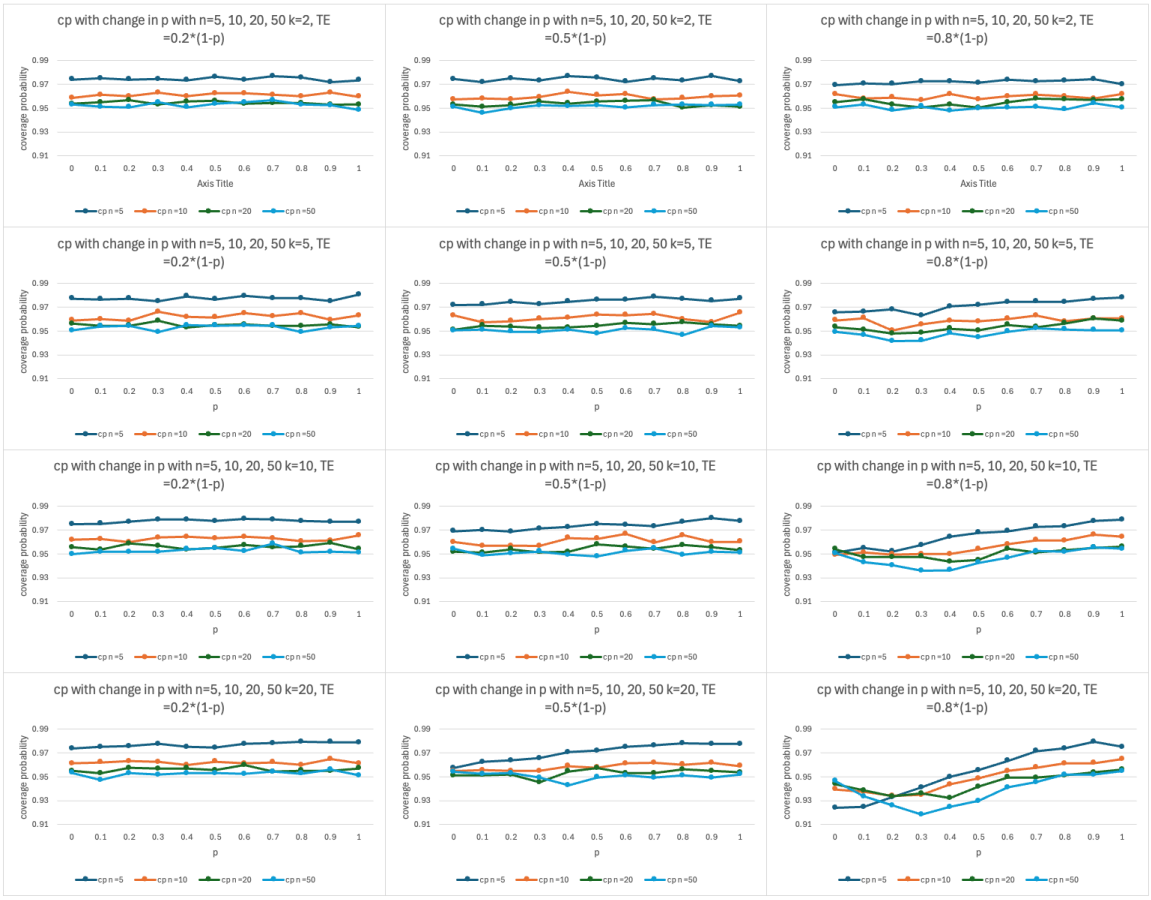

Figure 4.1.22 The coverage probability in the estimated treatment effect, as a function of p and n across different effect sizes, with a mixture of normals data using the standardized mean difference for a fixed value of k at all levels of 2, 5, 10, and 20.

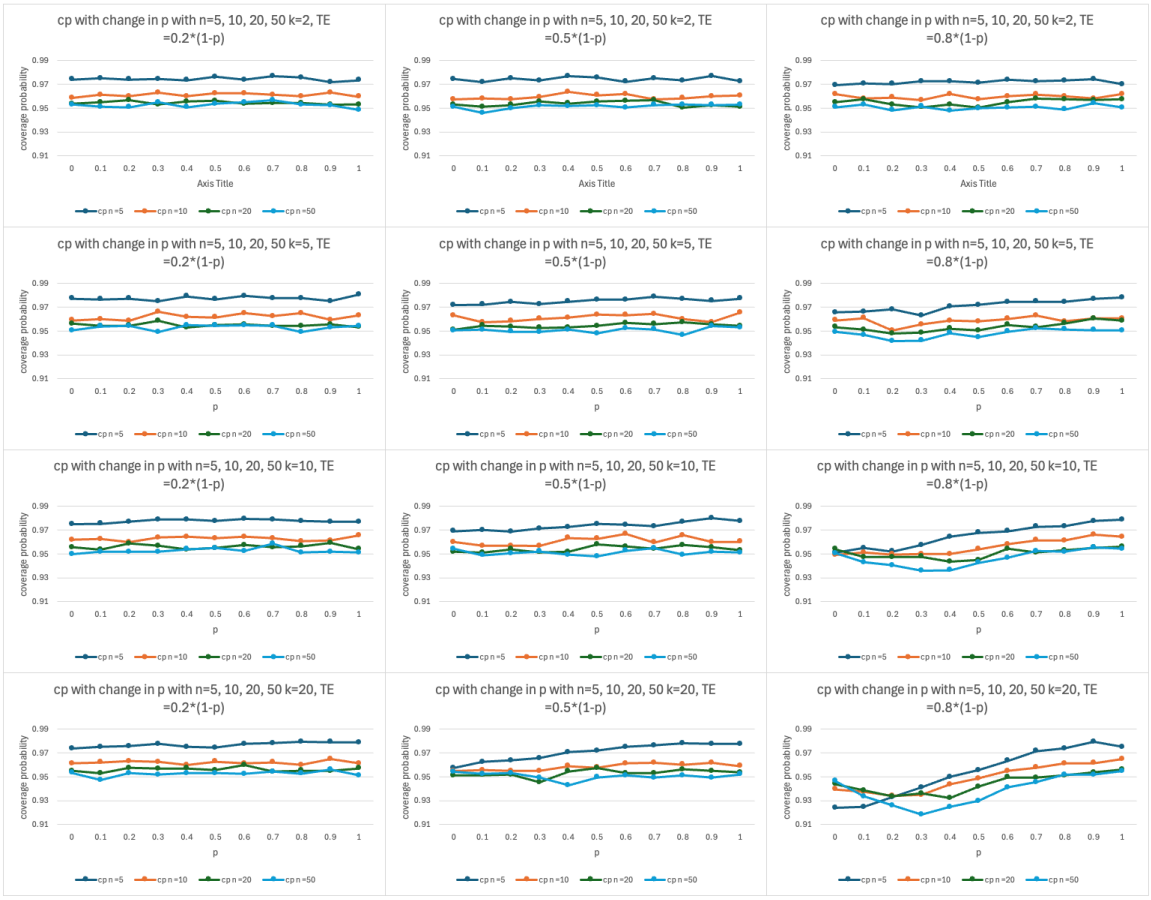

Figure 4.1.23 The coverage probability in the estimated treatment effect, as a function of p and k across different effect sizes, with a mixture of normals data using the standardized mean difference for a fixed value of n at all levels of 5, 10, 20, and 50.

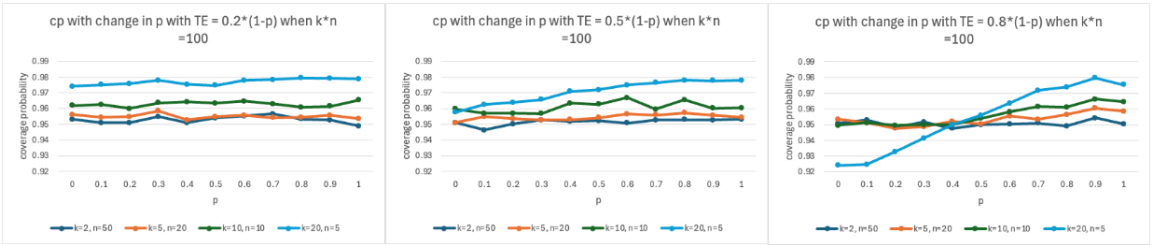

Figure 4.1.24 The coverage probability in the estimated treatment effect, as a function of p and different combinations of k and n across different effect sizes, with a mixture of normals data using the standardized mean difference for a fixed total sample size of 200.

The trend of power with changing n, k, effect sizes, and p is similar to the one using the mean difference, which is presented in Figure 4.1.25 and Figure 4.1.26. However, the

power is greater than 0.8 when the sample size exceeds 1000 for a small effect size. With a large effect size, the power exceeds 0.8 when the total sample size exceeds 80. As illustrated in Figure 4.1.27, with a fixed total sample size of 200, there is a slightly lower power with k=20 and n=5 for moderate and large effect sizes.

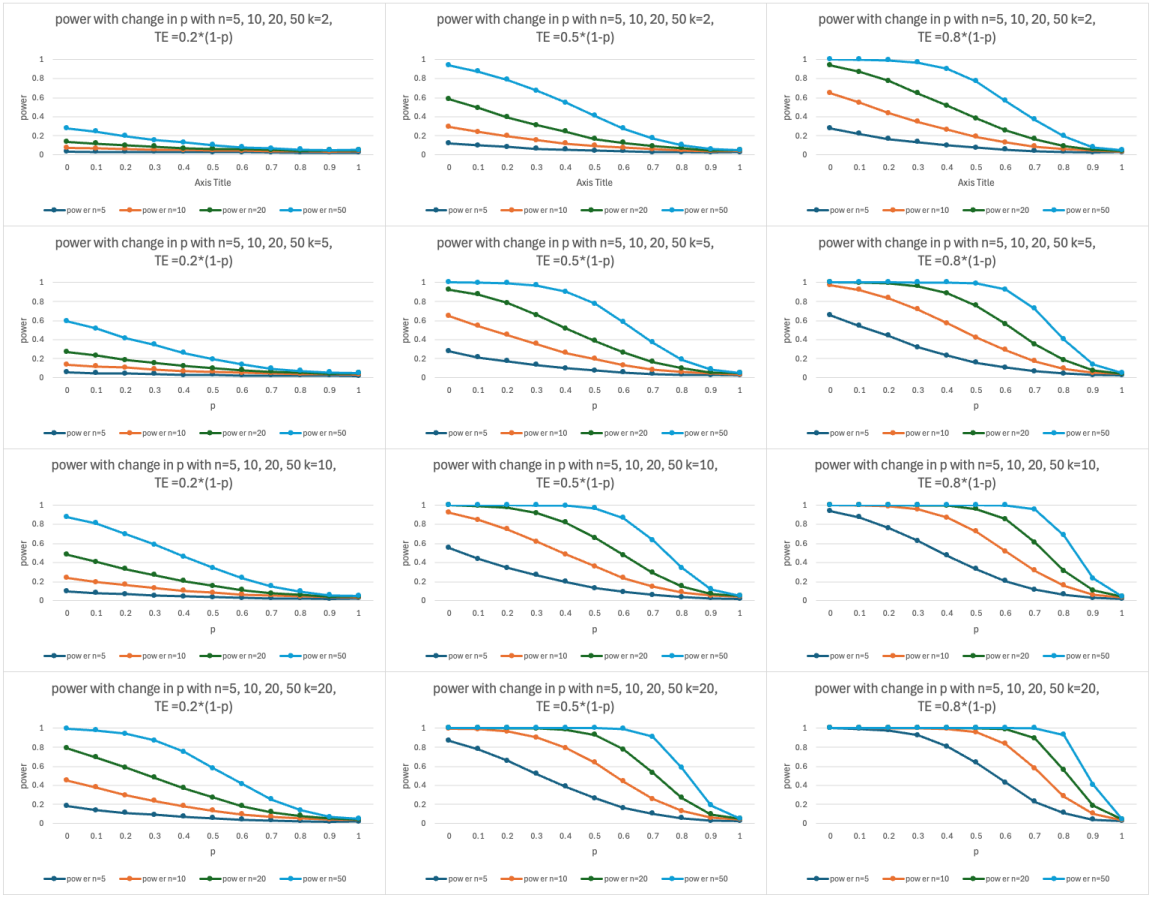

Figure 4.1.25 The power in the test, as a function of p and n across different effect sizes, with a mixture of normals data using the standardized mean difference for a fixed value of k at all levels of 2, 5, 10, and 20.

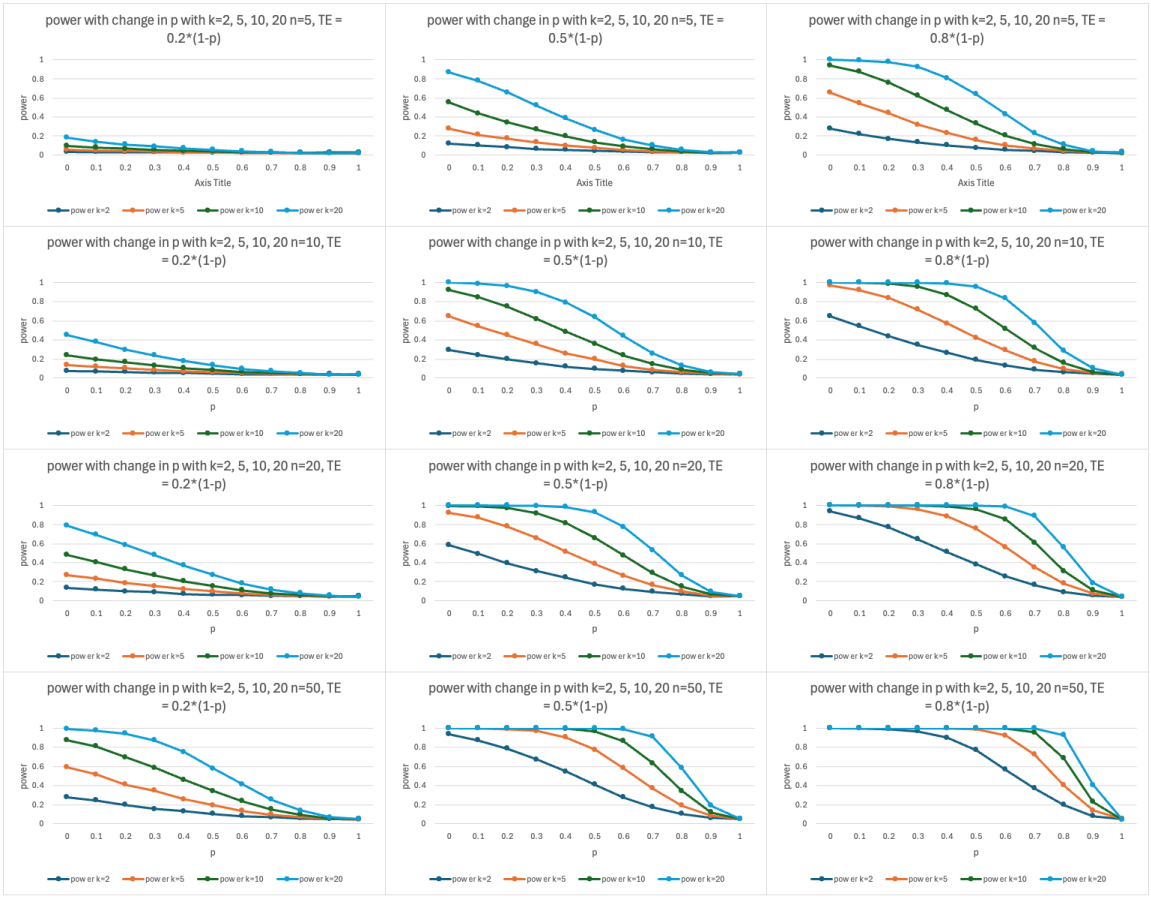

Figure 4.1.26 The power in the test, as a function of p and k across different effect sizes, with a mixture of normals data using the standardized mean difference for a fixed value of n at all levels of 5, 10, 20, and 50.

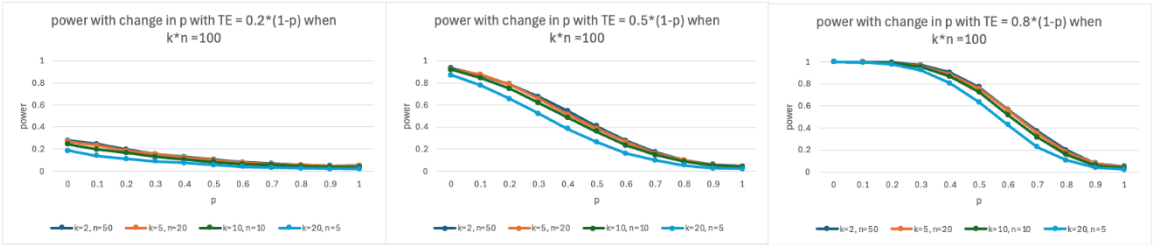

Figure 4.1.27 The power in the test as a function of p and different combinations of k and n across different effect sizes, with a mixture of normals data using the standardized mean difference for a fixed total sample size of 200.

When n is 5 and k is greater than 2, the type I error rate falls outside the acceptable range.

Moreover, it is below 0.05 in most cases.

4.1.2.2 *d* as the True Effect

From Figure 4.1.28, the relative bias is slightly worse when n decreases with a fixed value of k. k does not have a noticeable impact on the relative bias when n is fixed, as shown in Figure 4.1.29. The relative bias moves away from 0 to 100% as p increases. Furthermore, effect sizes have a negligible impact on relative bias. The relative bias is less than 5% only when p = 0. In addition, the relative bias is negative in most cases. As presented in Figure 4.1.30, with a fixed total sample size of 200, fewer studies and more patients per study give better relative bias, but the effect is small.

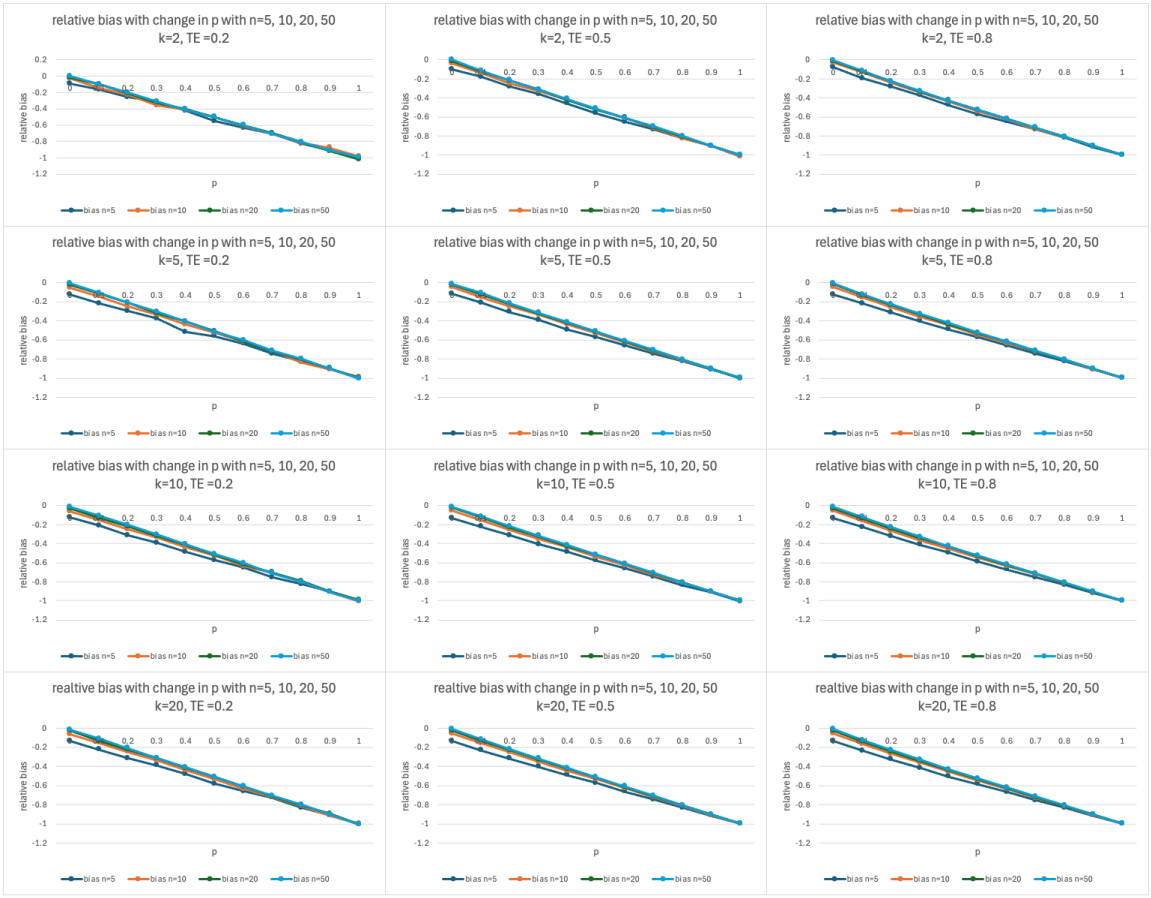

Figure 4.1.28 The relative bias in the estimated treatment effect, as a function of p and n across different effect sizes, with a mixture of normals data using the standardized mean difference for a fixed value of k at all levels of 2, 5, 10, and 20.

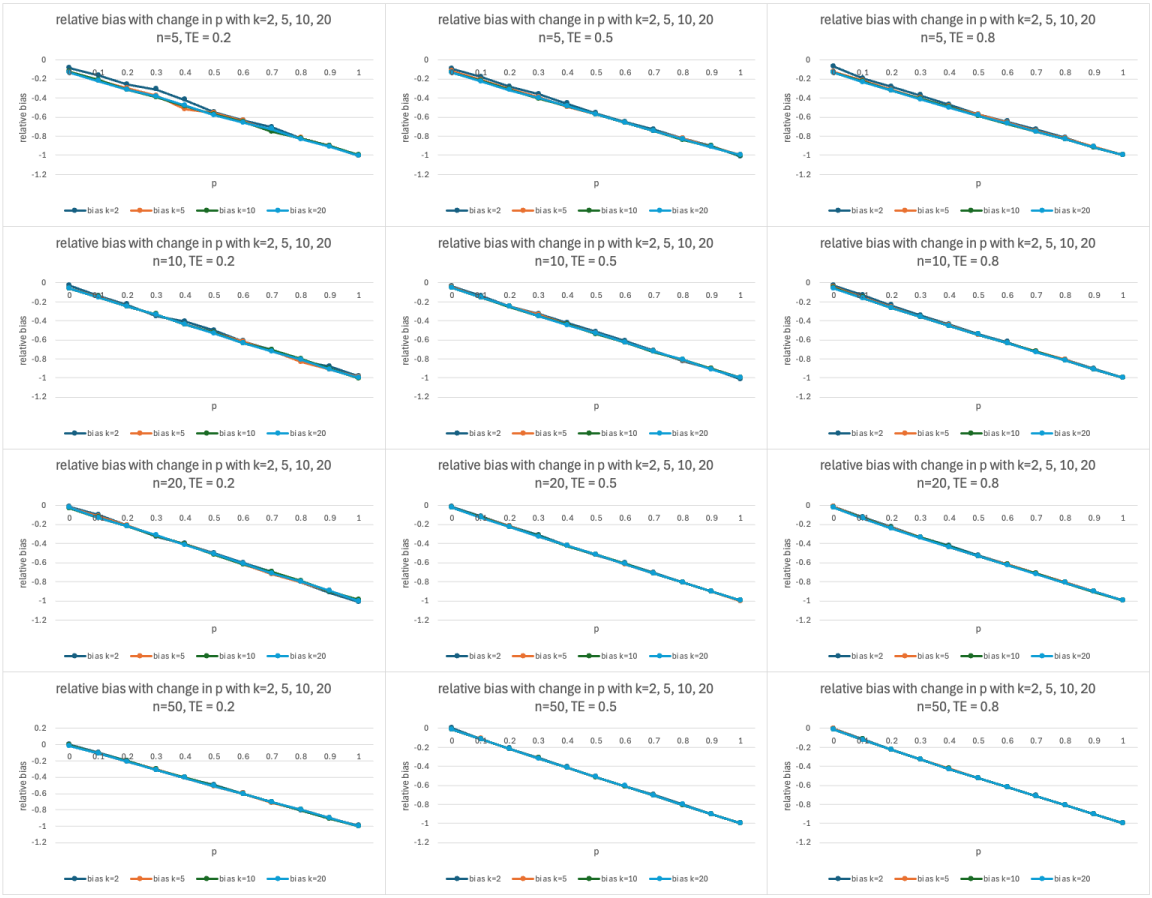

Figure 4.1.29 The relative bias in the estimated treatment effect, as a function of p and k across different effect sizes, with a mixture of normals data using the standardized mean difference for a fixed value of n at all levels of 5, 10, 20, and 50.

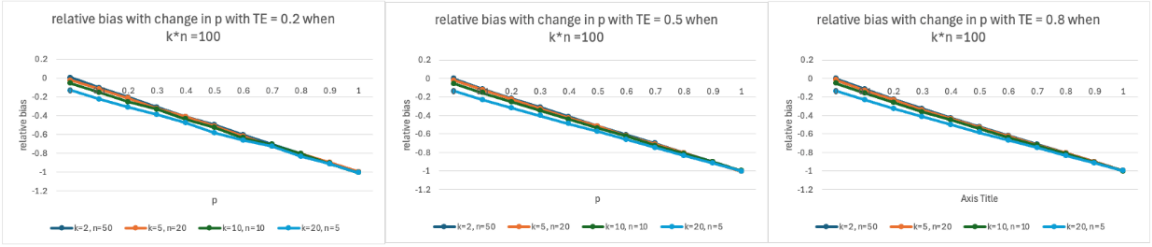

Figure 4.1.30 The relative bias in the estimated treatment effect, as a function of p and different combinations of k and n across different effect sizes, with a mixture of normals data using the standardized mean difference for a fixed total sample size of 200.

Figure 4.1.31 and Figure 4.1.32 indicate that the coverage probability declines as p increases, and this effect is more obvious as n, k, and effect sizes grow larger. In addition,

the coverage probability starts to drop at a lower p as n, k, and effect sizes increase. The number of cases where the coverage probability falls in the acceptable range is less than those outside. With a smaller p and a large total sample size, the coverage probability is within the acceptable range. With a fixed total sample size of 200, the coverage probability dropped sharply from 0% to 100% with larger effect sizes as p changes. Different combinations of k and n have an undetectable effect on the coverage probability, as illustrated in Figure 4.1.33.

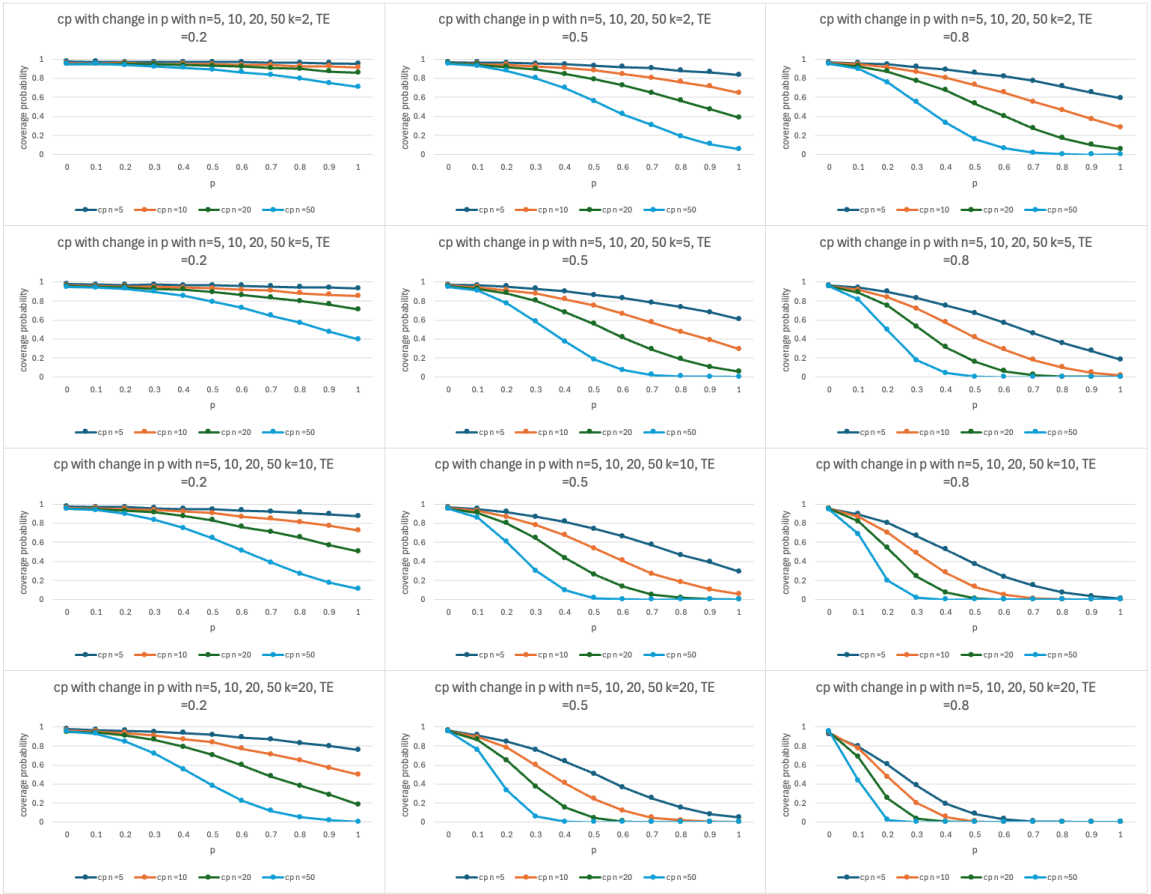

Figure 4.1.31 The coverage probability in the estimated treatment effect, as a function of p and n across different effect sizes, with a mixture of normals data using the standardized mean difference for a fixed value of k at all levels of 2, 5, 10, and 20.

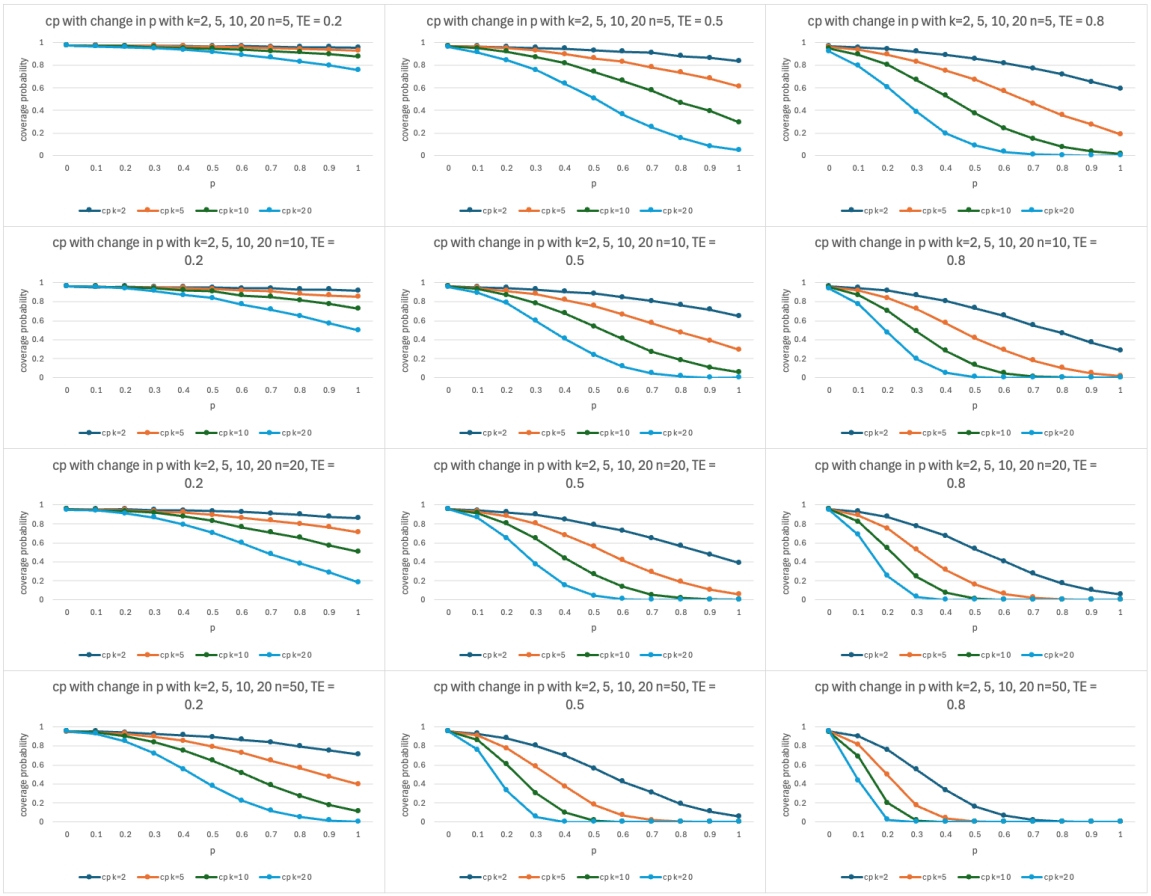

Figure 4.1.32 The coverage probability in the estimated treatment effect, as a function of p and k across different effect sizes, with a mixture of normals data using the standardized mean difference for a fixed value of n at all levels of 5, 10, 20, and 50.

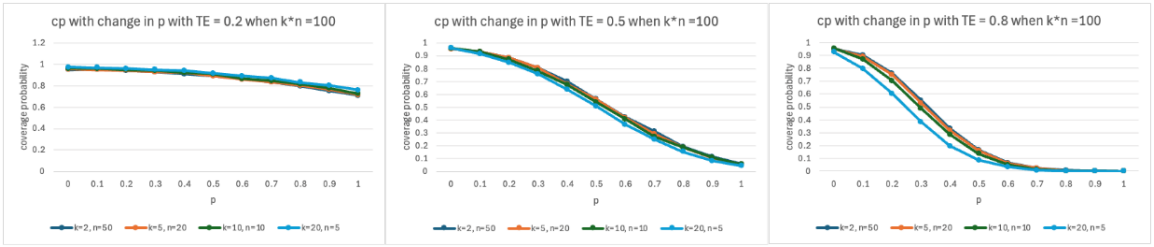

Figure 4.1.33 The coverage probability in the estimated treatment effect, as a function of p and different combinations of k and n across different effect sizes, with a mixture of normals data using the standardized mean difference for a fixed total sample size of 200.

The power is the same when using $(1-p)*(\mu_t-\mu_c)$, as shown in, Figure 4.1.34,

Figure 4.1.35, and Figure 4.1.36.

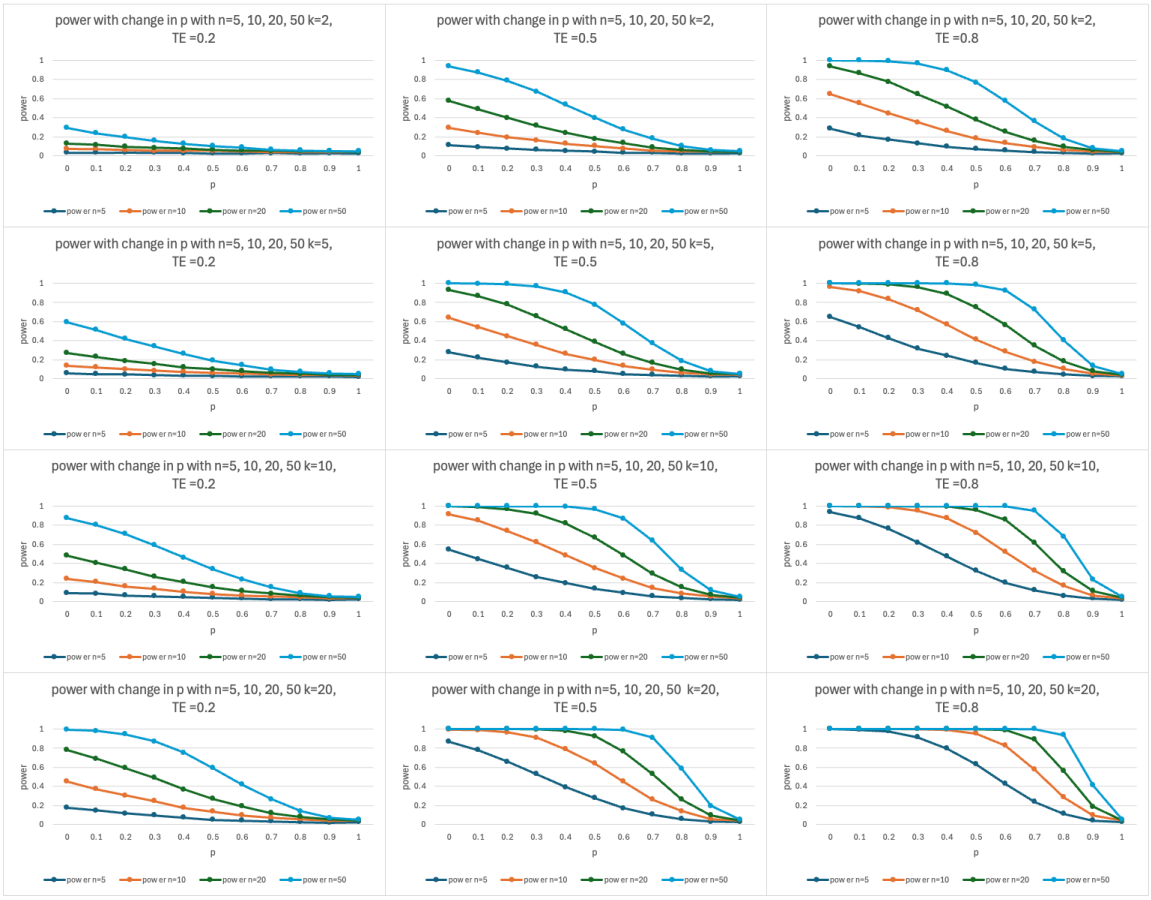

Figure 4.1.34 The power in the test, as a function of p and n across different effect sizes, with a mixture of normals data using the standardized mean difference for a fixed value of k at all levels of 2, 5, 10, and 20.

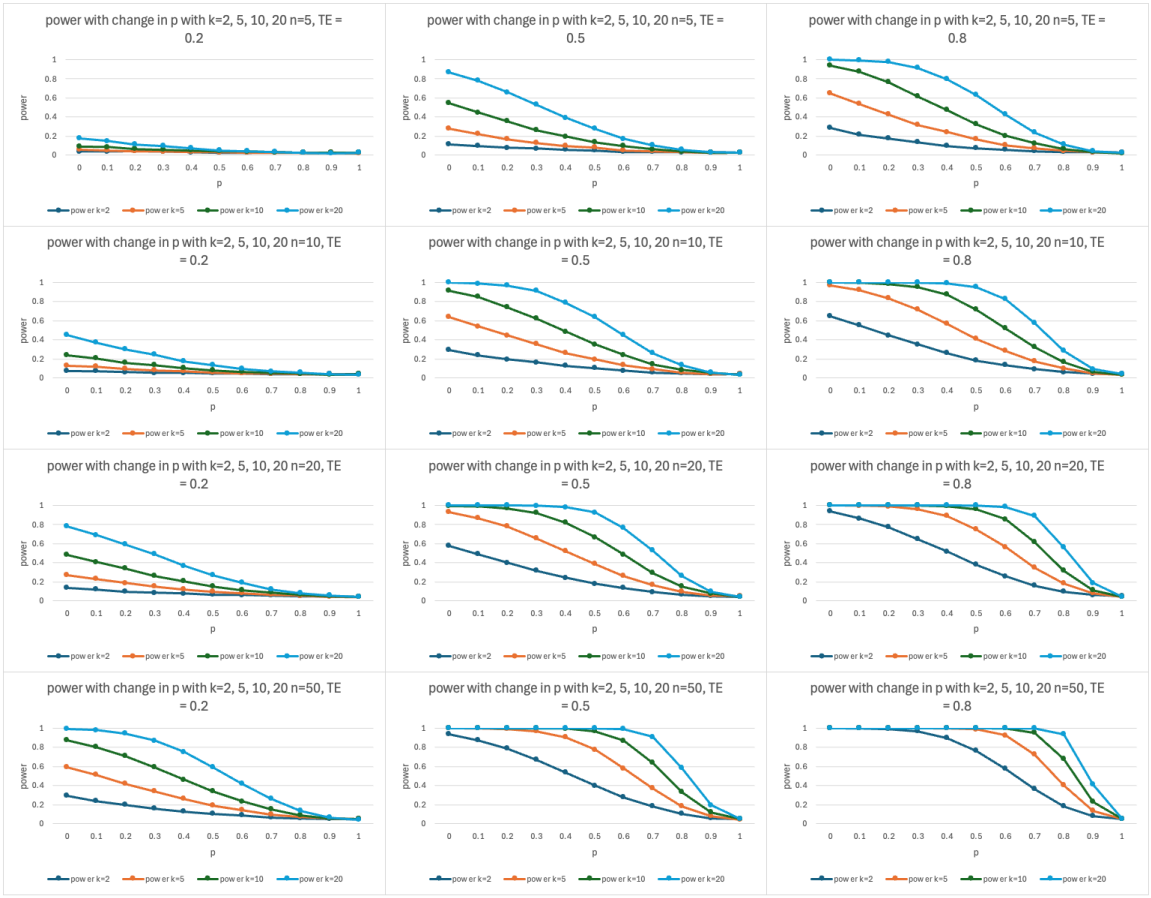

Figure 4.1.35 The power in the test, as a function of p and k across different effect sizes, with a mixture of normals data using the standardized mean difference for a fixed value of n at all levels of 5, 10, 20, and 50.

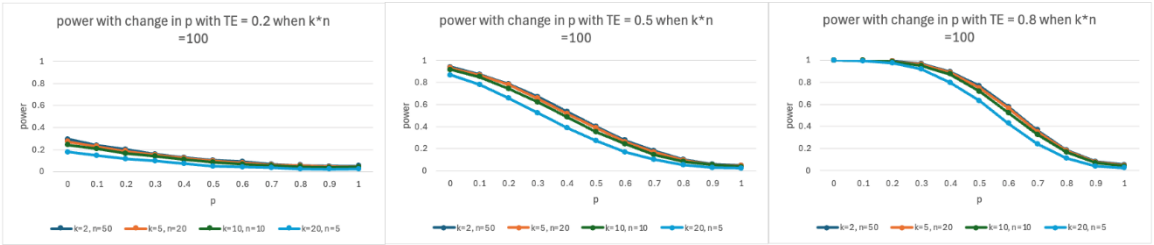

Figure 4.1.36 The power in the test, as a function of p and different combinations of k and n across different effect sizes, with a mixture of normals data using the standardized mean difference for a fixed total sample size of 200.

4.2 Lognormal Distribution

4.2.1 Mean Difference

4.2.1.1 Adding a Constant

As shown in Figure 4.2.1, with a fixed value of k at 20, the absolute relative bias is below 1% in all cases. With a fixed value of n at 50, the relative bias is within 1% with different values of k, effect sizes, and skewness, as presented in Figure 4.2.2. There is an undetectable effect of skewness. The effect sizes have little impact, whereas the relative bias is more stable and smaller with larger effect sizes. The absolute relative bias is below 3% in all cases. Figure 4.2.3 shows that different combinations of k and n do not have an important impact on the relative bias with a fixed total sample size of 200.

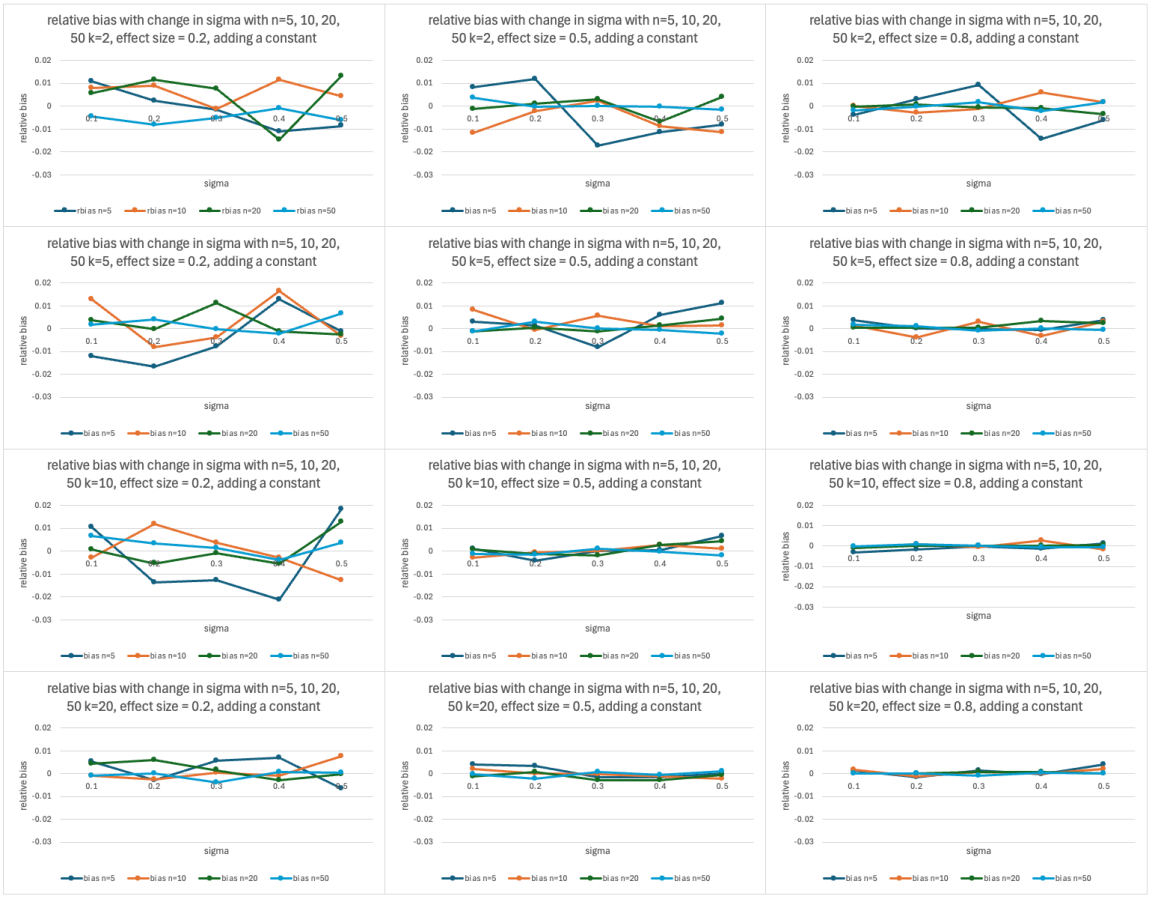

Figure 4.2.1 The relative bias in the estimated treatment effect, as a function of σ and n across different effect sizes, with adding a constant to lognormal data using the mean difference for a fixed value of k at all levels of 2, 5, 10, and 20.

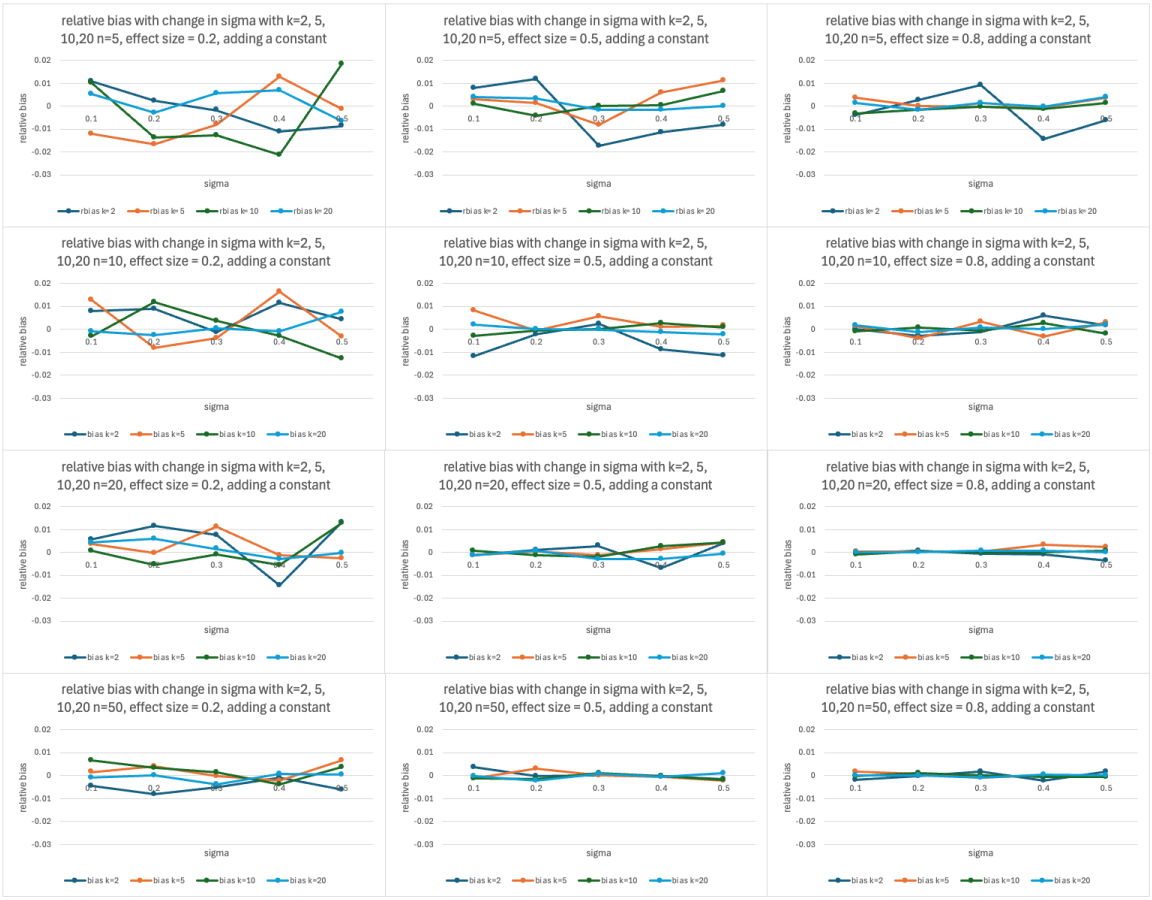

Figure 4.2.2 The relative bias in the estimated treatment effect, as a function of σ and k across different effect sizes, with adding a constant to lognormal data using mean difference for a fixed value of n at all levels of 5, 10, 20, and 50.

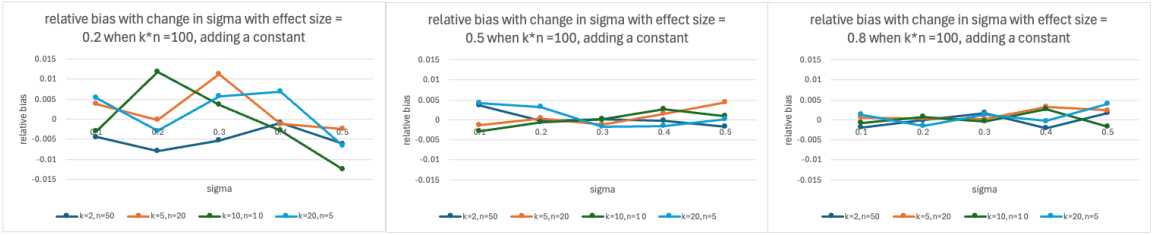

Figure 4.2.3 The relative bias in the estimated treatment effect, as a function of σ and different combinations of k and n across different effect sizes, with adding a constant to lognormal data using the mean difference for a fixed total sample size of 200.

With a fixed value of k at all levels, the coverage probability is the closest to 0.95 when n=50 and is the farthest from 0.95 when n=5, as illustrated in Figure 4.2.4. Figure 4.2.5

shows that with a fixed value of n at 5, the coverage probability increases as k decreases. However, when n=5, the coverage probability is below 0.925 in all cases. With other values of n, k does not have an important change in the coverage probability. Skewness and effect sizes have an undetectable impact on the coverage probability. In all cases, the coverage probability is below 0.95. With a fixed total sample size of 200, the coverage probability is ideal (closest to 0.95) when k=2 and n=50 and is below the acceptable range when k=10, n=10, and k=20, n=5, as shown in Figure 4.2.6.

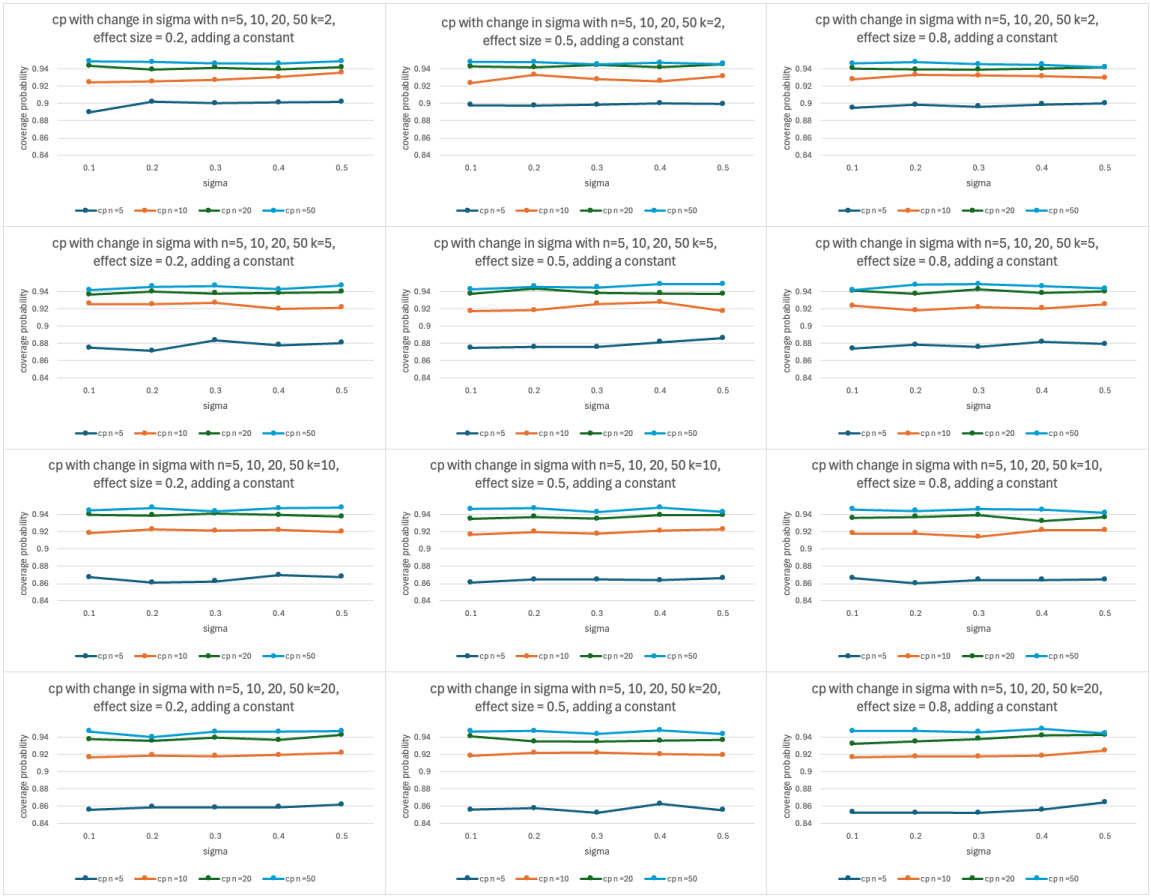

Figure 4.2.4 The coverage probability in the estimated treatment effect, as a function of σ and n across different effect sizes, with adding a constant to lognormal data using the mean difference for a fixed value of k at all levels of 2, 5, 10, and 20.

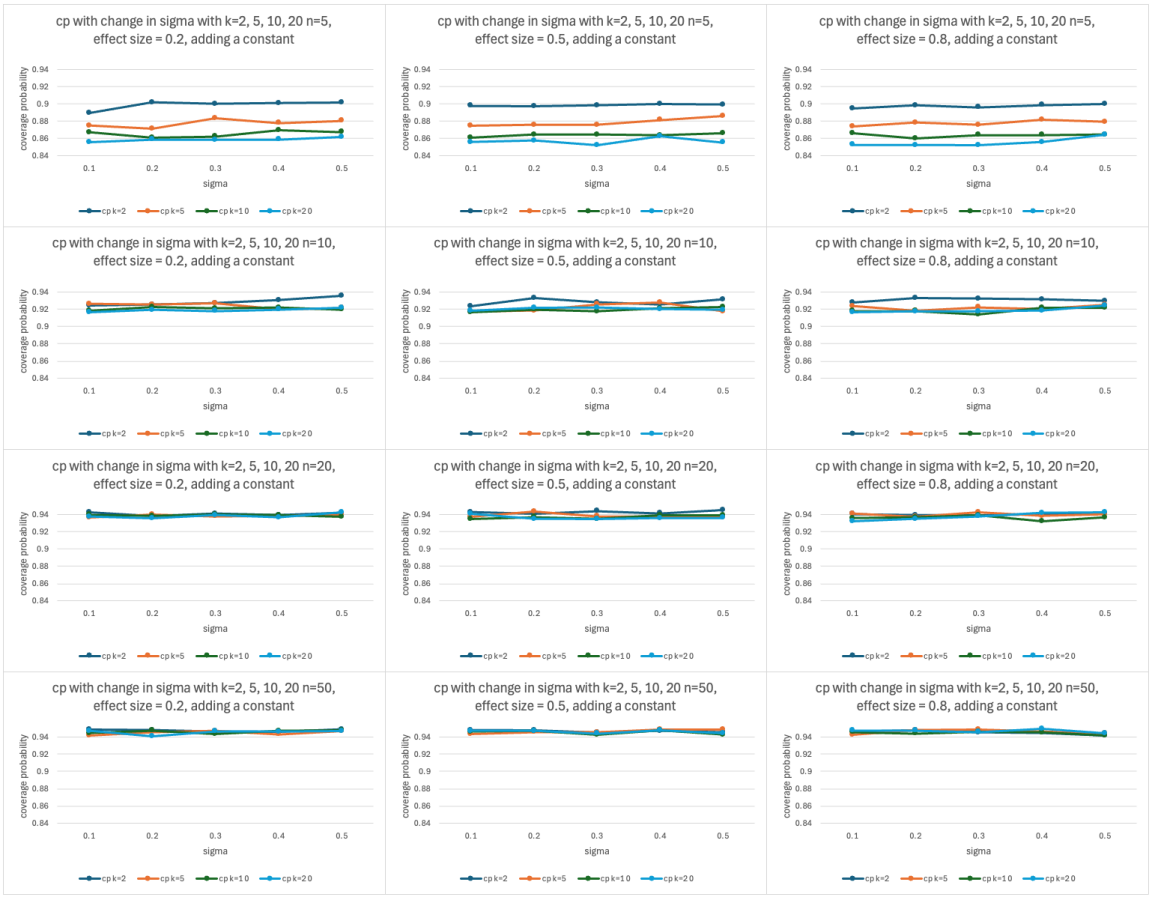

Figure 4.2.5 The coverage probability in the estimated treatment effect, as a function of σ and k across different effect sizes, with adding a constant to lognormal data using the mean difference for a fixed value of n at all levels of 5, 10, 20, and 50.

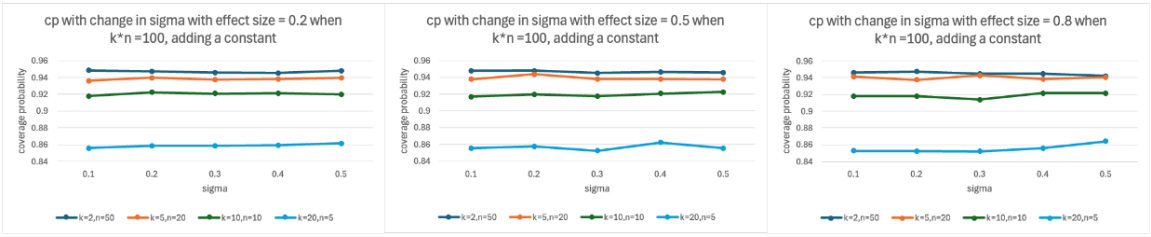

Figure 4.2.6 The coverage probability in the estimated treatment effect, as a function of σ and different combinations of k and n across different effect sizes, with adding a constant to lognormal data using the mean difference for a fixed total sample size of 200.

Figure 4.2.7 and Figure 4.2.8 indicate that the power increases as the number of studies, patients per study, and the effect size increases. With a fixed value of k, there is a slight

increase in power as skewness grows when n=5 with moderate and large effect sizes. With a fixed value of n, there is a slight increase in power as skewness increases with smaller k values. With a small effect size, the power exceeds 0.8 when the total sample size exceeds 800. With a moderate effect size, the power exceeds 0.8 when the total sample size exceeds 200. With a large effect size, the power exceeds 0.8 when the total sample size exceeds 50. With a fixed total sample size of 200, there is no detectable impact from varied combinations of k and n on power with moderate and large effect sizes, as shown in Figure 4.2.9.

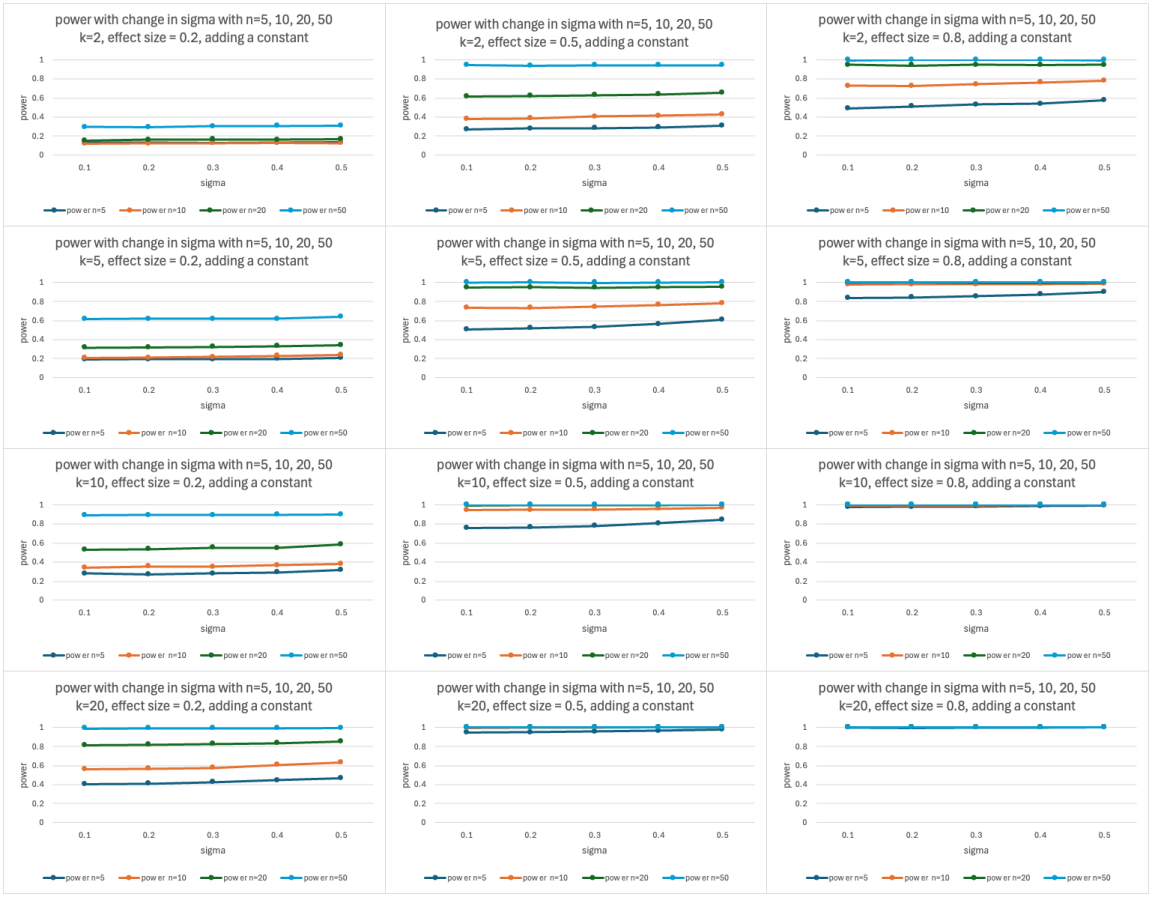

Figure 4.2.7 The power in the test, as a function of σ and n across different effect sizes, with adding a constant to lognormal data using the mean difference for a fixed value of k at all levels of 2, 5, 10, and 20.

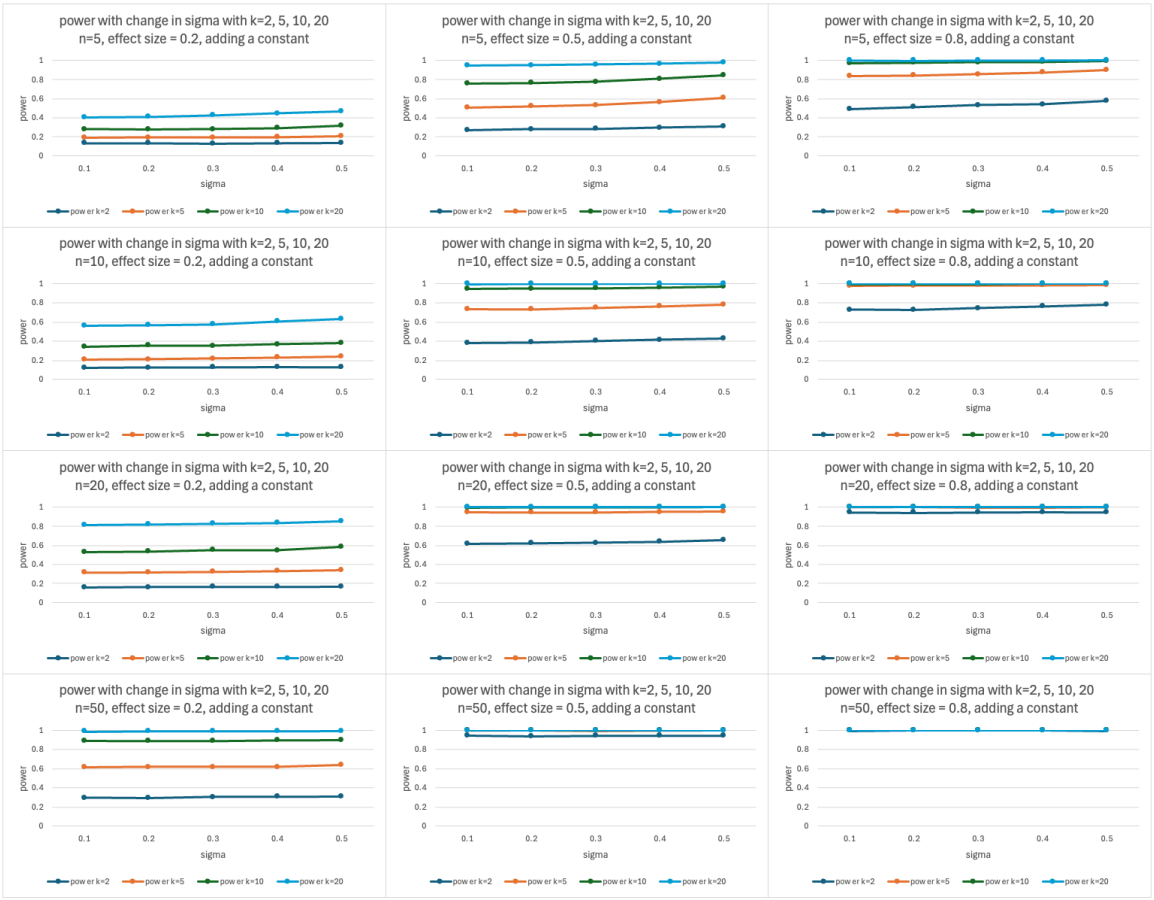

Figure 4.2.8 The power in the estimated treatment effect, as a function of σ and k across different effect sizes, with adding a constant to lognormal data using the mean difference for a fixed value of n at all levels of 5, 10, 20, and 50.

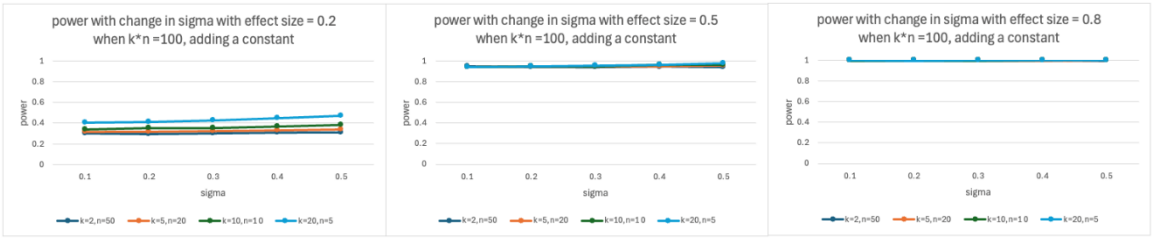

Figure 4.2.9 The power in the test, as a function of σ and different combinations of k and n across different effect sizes, with adding a constant to lognormal data using the mean difference for a fixed total sample size of 200.

If n=20, 50, the type I error rate is acceptable. The type I error rate are above 0.05 in all situations.

4.2.1.2 Both Lognormal with Different μs

With a fixed value of k at all levels, the relative bias moves away from zero as n decreases and skewness increases; moreover, the relative bias goes worse faster with smaller n (n=5, 10), shown in Figure 4.2.10. In Figure 4.2.11, with a fixed value of small n (n=5, 10), the relative bias is better when k is small. With other fixed values of n, the relative bias has a negligible change with k. Whenever the absolute value of relative bias is greater than 10%, n is 5, and skewness is the largest. Effect sizes have an unimportant impact on relative bias. With a fixed total sample size of 200, the relative bias is more robust around zero against skewness when k=2, n=50, and is the worst when k=20, n=5, as presented in Figure 4.2.12.

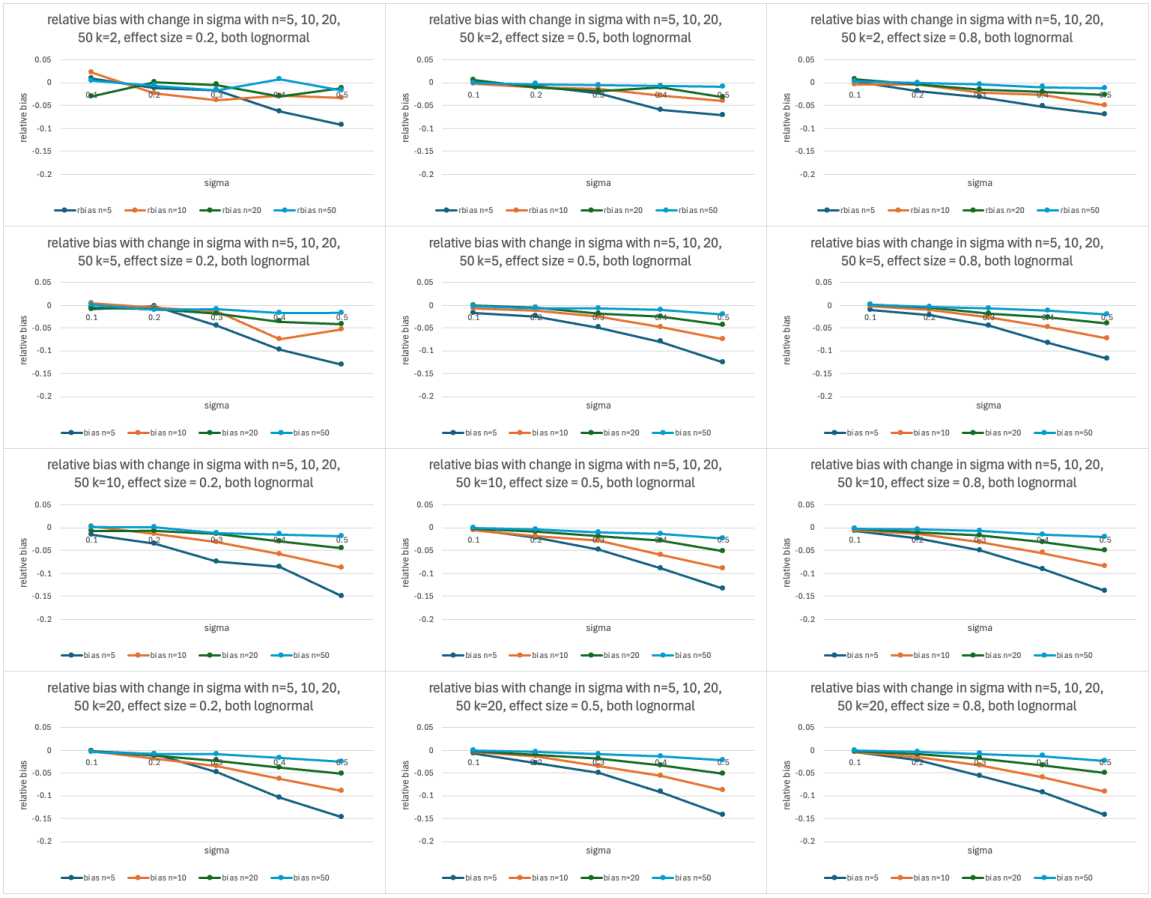

Figure 4.2.10 The relative bias in the estimated treatment effect, as a function of σ and n across different effect sizes, with simulating two lognormal distributions with different parameters using the mean difference for a fixed value of k at all levels of 2, 5, 10, and 20.

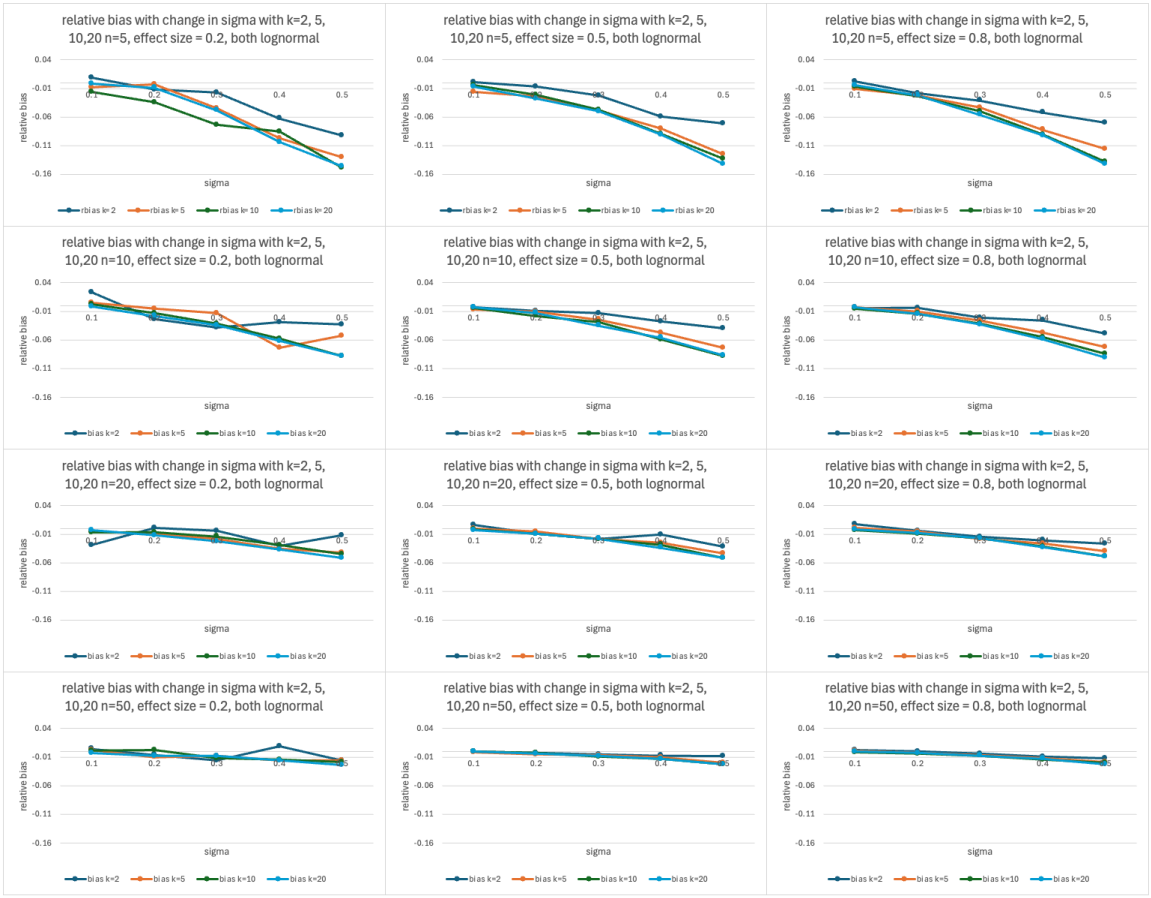

Figure 4.2.11 The relative bias in the estimated treatment effect, as a function of σ and k across different effect sizes, with simulating two lognormal distributions with different parameters using the mean difference for a fixed value of n at all levels of 5, 10, 20, and 50.

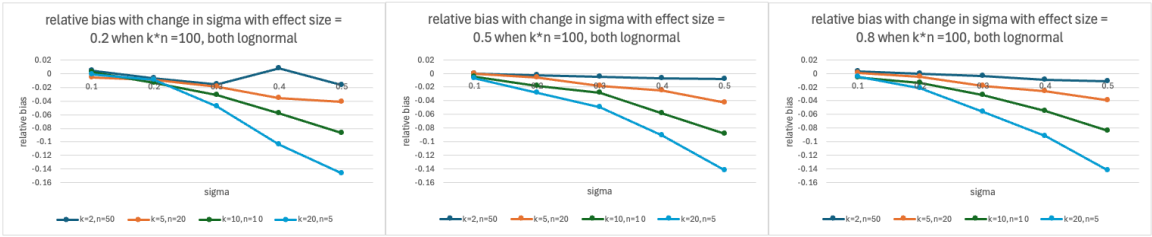

Figure 4.2.12 The relative bias in the estimated treatment effect, as a function of σ and different combinations of k and n across different effect sizes, with simulating two lognormal distributions with different parameters using the mean difference for a fixed total sample size of 200.

As shown in Figure 4.2.13, with a fixed value of k at all levels, the coverage probability increases as n increases. With a fixed value of n at 5, the coverage probability increases as k decreases. The same trend is observed with all n values at a large effect size (d=0.8), as presented in Figure 4.2.14. The effect of skewness on the coverage probability is minor for a small effect size. However, the coverage probability decreases as skewness increases with larger k, smaller n for moderate and large effect sizes (d=0.5, 0.8). When n=5, the coverage probability is below 0.925 in all cases. From Figure 4.2.15, with a fixed total sample size of 200, the coverage probability is closest to 0.95 when k=2 and n=50 and is below the acceptable range when k=20 and n=5.

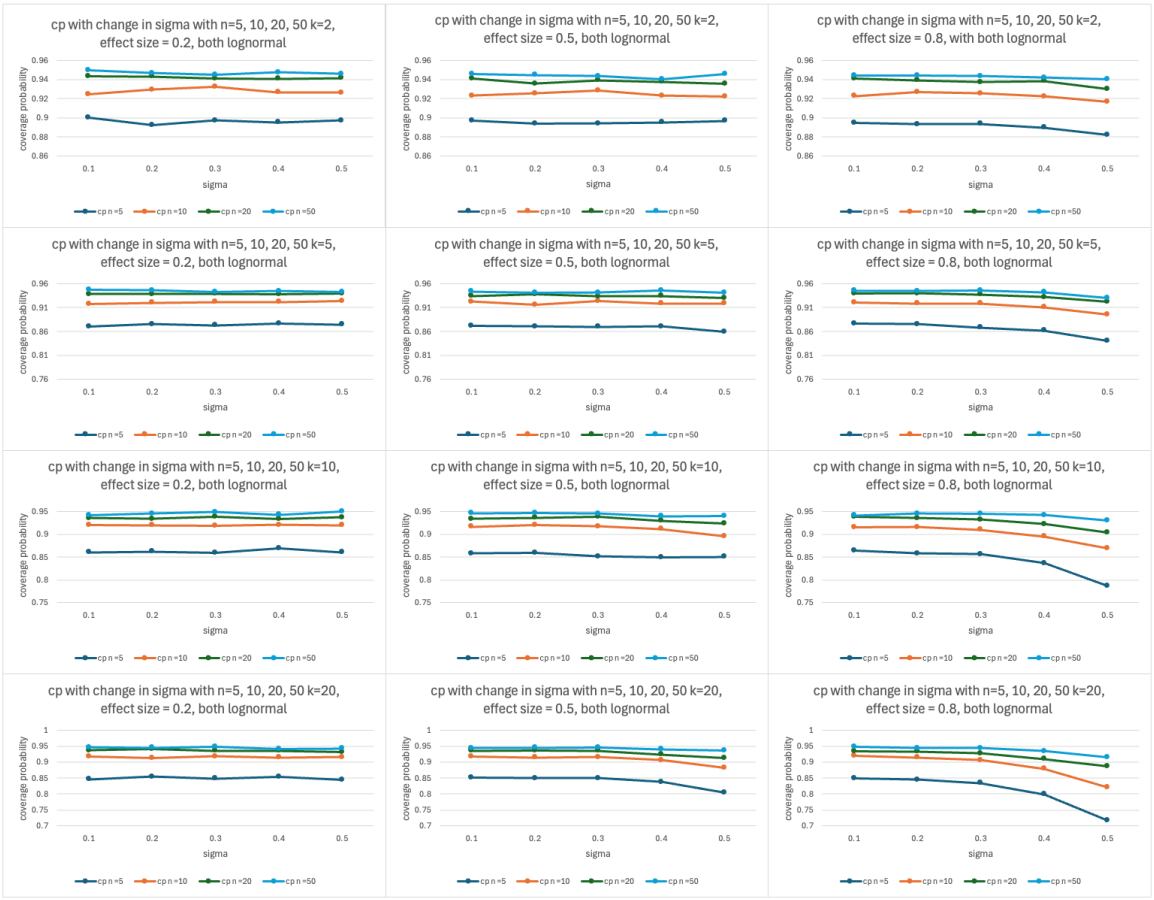

Figure 4.2.13 The coverage probability in the estimated treatment effect, as a function of σ and n across different effect sizes, with simulating two lognormal distributions with different parameters using the mean difference for a fixed value of k at all levels of 2, 5, 10, and 20.

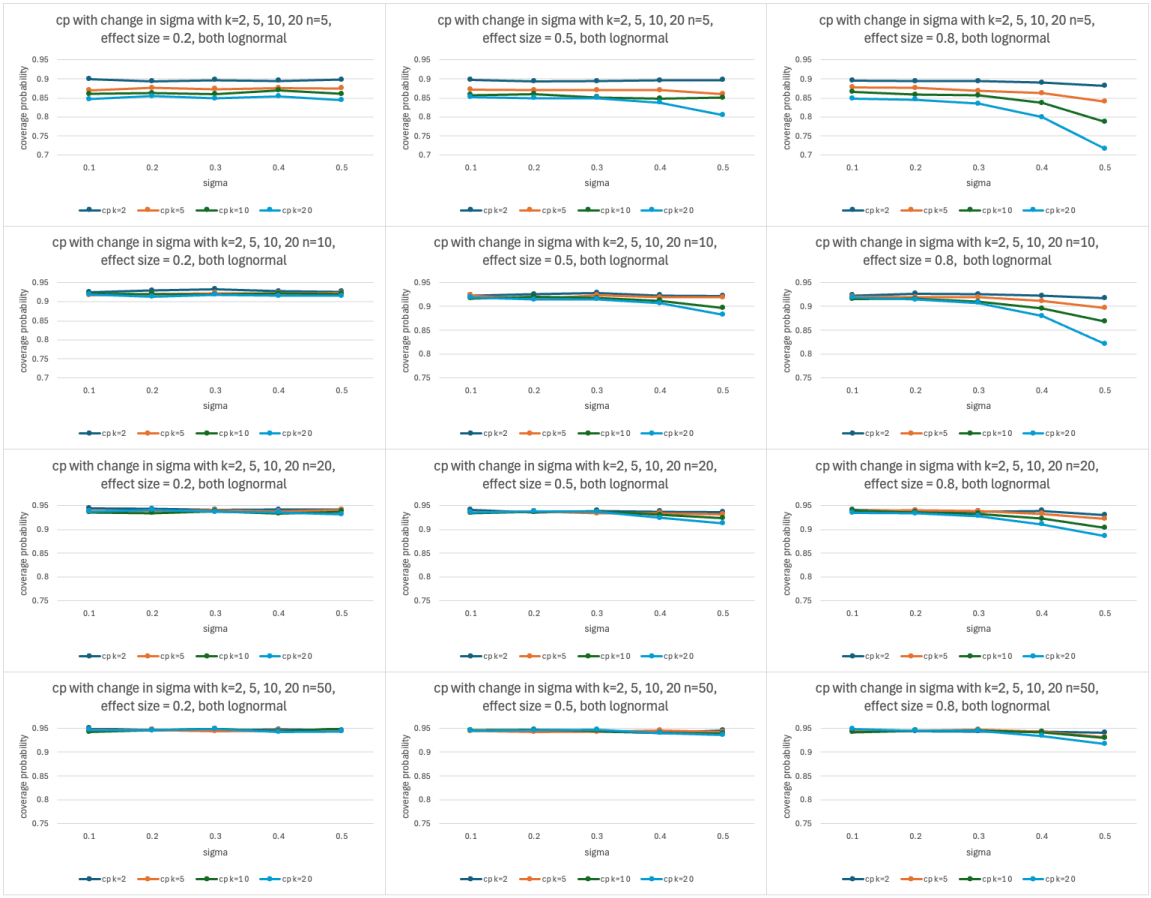

Figure 4.2.14 The coverage probabilities in the estimated treatment effect, as a function of σ and k across different effect sizes, with simulating two lognormal distributions with different parameters using the mean difference for a fixed value of n at all levels of 5, 10, 20, and 50.

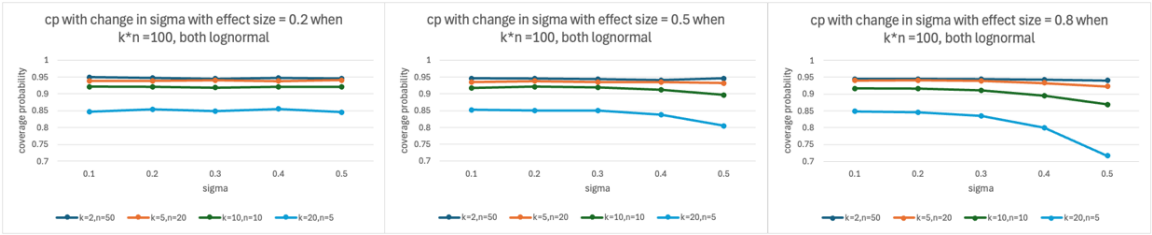

Figure 4.2.15 The coverage probability in the estimated treatment effect, as a function of σ and different combinations of k and n across different effect sizes, with simulating two lognormal distributions with different parameters using the mean difference for a fixed total sample size of 200.

Figure 4.2.16 and Figure 4.2.17 show that the power increases as the number of studies, patients per study, and the effect size increases. There is no important effect of skewness on the power. With a small effect size, the power exceeds 0.8 when the total sample size exceeds 1000. With a moderate effect size, the power exceeds 0.8 when the total sample size exceeds 200. With a large effect size, the power exceeds 0.8 when the total sample size exceeds 50. As illustrated in Figure 4.2.18, with a fixed total sample size of 200, varied combinations of k and n do not have an important impact on the power with moderate and large effect sizes.

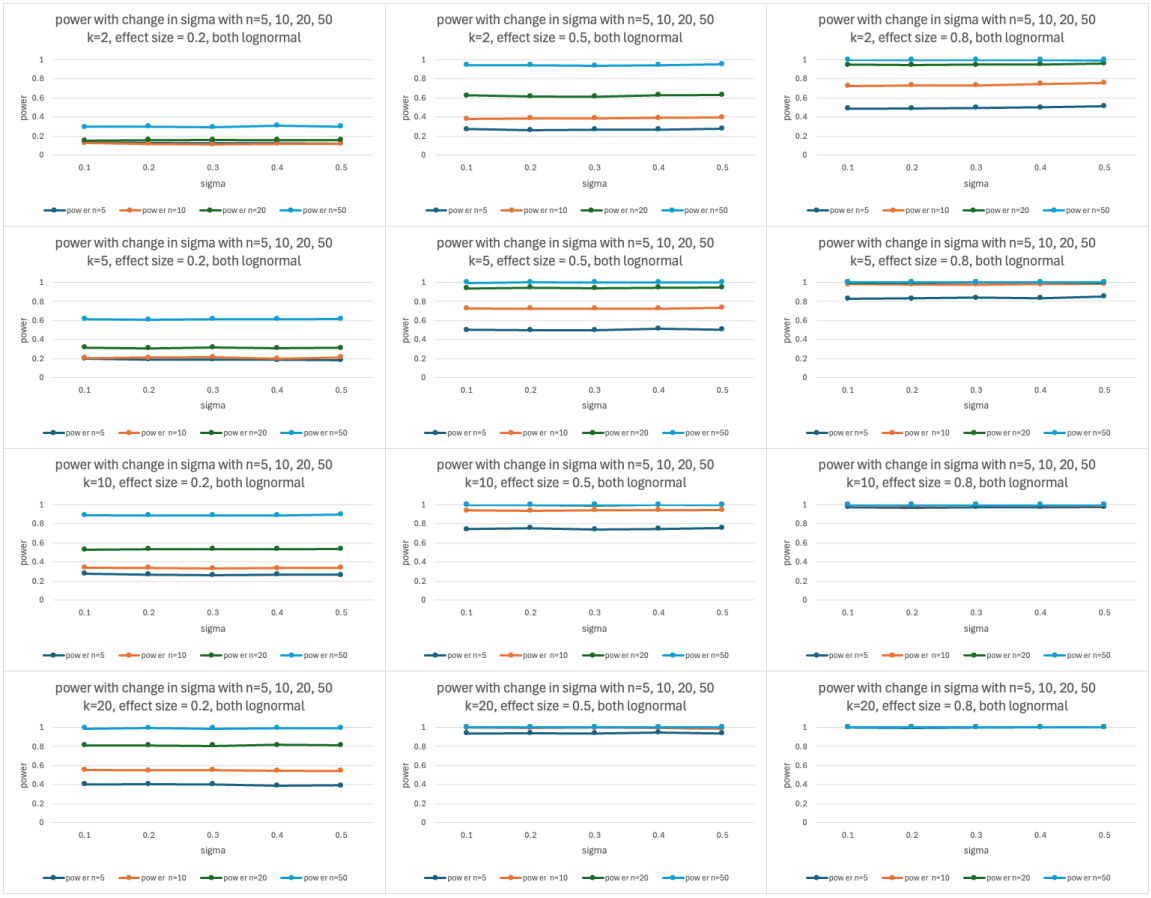

Figure 4.2.16 The power in the test, as a function of σ and n across different effect sizes, with simulating two lognormal distributions with different parameters using the mean difference for a fixed value of k at all levels of 2, 5, 10, and 20.

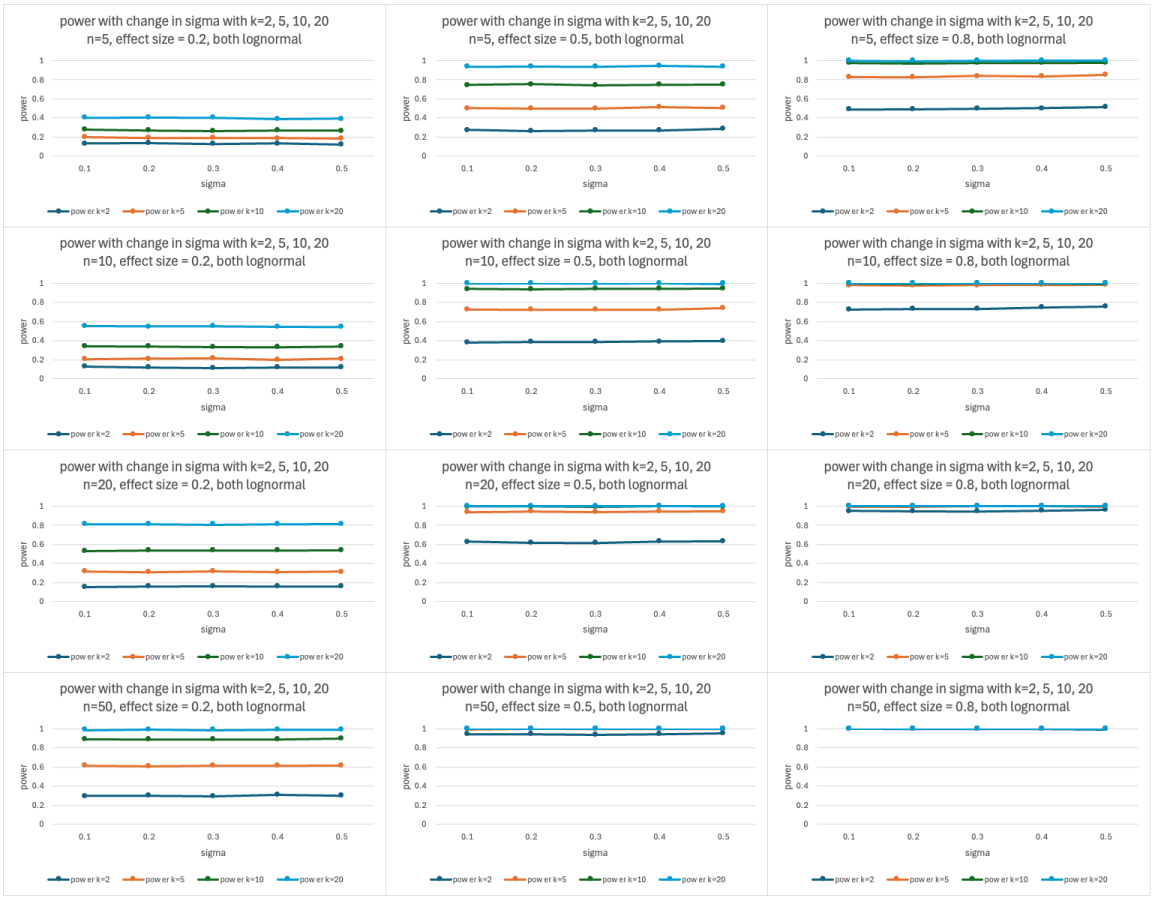

Figure 4.2.17 The power in the test, as a function of σ and k across different effect sizes, with simulating two lognormal distributions with different parameters using the mean difference for a fixed value of n at all levels of 5, 10, 20, and 50.

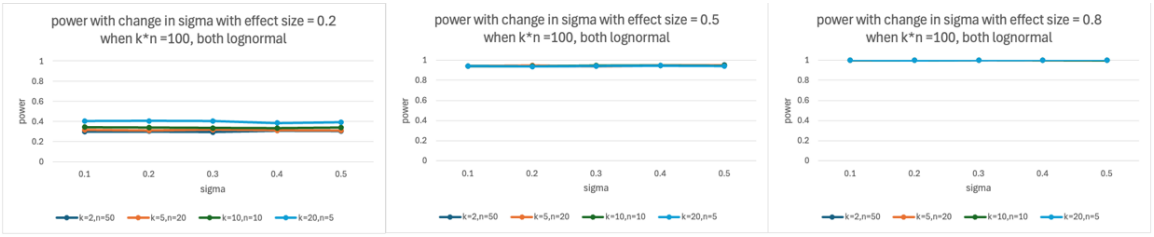

Figure 4.2.18 The power in the test, as a function of σ and different combinations of k and n across different effect sizes, with simulating two lognormal distributions with different parameters using the mean difference for a fixed total sample size of 200.

4.2.2 Standardized Mean Difference

4.2.2.1 Adding a Constant

From Figure 4.2.19, with a fixed value of k at all levels, the relative bias is closest to zero as n increases. With a fixed value of n at all levels, the relative bias is closest to zero as k decreases with skewness less than 1 ($\sigma = 0.1, 0.2, and\ 0.3$), as shown in Figure 4.2.20. The skewness has a greater effect on the relative bias with smaller n (n=5), and the relative bias moves closer to zero as the skewness increases. Effect sizes have an undetectable effect on the relative bias. There are only 4 cases where the absolute relative bias is more than 5%, when k=5 and n=5, with one case having a moderate effect with skewness of 0.3 ($\sigma = 0.1$), and three cases having a large effect with skewness of 0.3, 0.61 ($\sigma = 0.2$), and 0.95 ($\sigma = 0.3$) respectively. Most relative bias is negative, indicating an underestimation of the estimated effect. As presented in Figure 4.2.21, with a fixed total sample size of 200, the relative bias is the farthest from 0 when k=20 and n=5.

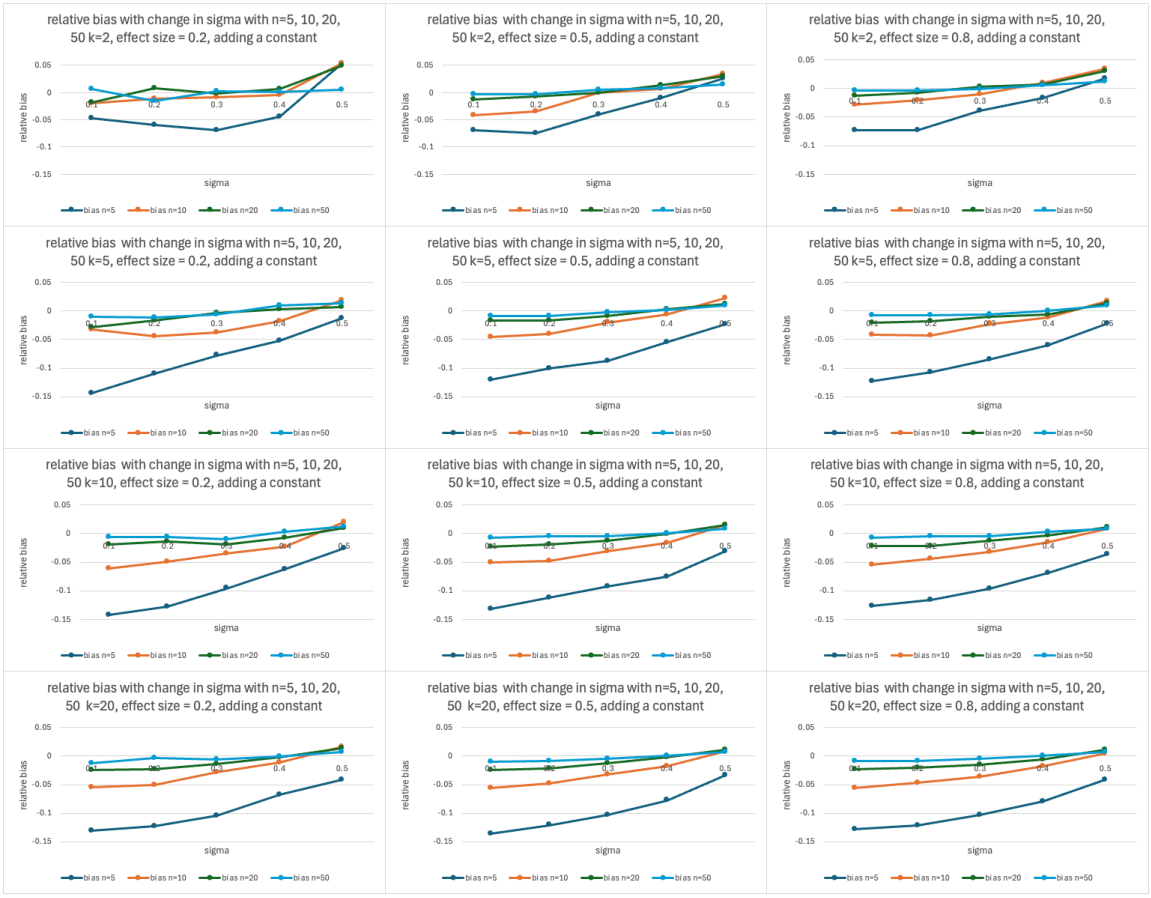

Figure 4.2.19 The relative bias in the estimated treatment effect, as a function of σ and n across different effect sizes, with adding a constant to lognormal data using the standardized mean difference for a fixed value of k at all levels of 2, 5, 10, and 20.

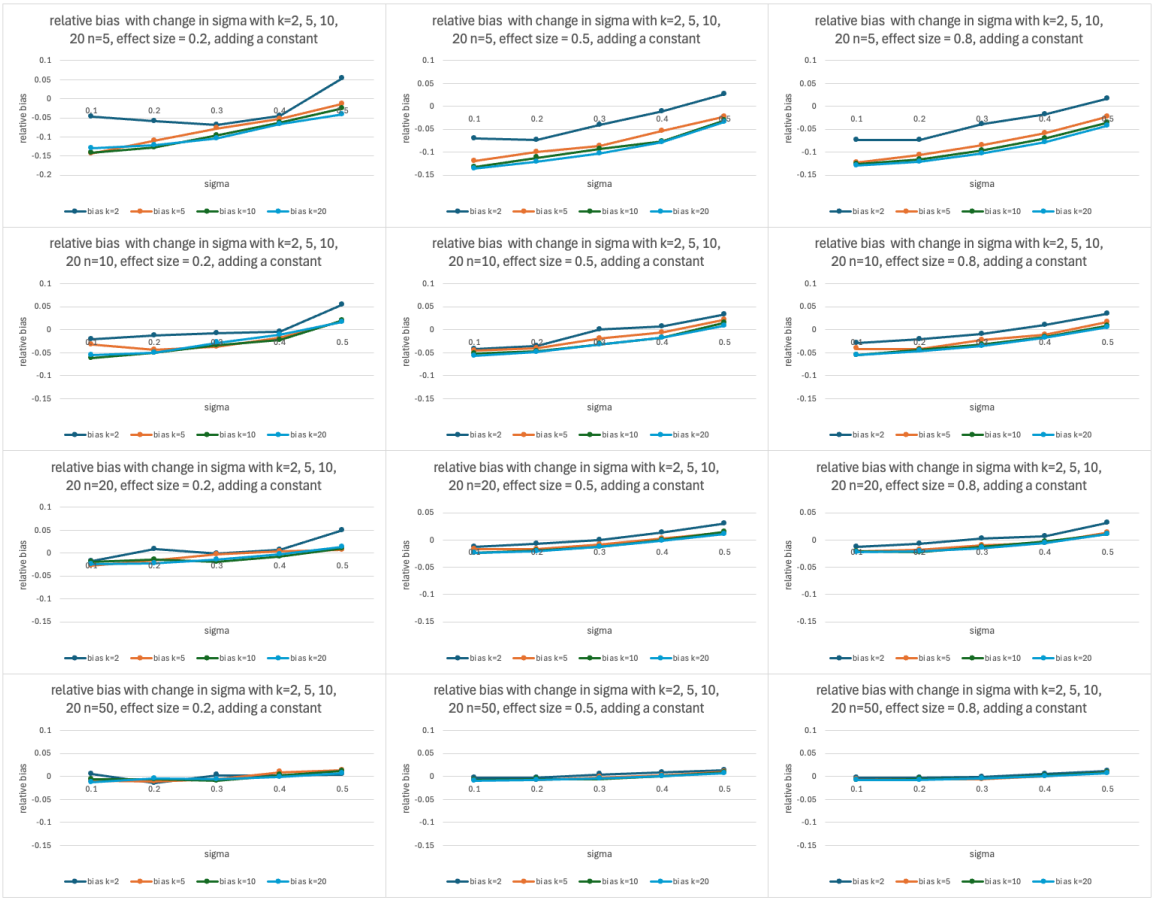

Figure 4.2.20 The relative bias in the estimated treatment effect, as a function of σ and k across different effect sizes, with adding a constant to lognormal data using the standardized mean difference for a fixed value of n at all levels of 5, 10, 20, and 50.

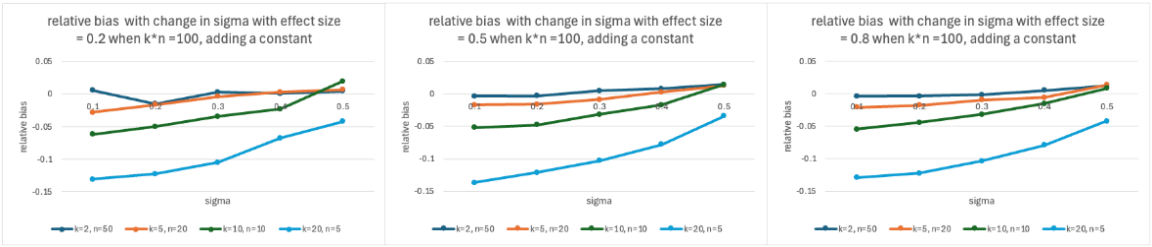

Figure 4.2.21 The relative bias in the estimated treatment effect, as a function of σ and different combinations of k and n across different effect sizes, with adding a constant to lognormal data using the standardized mean difference for a fixed total sample size of 200.

With a fixed value of k at all levels, the coverage probability is closest to 0.95 with larger n and smaller skewness, as shown in Figure 4.2.22. The skewness does not noticeably

affect the coverage probability when the effect size is small (d=0.2). For moderate and large effect sizes (d=0.5, 0.8), the coverage probability increases as skewness increases when n = 5, but it decreases as skewness increases for other n values. As presented in Figure 4.2.23, with a fixed value of small n, a larger k results in a less stable coverage probability as skewness changes. Conversely, with a fixed value of large n, the coverage probability remains stable across all k values as skewness increases at a small effect size. When n = 20 and 50, the coverage probability decreases from 0.95 as skewness increases for moderate and large effect sizes. Whenever the coverage probability is outside the acceptable range, n = 5, and d = 0.2. Furthermore, the coverage probability does not drop below 0.925 when it falls outside the acceptable range. From Figure 4.2.24, with a fixed total sample size of 200, the coverage probability is closest to 0.95 when k = 2 and n = 50, given small and moderate effect sizes. Additionally, when k = 10 and n = 10, the coverage probability is robust against changes in skewness for a large effect size.

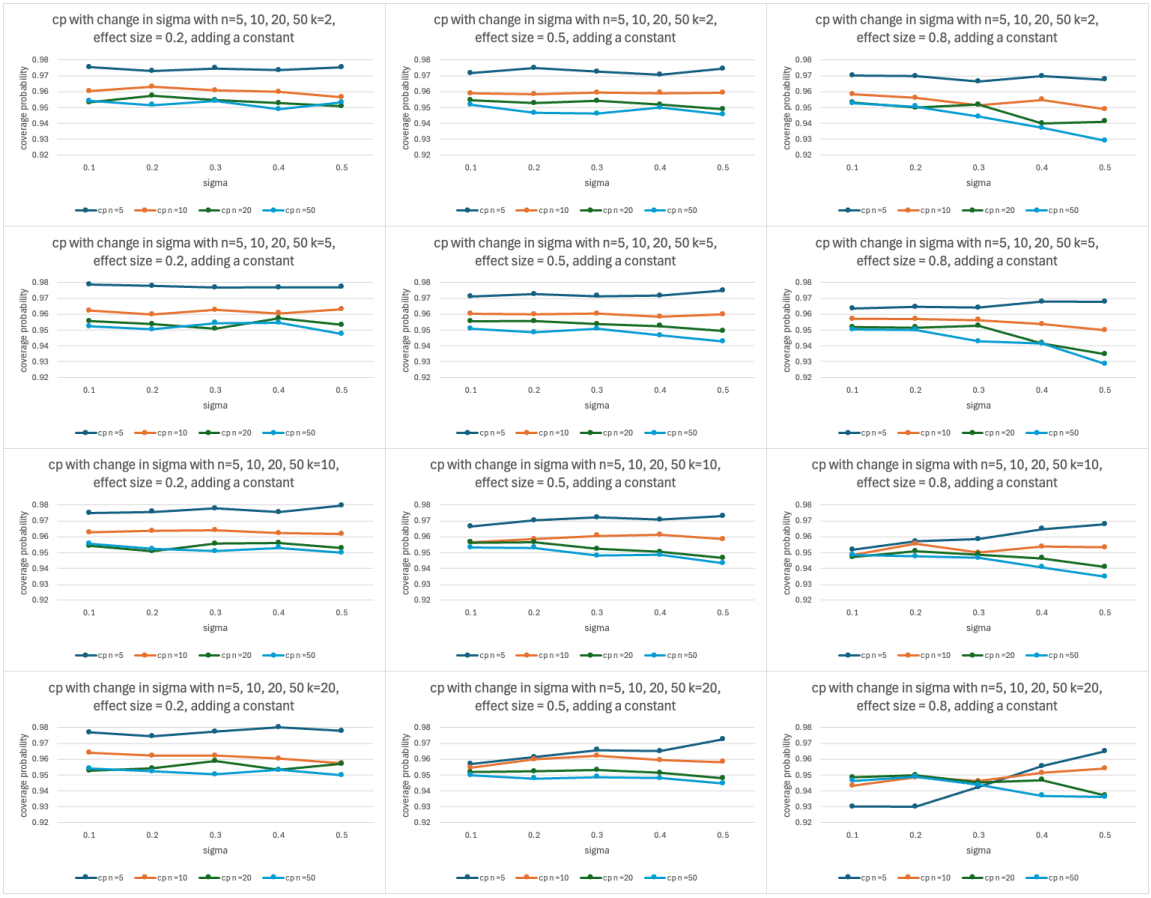

Figure 4.2.22 The coverage probability in the estimated treatment effect, as a function of σ and n across different effect sizes, with adding a constant to lognormal data using the standardized mean difference for a fixed value of k at all levels of 2, 5, 10, and 20.

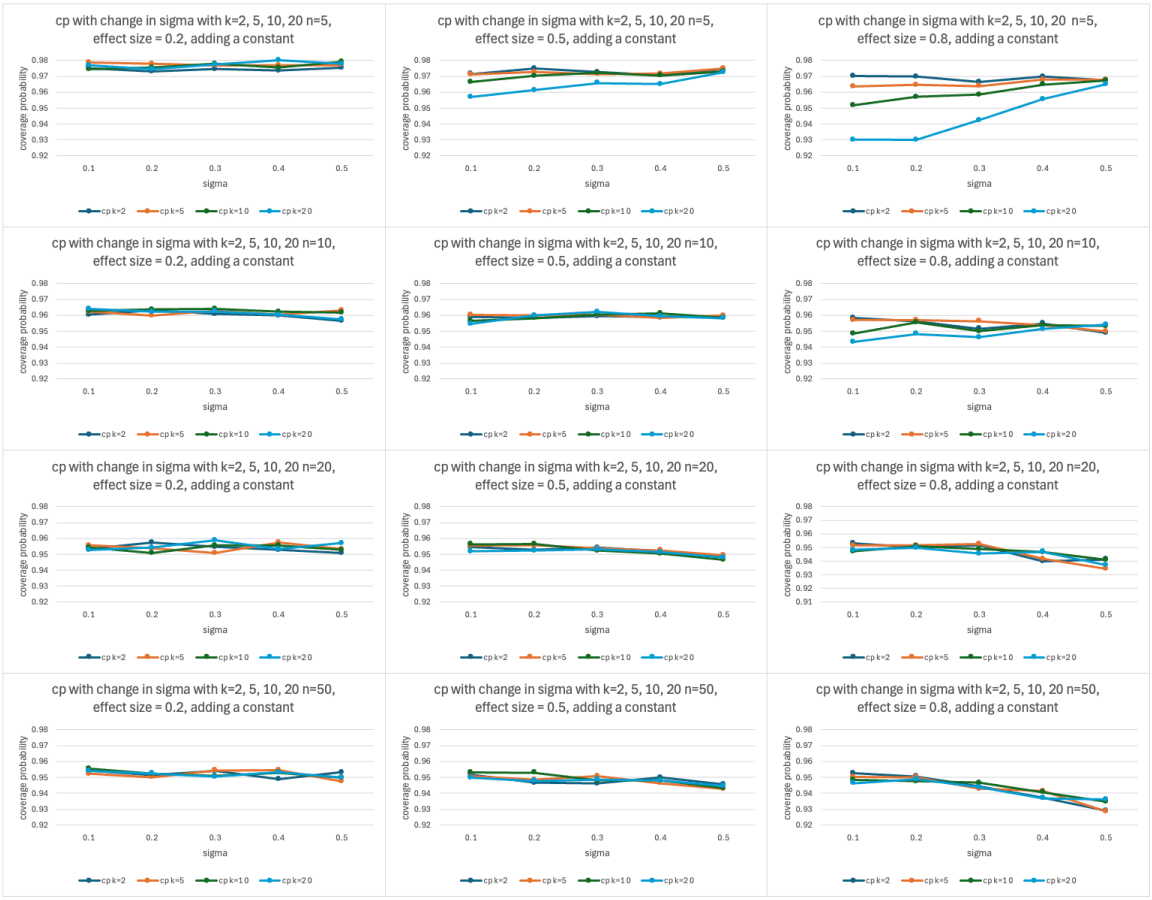

Figure 4.2.23 The coverage probability in the estimated treatment effect, as a function of σ and k across different effect sizes, with adding a constant to lognormal data using the standardized mean difference for a fixed value of n at all levels of 5, 10, 20, and 50.

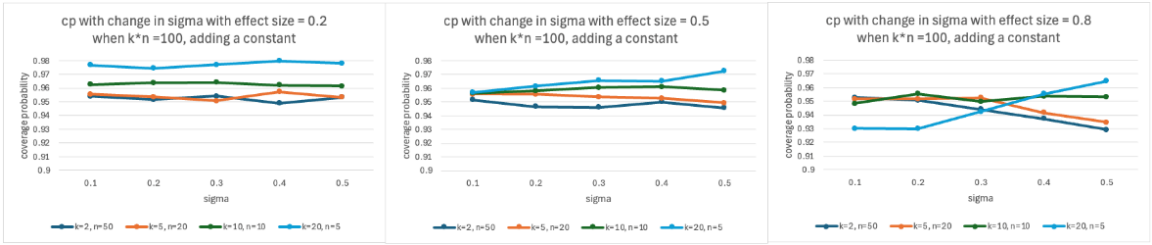

Figure 4.2.24 The coverage probability in the estimated treatment effect, as a function of σ and different combinations of k and n across different effect sizes, with adding a constant to lognormal data using the standardized mean difference for a fixed total sample size of 200.

Figure 4.2.25 and Figure 4.2.26 indicate that the power increases with the number of studies, patients per study, and effect size. Skewness has a negligible effect on the power.

For a small effect size, the power exceeds 0.8 when the total sample size is 800 or more. For a moderate effect size, the power exceeds 0.8 when the total sample size is 200 or more. For a large effect size, the power exceeds 0.8 when the total sample size exceeds 80.

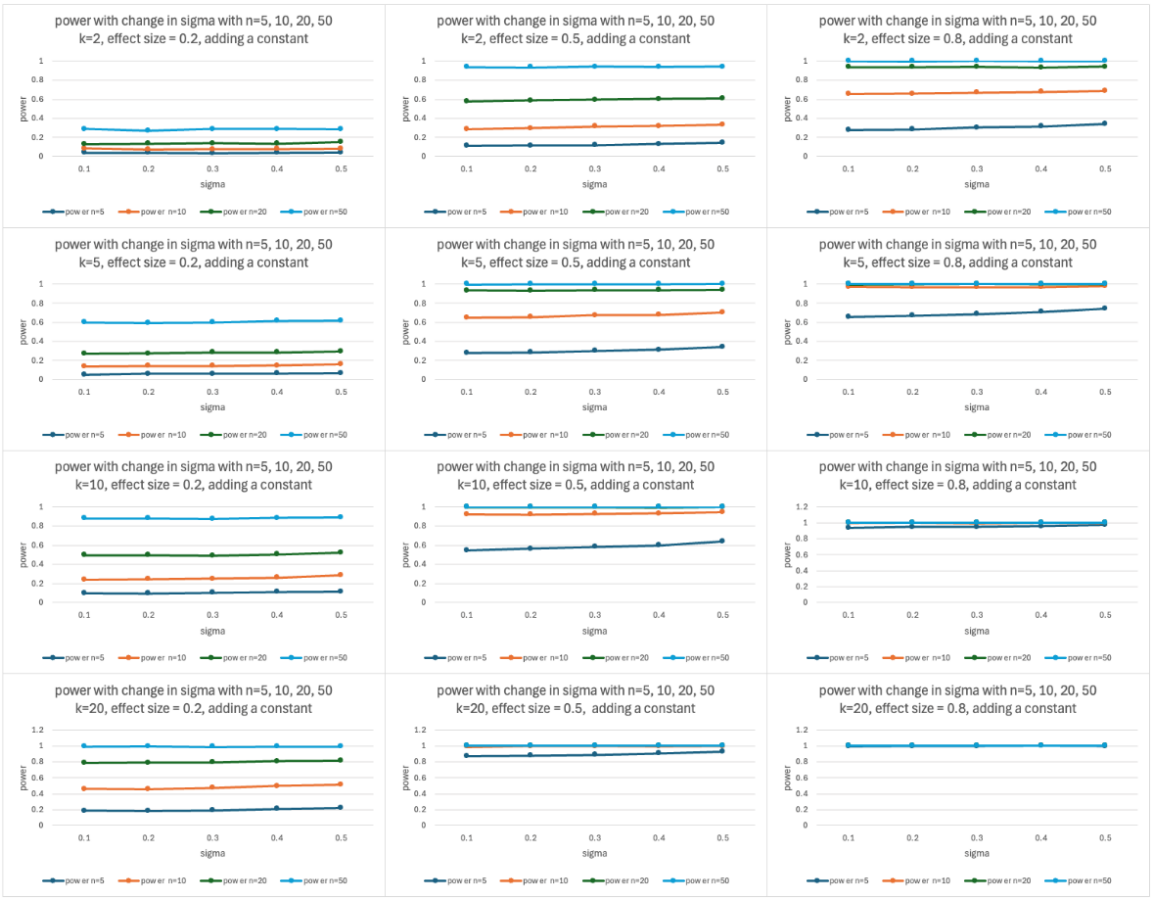

Figure 4.2.25 The power in the test, as a function of σ and n across different effect sizes, with adding a constant to lognormal data using the standardized mean difference for a fixed value of k at all levels of 2, 5, 10, and 20.

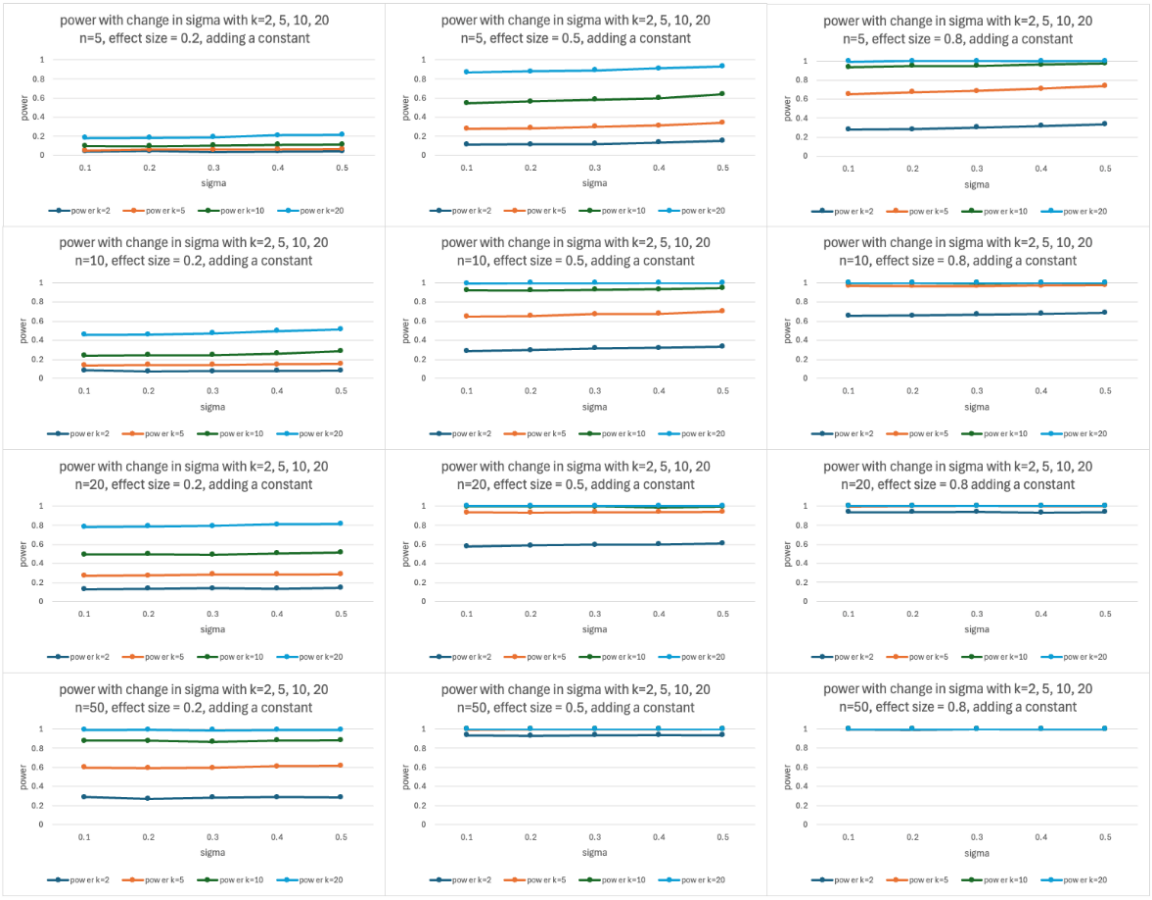

Figure 4.2.26 The power in the test, as a function of σ and k across different effect sizes, with adding a constant to lognormal data using the standardized mean difference for a fixed value of n at all levels of 5, 10, 20, and 50.

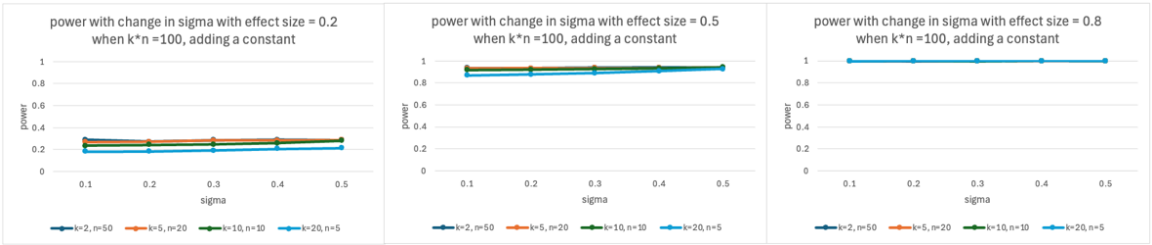

Figure 4.2.27 The power in the test, as a function of σ and different combinations of k and n across different effect sizes, with adding a constant to lognormal data using the standardized mean difference for a fixed total sample size of 200.

When using SMD, if k = 2 and n = 5 and 10, the type I error rate is above 0.075, and if k = 5, 10, 20, and n = 5, it is below 0.025. Skewness has no undetectable effect on the type I error rate.

4.2.2.2 Both Lognormal with Different μs

From Figure 4.2.28, with a fixed value of k at all levels, a larger n gives better relative bias. With a fixed value of n at all levels, a smaller k gives better relative bias, as shown in Figure 4.1.29. The skewness does not have a detectable effect on the relative bias with a small effect size and larger n (n=20, 50). There is a negligible trend of relative bias with effect sizes. Whenever the absolute relative bias is greater than 10%, n is 5. Moreover, most cases give a negative relative bias. With a fixed total sample size of 200, the relative bias is closest to zero when k=2 and n=50 and is the farthest from zero when k=20 and n=5, as presented in Figure 4.2.30.

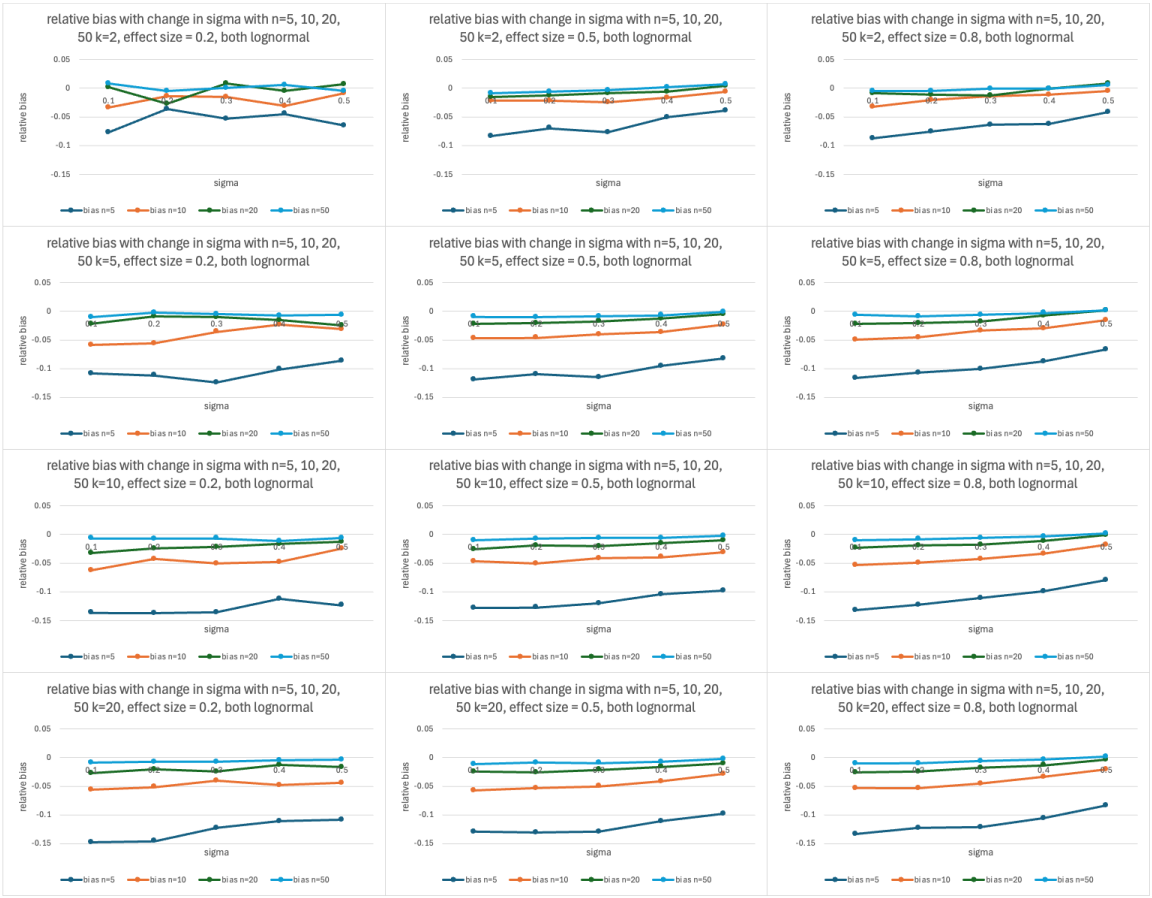

Figure 4.2.28 The relative bias in the estimated treatment effect, as a function of σ and n across different effect sizes, with simulating two lognormal distributions with different parameters using the standardized mean difference for a fixed value of k at all levels of 2, 5, 10, and 20.

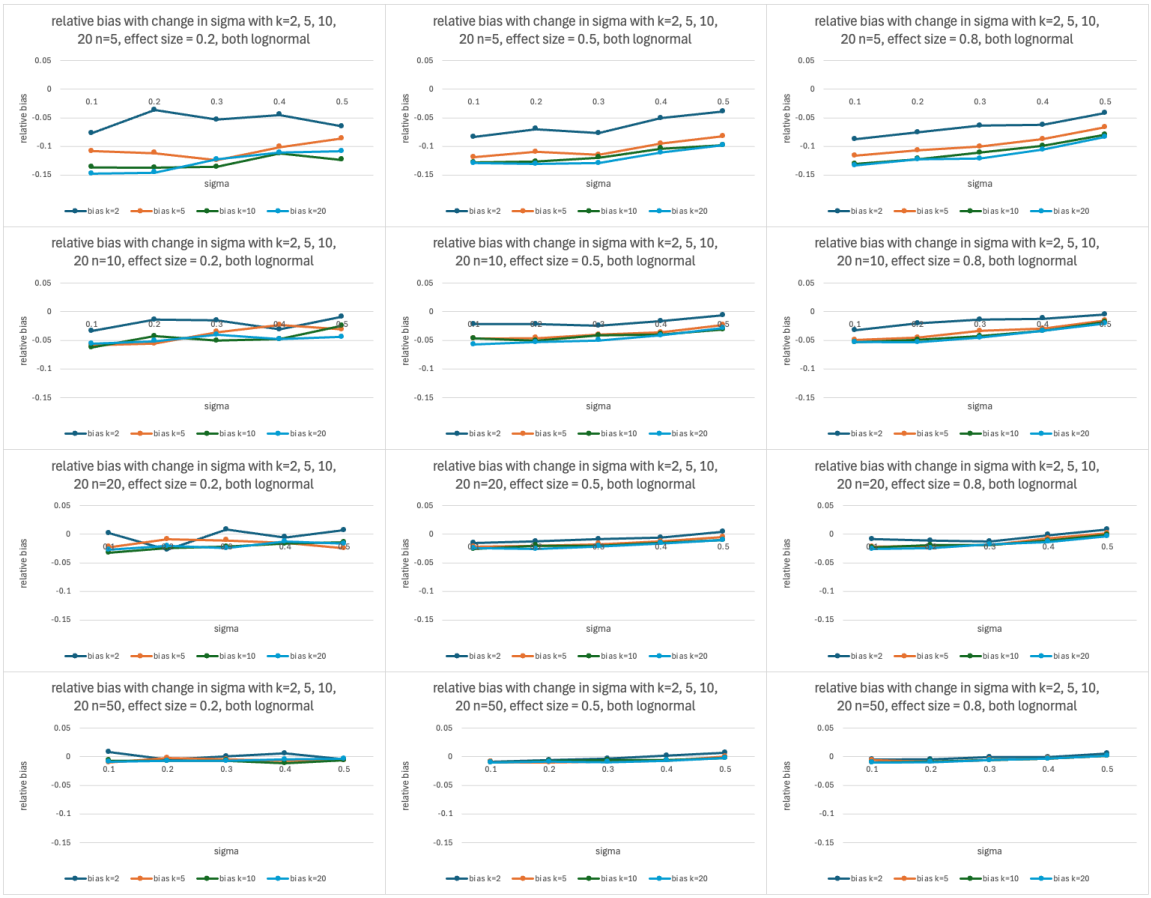

Figure 4.2.29 The relative bias in the estimated treatment effect, as a function of σ and k across different effect sizes, with simulating two lognormal distributions with different parameters using the standardized mean difference for a fixed value of n at all levels of 5, 10, 20, and 50.

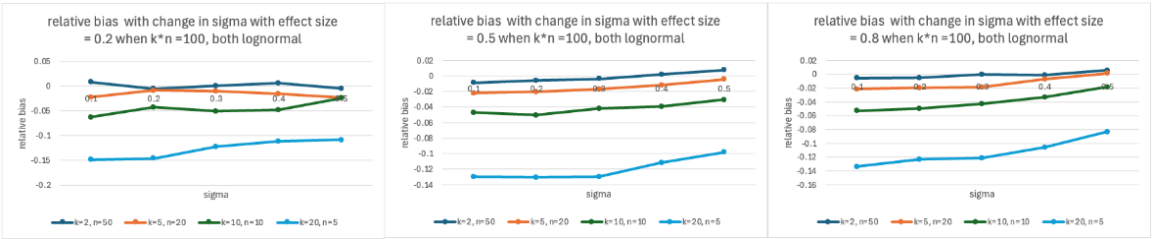

Figure 4.2.30 The relative bias in the estimated treatment effect, as a function of σ and different combinations of k and n across different effect sizes, with simulating two lognormal distributions with different parameters using the standardized mean difference for a fixed value of n at all levels of 200.

From Figure 4.2.31, with a fixed value of k at all levels, the coverage probability is around 0.95 with large n. Moreover, it increases from 0.95 as n decreases and skewness increases, especially for a large effect size. From Figure 4.2.32, with a fixed value of n at all levels, the larger k results in a better coverage probability and more stable against skewness for small and moderate effect sizes (d=0.2, 0.5). The coverage probability increases from 0.95 for a large effect size as skewness increases when n = 20 and 50. Whenever the coverage probability is above 0.975 or below 0.925, n is 5. The coverage probability is usually above 0.975 if it falls outside of the acceptable range. Figure 4.2.33 indicates that the coverage probability is ideal for a fixed total sample size of 200 when more patients per study and fewer studies are included.

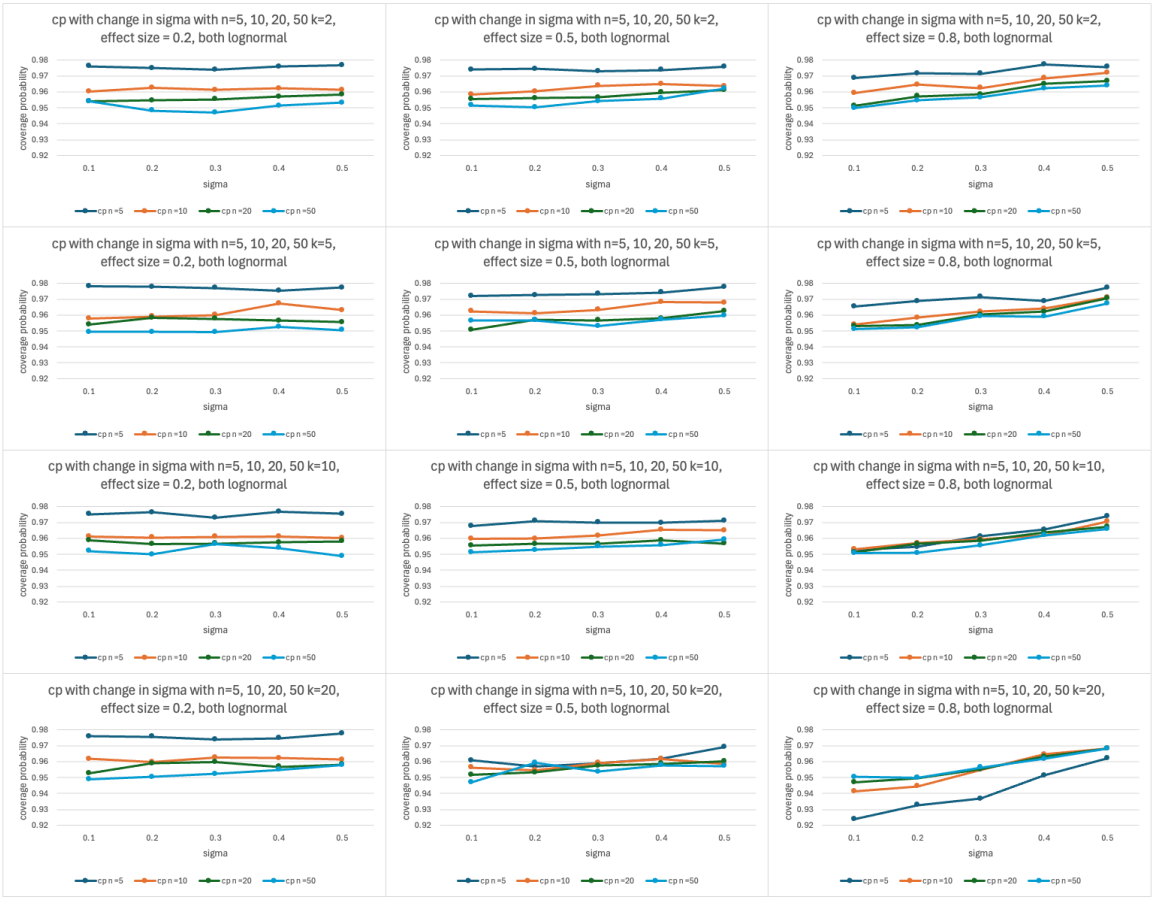

Figure 4.2.31 The coverage probability in the estimated treatment effect, as a function of σ and n across different effect sizes, with simulating two lognormal distributions with different parameters using the standardized mean difference for a fixed value of k at all levels of 2, 5, 10, and 20.

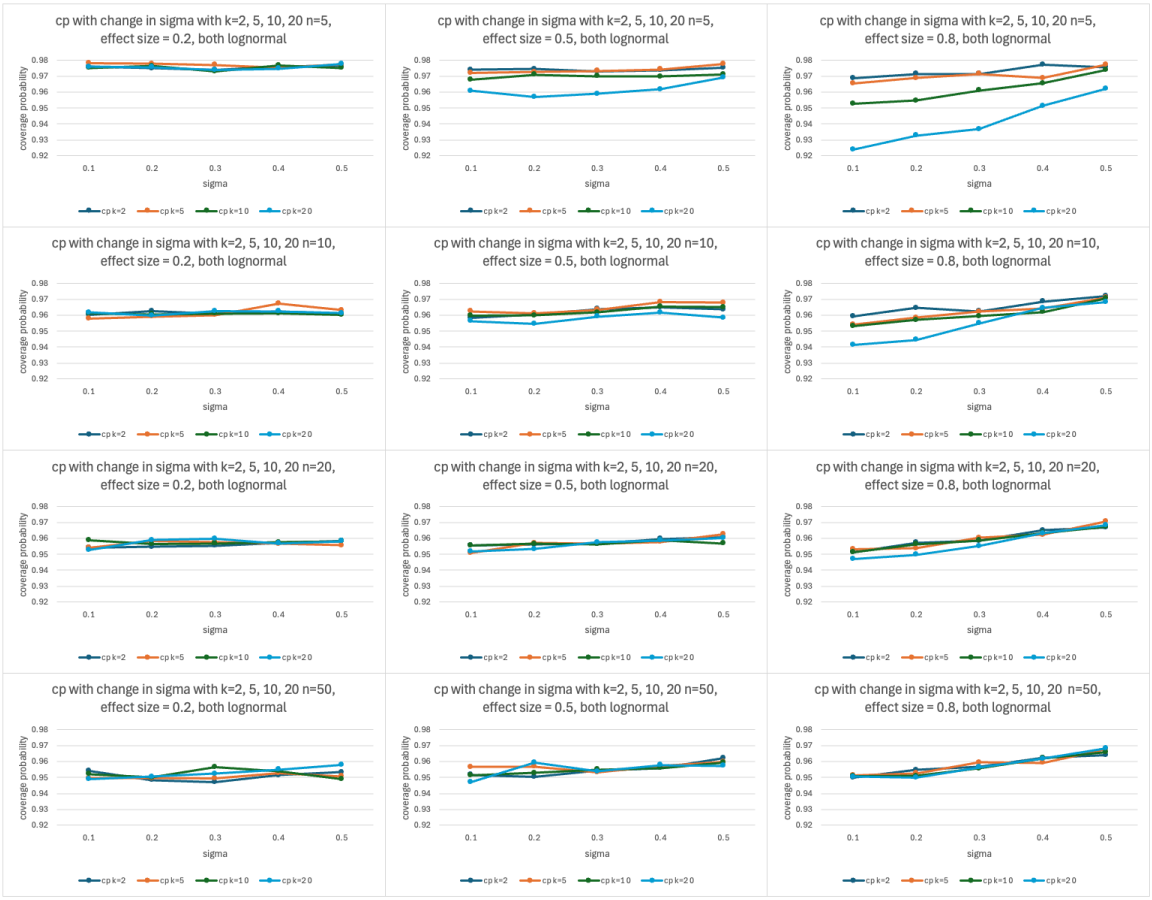

Figure 4.2.32 The coverage probability in the estimated treatment effect, as a function of σ and k across different effect sizes, with simulating two lognormal distributions with different parameters using the standardized mean difference for a fixed value of n at all levels of 5, 10, 20, and 50.

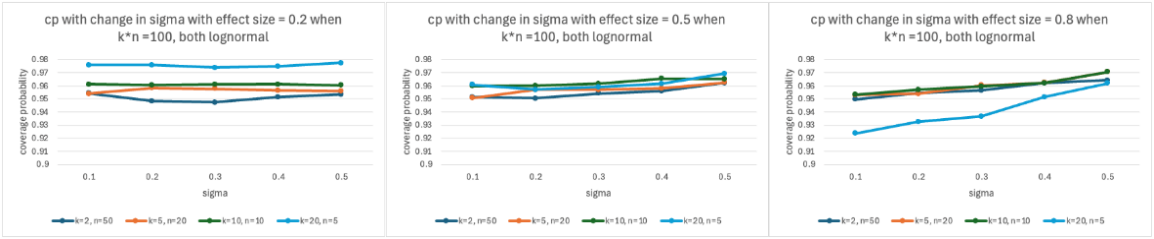

Figure 4.2.33 The coverage probability in the estimated treatment effect, as a function of σ and different combinations of k and n across different effect sizes, with simulating two lognormal distributions with different parameters using the standardized mean difference for a fixed total sample size of 200.

As shown in Figure 4.1.34 and Figure 4.1.35, the power increases as the number of studies, patients per study, and effect size increase. Skewness has a negligible effect on the power. For a small effect size, the power exceeds 0.8 when the total sample size is 1000 or more. For a moderate effect size, the power exceeds 0.8 when the total sample size is 200 or more. For a large effect size, the power exceeds 0.8 when the total sample size exceeds 80.

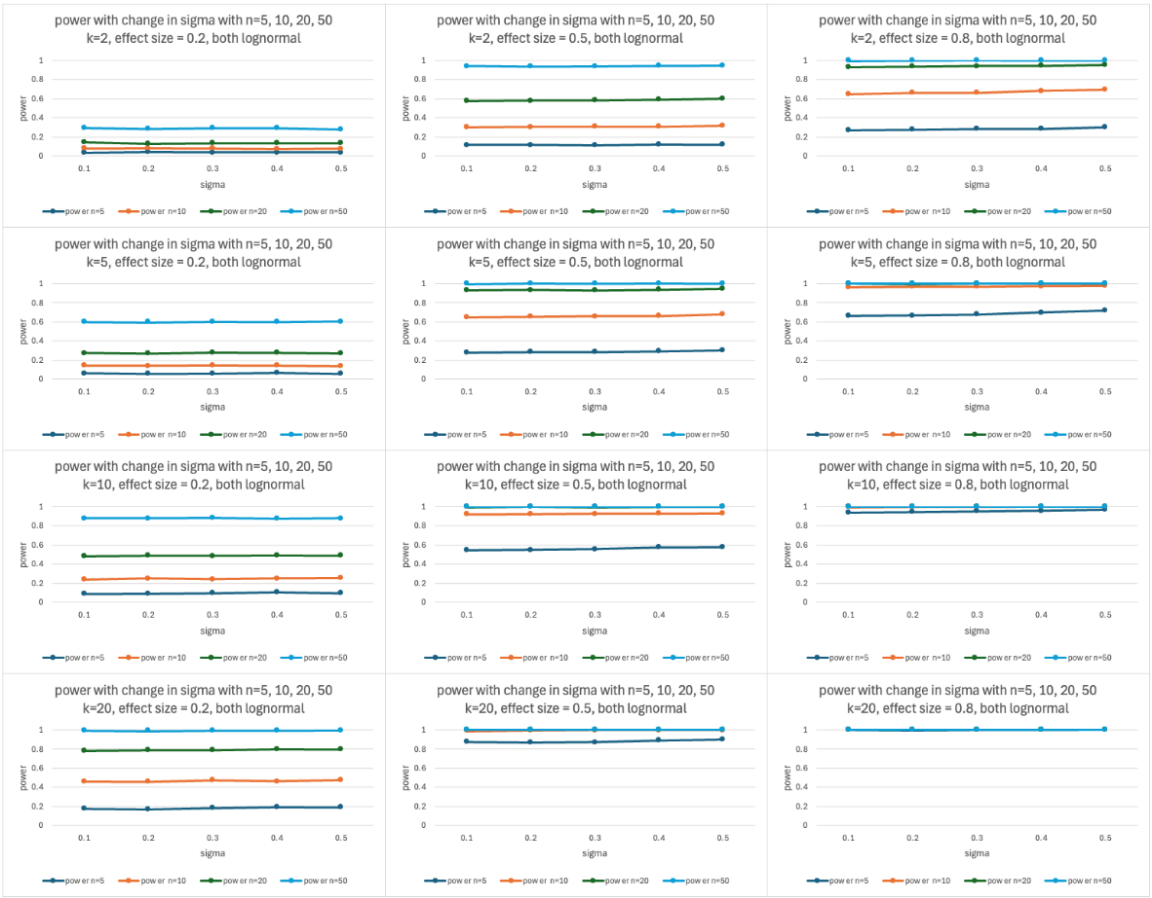

Figure 4.2.34 The power in the test, as a function of σ and n across different effect sizes, with simulating two lognormal distributions with different parameters using the standardized mean difference for a fixed value of k at all levels. of 2, 5, 10, and 20

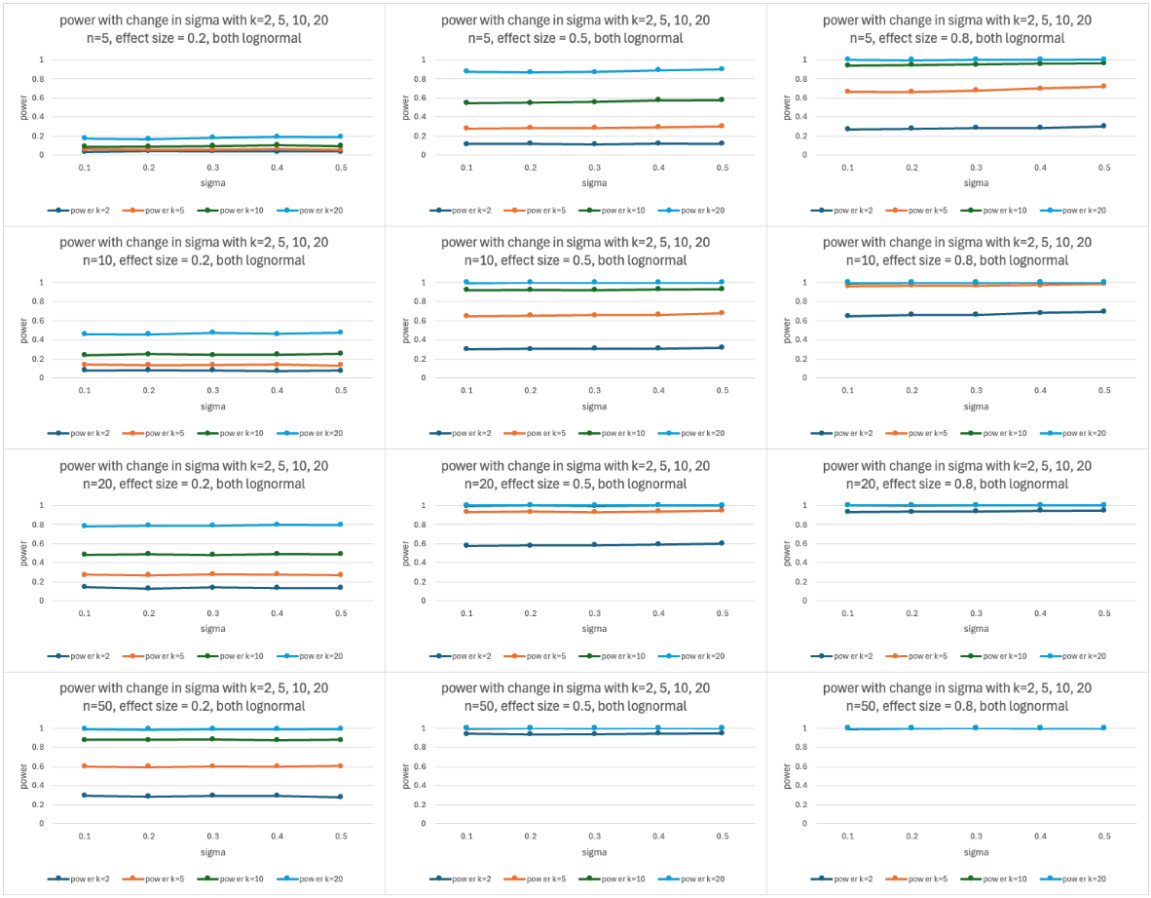

Figure 4.2.35 The power in the test, as a function of σ and k across different effect sizes, with simulating two lognormal distributions with different parameters using the standardized mean difference for a fixed value of n at all levels of 5, 10, 20, and 50.

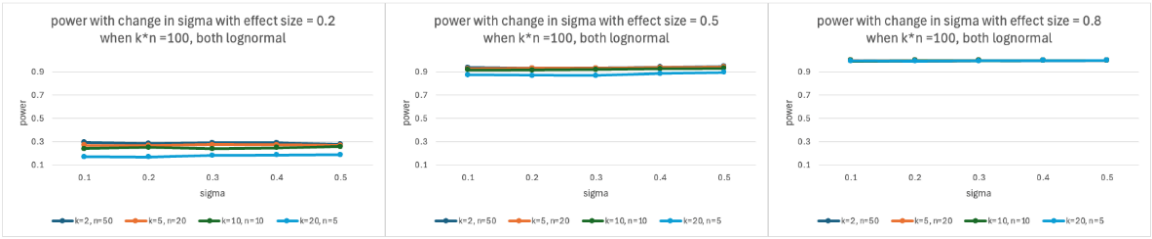

Figure 4.2.36 The power in the test, as a function of σ and different combinations of k and n across different effect sizes, with simulating two lognormal distributions with different parameters using the standardized mean difference for a fixed total sample size of 200.

4.3 Gamma Distribution

4.3.1 Mean Difference

4.3.1.1 Adding a Constant

From Figure 4.3.1, with a fixed value of k at 2 and 5, the absolute relative bias is slightly greater when n=5 and 10 with a small effect size. The relative bias becomes smaller and more stable when k=10, 20 and the effect size increases. From Figure 4.3.2, with a fixed value of n = 5 and 10, the absolute relative bias is larger when k=2 and shows greater fluctuation with the change in skewness. When n=20, 50 and the effect size increases, there is a negligible effect of k on the relative bias. From Figure 4.3.3, with a fixed total sample size of 200, the absolute relative bias is below 2% and is slightly smaller with moderate and large effect sizes. There is an undetectable effect of different combinations of k and n and skewness on the relative bias for a fixed total sample size of 200. When the effect size is 0.8, the relative bias is below 1% for all cases. When two identical gamma distributions are simulated and a constant is added to the treatment group, the skewness has an undetectable effect on the relative bias for most cases.

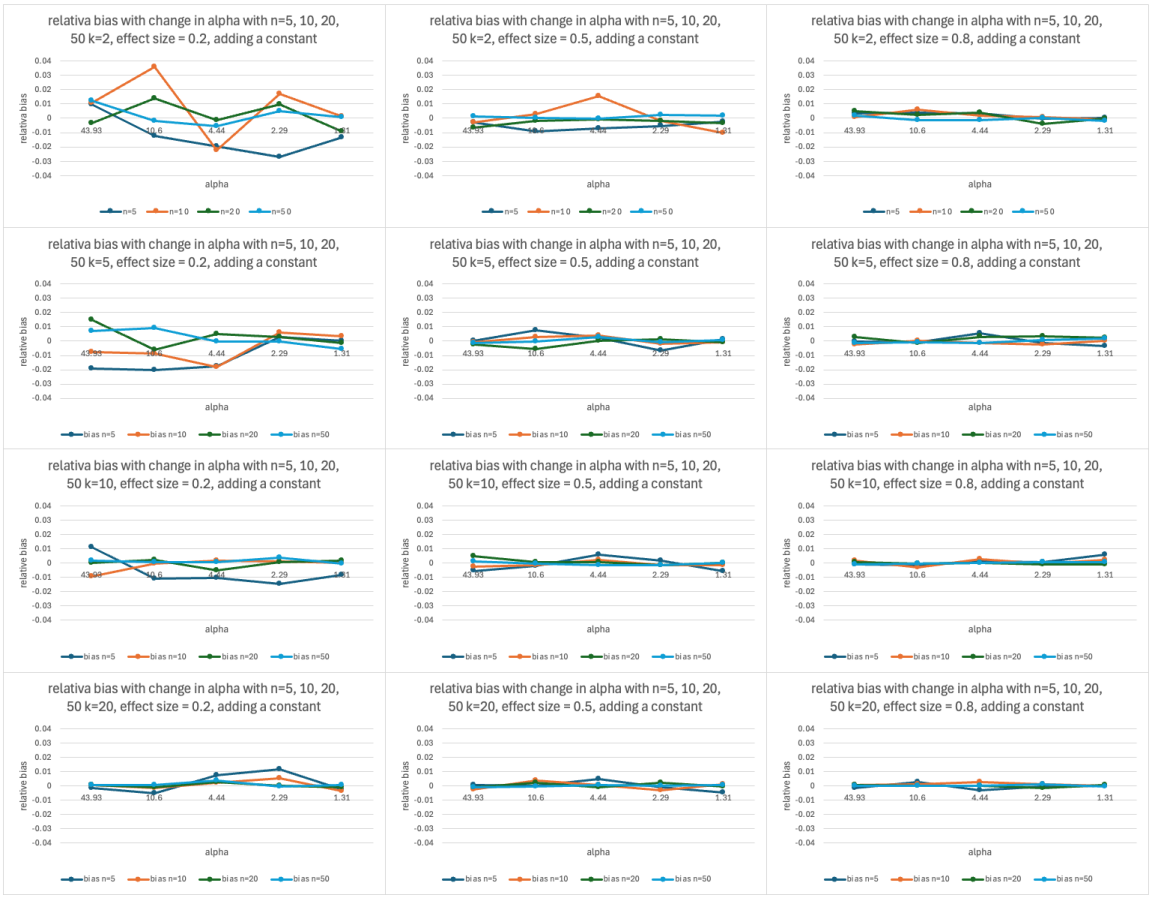

Figure 4.3.1 The relative bias in the estimated treatment effect, as a function of α and n across different effect sizes, with adding a constant to gamma data using the mean difference for a fixed value of k at all levels of 2, 5, 10, and 20.

Figure 4.3.2 The relative bias in the estimated treatment effect, as a function of α and k across different effect sizes, with adding a constant to gamma data using the mean difference for a fixed value of n at all levels of 5, 10, 20, and 50.

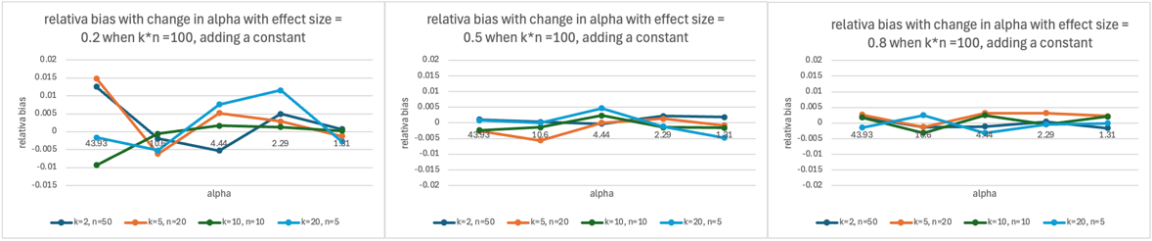

Figure 4.3.3 The relative bias in the estimated treatment effect, as a function of α and different combinations of k and n across different effect sizes, with adding a constant to gamma data using the mean difference for a fixed total sample size of 200.

From Figure 4.3.4, with a fixed value of k at all levels, the coverage probability is the lowest when n=5 and is lower than 0.9. There is a slight increase in coverage probability

with increasing skewness when n=5. As presented in Figure 4.3.5, with a fixed value of n at 5, the coverage probability falls below 0.925 and increases as k decreases. For other fixed n values, there is no important trend in coverage probability with k, skewness and effect sizes. There are only three cases when the coverage probability is around 0.95, which are k=2, 10, n=50, d=0.2 and k=5, n=50, d=0.8. These three cases all have a small skewness. Except for these three cases, the coverage probability is all below 0.95. Moreover, whenever the coverage probability is below 0.925, n = 5 and 10. With a fixed total sample size of 200, the coverage probability is the best (closest to 0.95) with the least number of studies and the most patients per study, as shown in Figure 4.3.6.

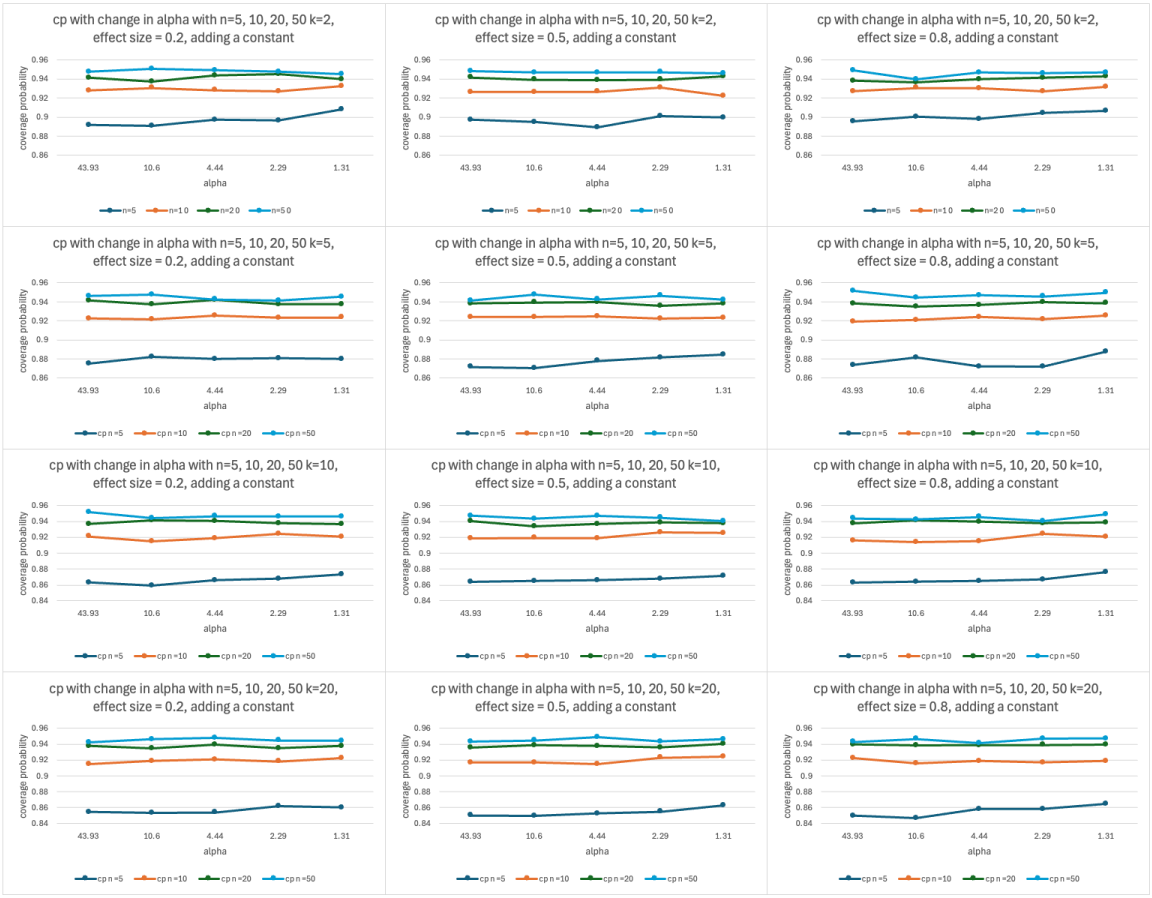

Figure 4.3.4 The coverage probability in the estimated treatment effect, as a function of α and n across different effect sizes, with adding a constant to gamma data using the mean difference for a fixed value of k at all levels of 2, 5, 10, and 20.

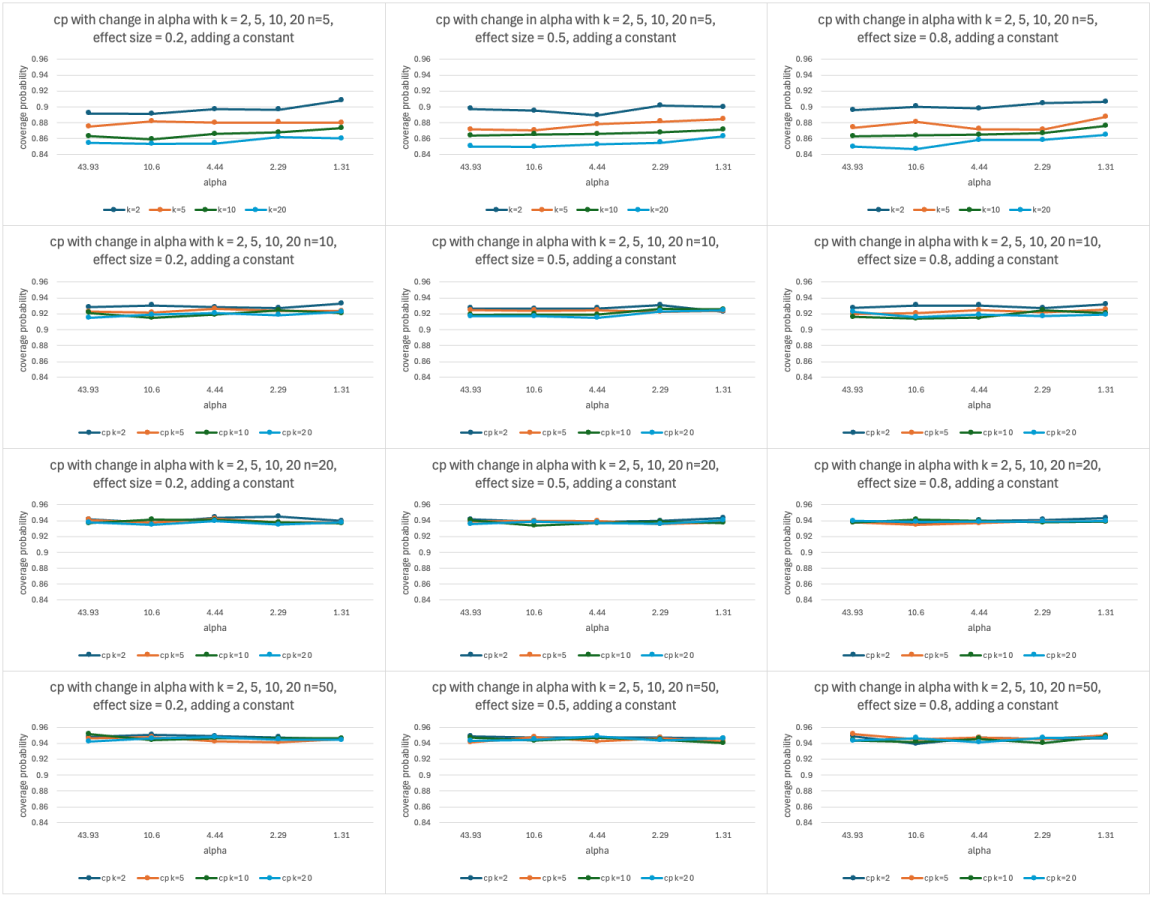

Figure 4.3.5 The coverage probability in the estimated treatment effect, as a function of α and k across different effect sizes, with adding a constant to gamma data using the mean difference for a fixed value of n at all levels of 5, 10, 20, and 50.

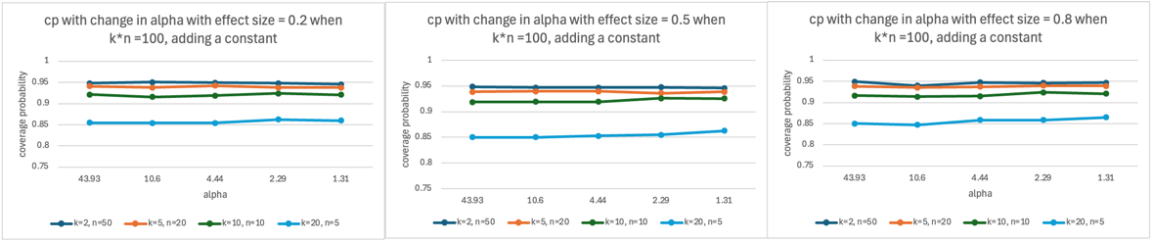

Figure 4.3.6 The coverage probability in the estimated treatment effect, as a function of α and different combinations of k and n across different effect sizes, with adding a constant to gamma data using the mean difference for a fixed value of k at all levels of 200.

The power increases as the number of studies, patients per study, and the effect size increases. As shown in Figure 4.3.7, with a fixed value of k, there is a slight increase in

power as skewness increases when n=5 for moderate and large effect sizes. As presented in Figure 4.3.8, with a fixed value of n, there is a slight increase in power as skewness increases with different values of k. With a small effect size, the power exceeds 0.8 when the total sample size equals or exceeds 800. With a moderate effect size, the power exceeds 0.8 when the total sample size equals or exceeds 200. With a large effect size, the power exceeds 0.8 when the total sample size exceeds 50.

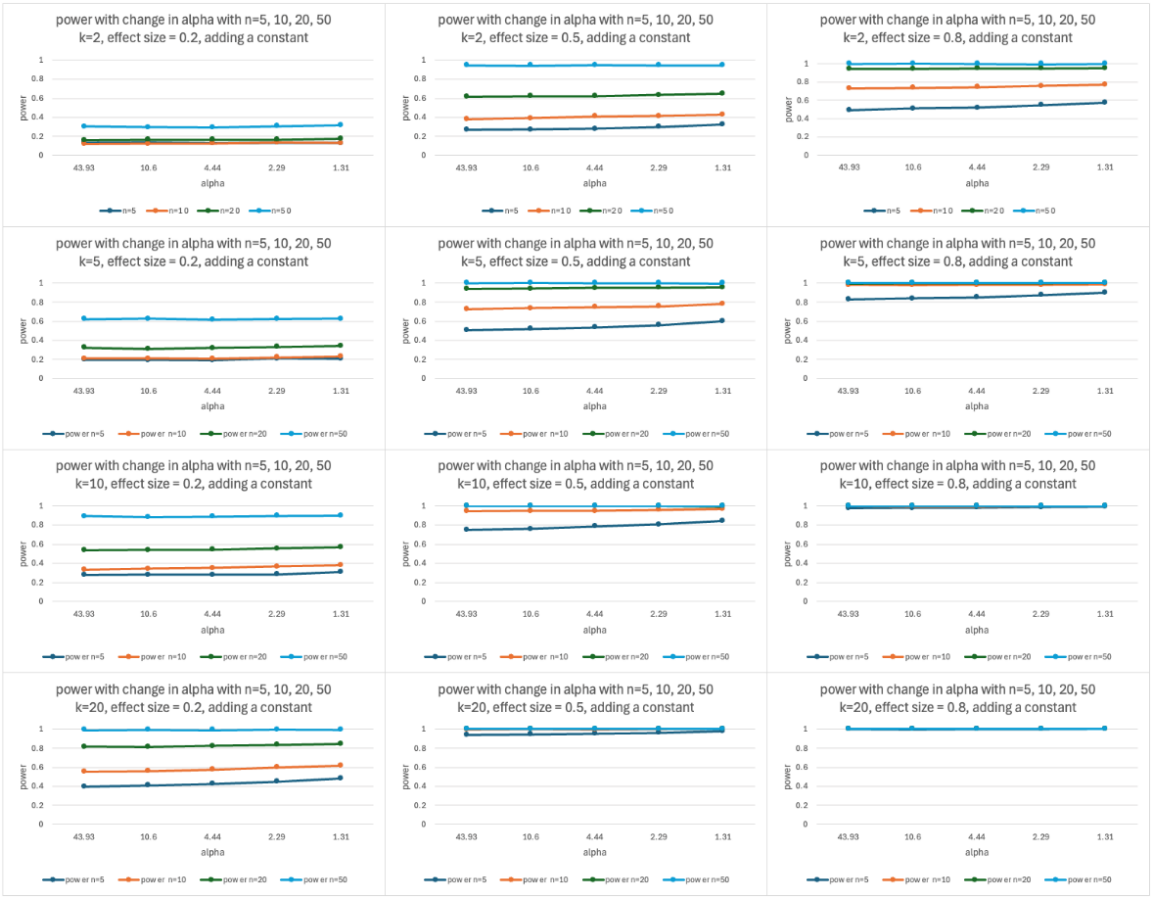

Figure 4.3.7 The power in the test, as a function of α and n across different effect sizes, with adding a constant to gamma data using the mean difference for a fixed value of k at all levels of 2, 5, 10, and 20.

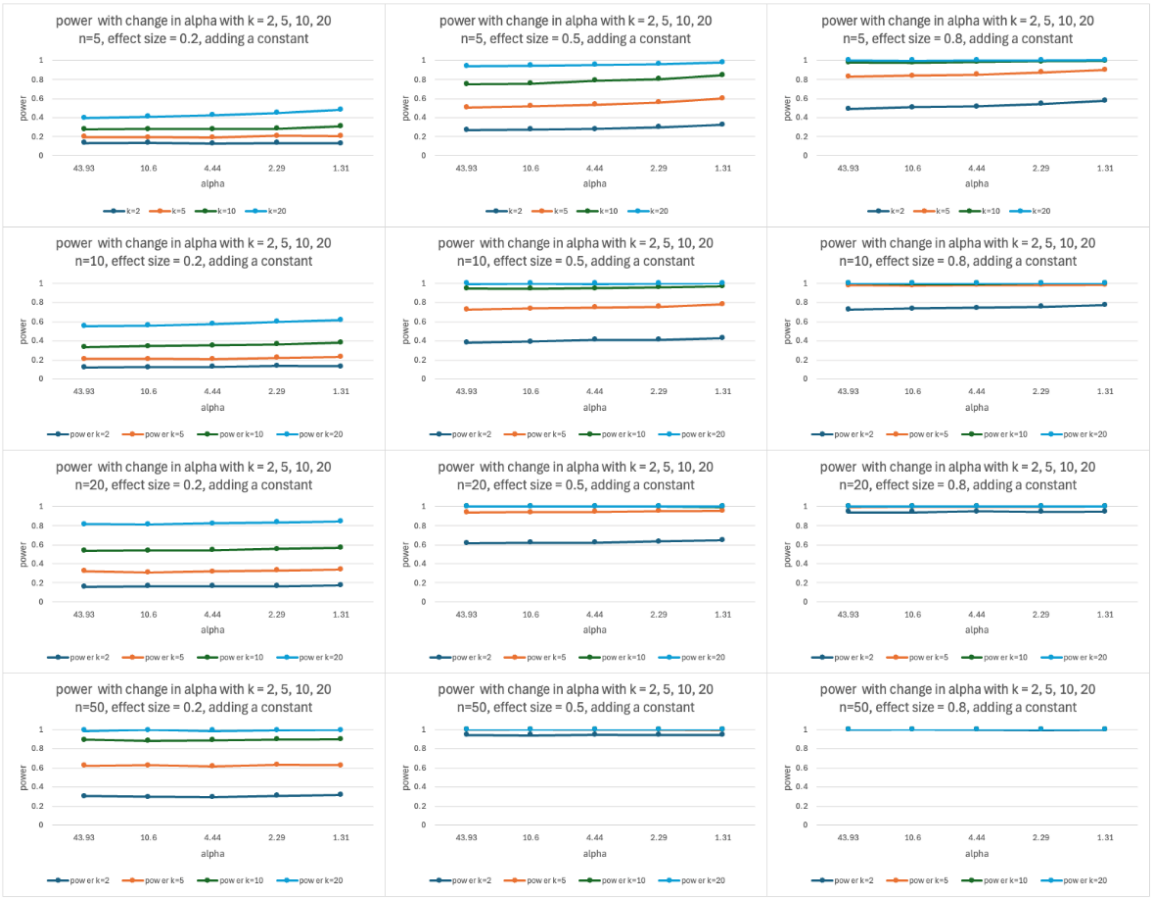

Figure 4.3.8 The power in the test, as a function of α and k across different effect sizes, with adding a constant to gamma data using the mean difference for a fixed value of n at all levels of 5, 10, 20, and 50.

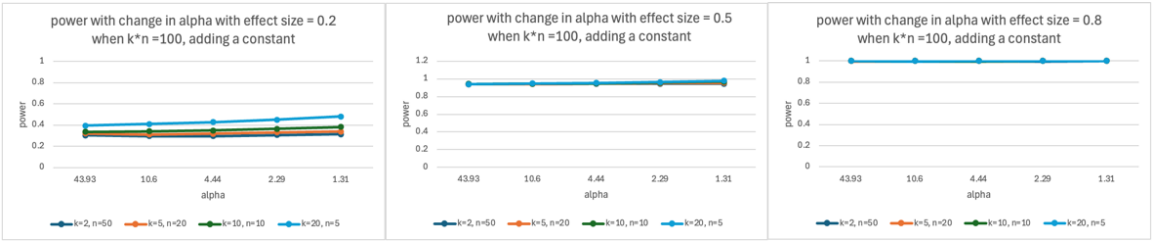

Figure 4.3.9 The power in the test, as a function of α and different combinations of k and n across different effect sizes, with adding a constant to gamma data using the mean difference for a fixed total sample size of 200.

The type I error rate is acceptable if k=5, 10, 20 and n=20, 50. When k=2, n=50, with skewness greater than 1 and when k=5, n=10, with skewness greater than 0.05, the type I

error rate is also in the acceptable range. The type I error rate are above 0.05 in all situations. In addition, the type I error rate decreases as the skewness increases.

4.3.1.2 Both Gamma with Different βs

From Figure 4.3.10, with a fixed value of k at all levels, the relative bias is closer to zero with larger n. From Figure 4.3.11, with a fixed value of n at all levels, more studies introduce a greater relative bias. When n=20 and 50, the absolute relative bias is below 10% for all k and effect sizes. In other situations, the absolute relative bias can reach 30% with highly skewed ($\alpha = 1.31$) outcome variables. As shown in Figure 4.3.12, with a fixed total sample size of 200, the relative bias is closer to zero with fewer studies and more patients per study included. The relative bias is negative when the effect size is small and moderate (d=0.2, 0.5) and is positive when the effect size is large in all cases.

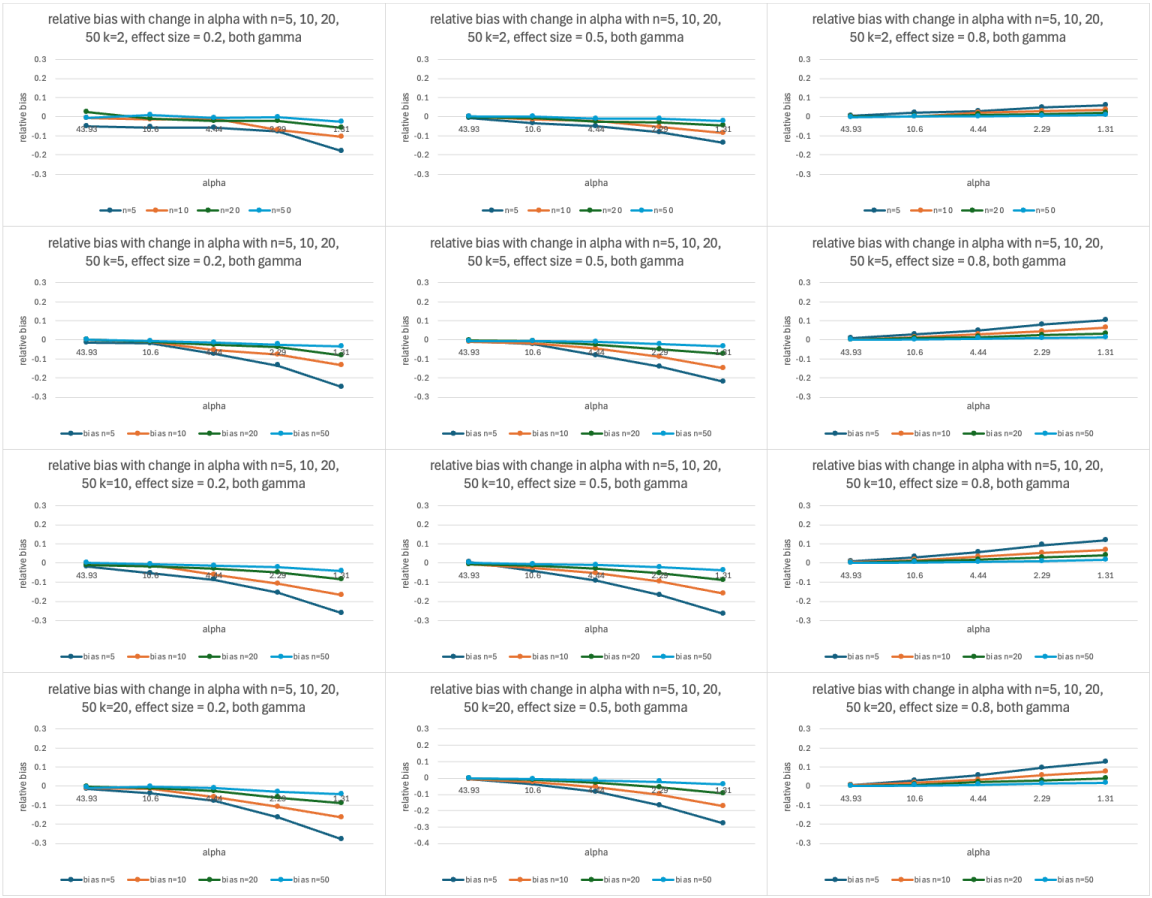

Figure 4.3.10 The relative bias in the estimated treatment effect, as a function of α and n across different effect sizes, with simulating two gamma distributions with different parameters using the mean difference for a fixed value of k at all levels of 2, 5, 10, and 20.

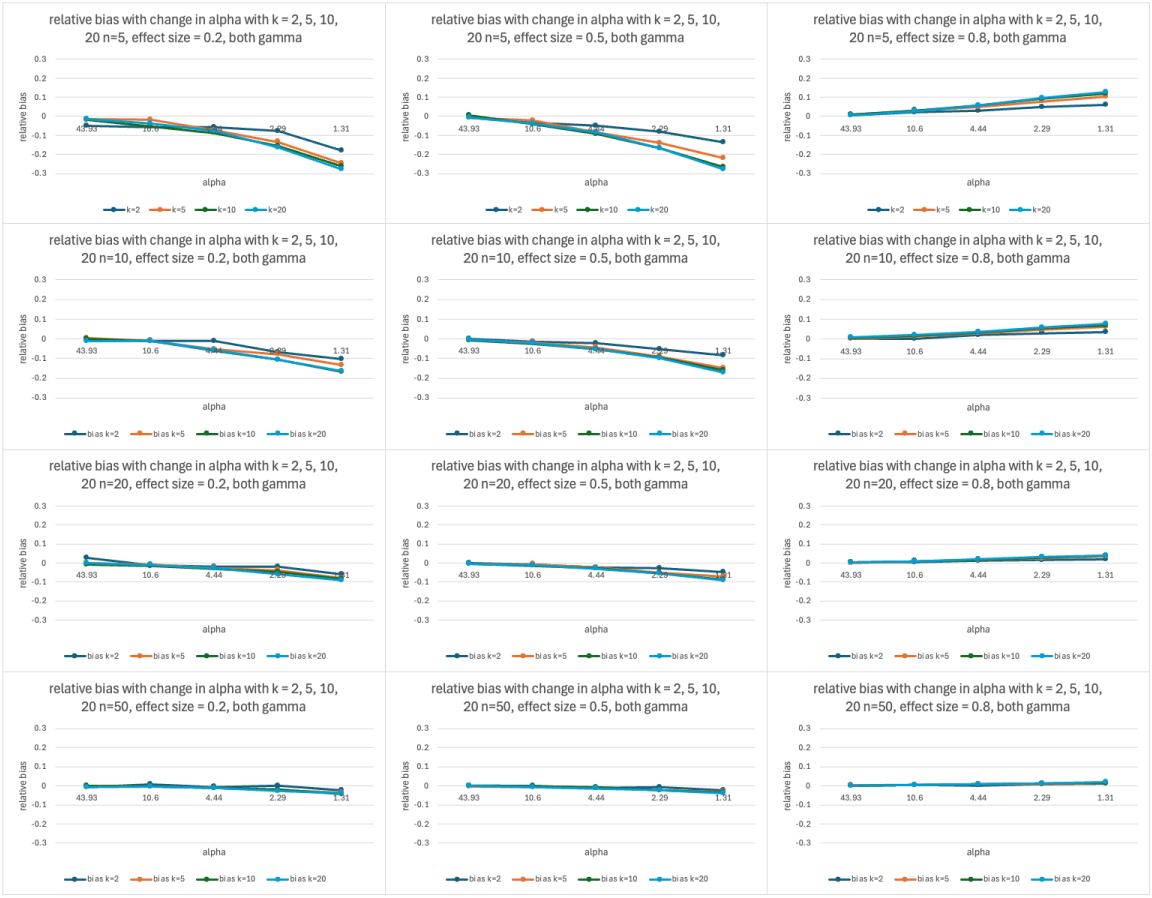

Figure 4.3.11 The relative bias in the estimated treatment effect, as a function of α and k across different effect sizes, with simulating two gamma distributions with different parameters using the mean difference for a fixed value of n at all levels of 5, 10, 20, and 50.

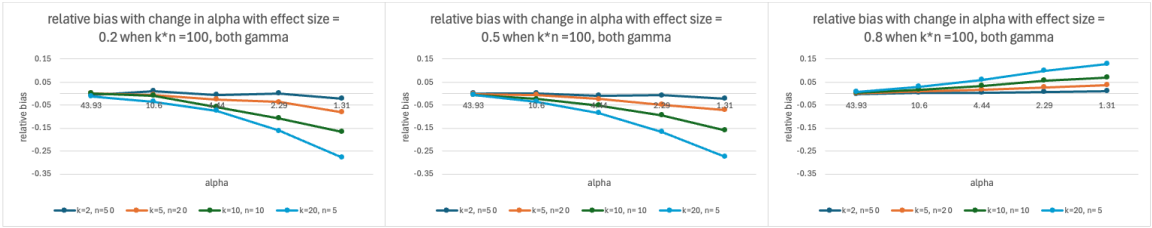

Figure 4.3.12 The relative bias in the estimated treatment effect, as a function of α and different combinations of k and n across different effect sizes, with simulating two gamma distributions with different parameters using the mean difference for a fixed total sample size of 200.

With a fixed value of k at all levels, the coverage probability increases towards 0.95 as n increases, as illustrated in Figure 4.3.13. With a fixed value of n at all levels, the coverage

probability increases with smaller k as shown in Figure 4.3.14. For all cases, there is a negligible trend in the coverage probability with increasing skewness for a small effect size. However, with moderate and large effect sizes, the coverage probability decreases as skewness increases. Whenever the coverage probability is below 0.925, n=5 and 10 for most cases. With other values of n and a skewness greater than 1 ($\alpha = 2.29, 1.31$), the coverage probability is below 0.925. The lowest coverage probability can be as low as 0.4 with k=20, n=5, and d=0.8, with the largest skewness. All the cases have a coverage probability under 0.95. Given in Figure 4.3.15, with a fixed total sample size of 200, the coverage probability is ideal and does not change with skewness when k=2, n=50. When k=20, n=5 and k=10, n=10, the coverage probability decreases as the skewness increases.

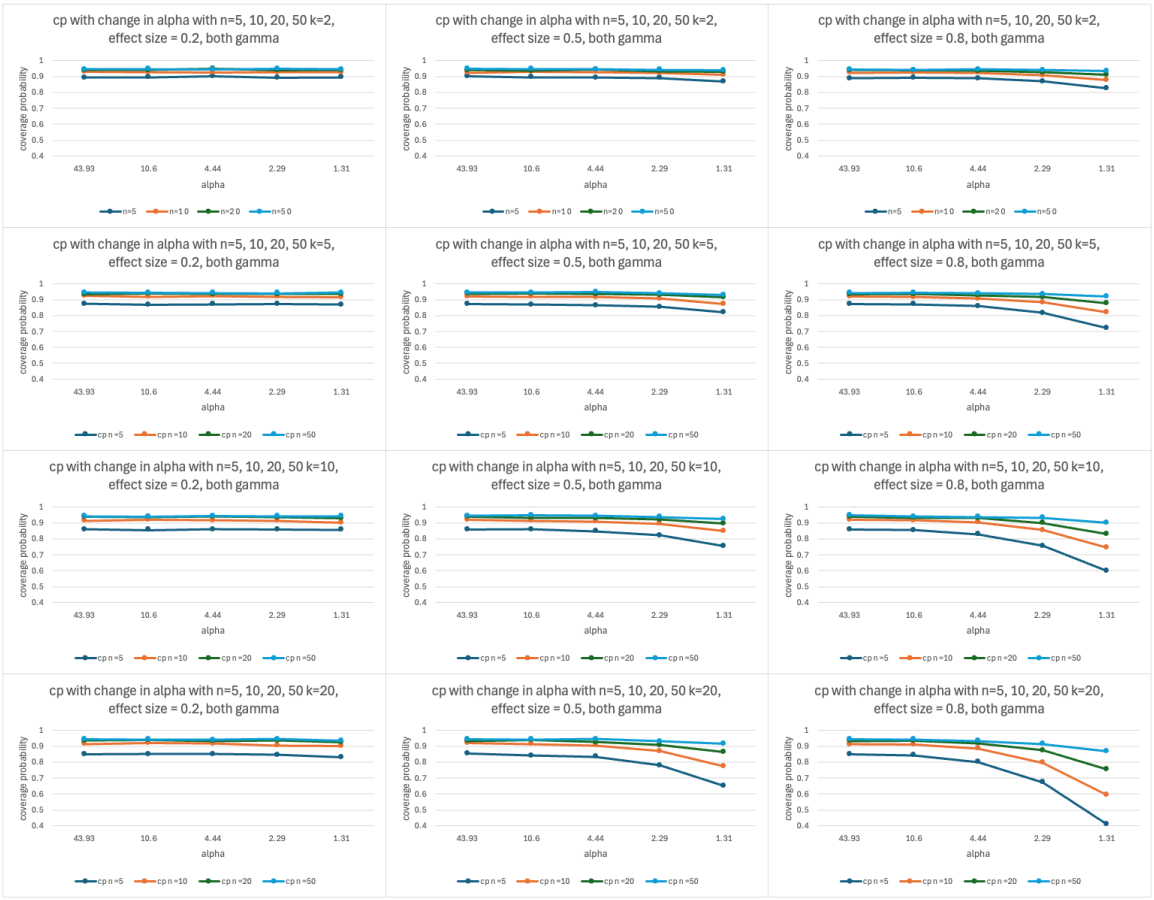

Figure 4.3.13 The coverage probability in the estimated treatment effect, as a function of α and n across different effect sizes, with simulating two gamma distributions with different parameters using the mean difference for a fixed value of k at all levels of 2, 5, 10, and 20.

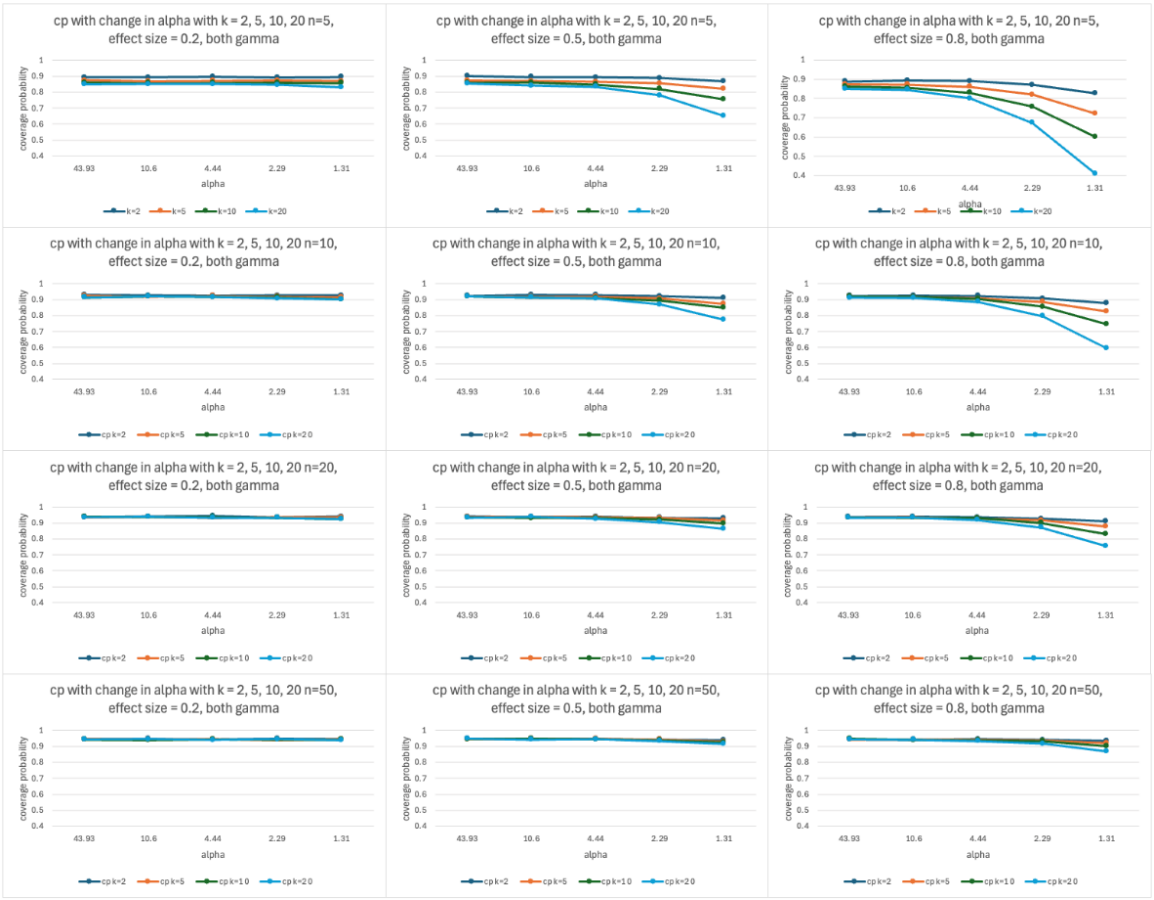

Figure 4.3.14 The coverage probability in the estimated treatment effect, as a function of α and k across different effect sizes, with simulating two gamma distributions with different parameters using the mean difference for a fixed value of n at all levels of 5, 10, 20, and 50.

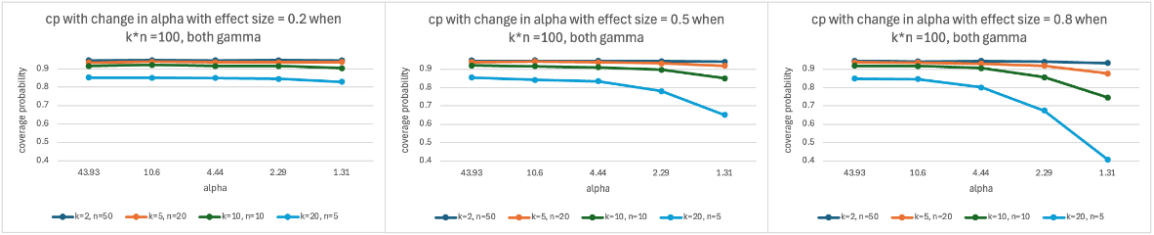

Figure 4.3.15 The coverage probability in the estimated treatment effect, as a function of α and different combinations of k and n across different effect sizes, with simulating two gamma distributions with different parameters using the mean difference for a fixed total sample size of 200.

The power increases as the number of studies, patients per study, and the effect size increases. As shown in Figure 4.3.16, with a fixed value of k, there is a slight decrease in

power as skewness increases when n=5 for all effect sizes. Given in Figure 4.3.17, With a fixed value of n, there is an undetectable trend in power as skewness increases with different values of k. With a small effect size, the power exceeds 0.8 when the total sample size equals or exceeds 800. With a moderate effect size, the power exceeds 0.8 when the total sample size equals or exceeds 200. With a large effect size, the power exceeds 0.8 when the total sample size exceeds 50.

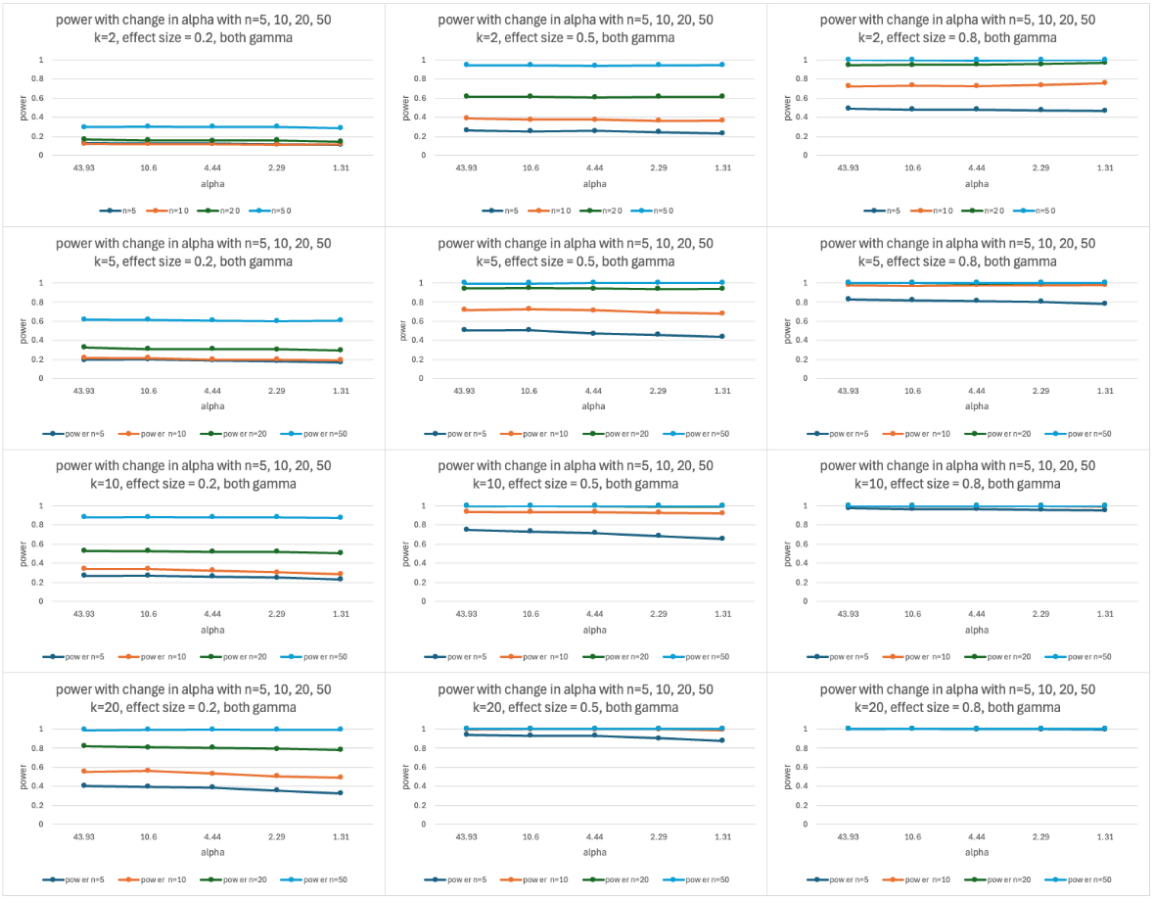

Figure 4.3.16 The power in the test, as a function of α and n across different effect sizes, with simulating two gamma distributions with different parameters using the mean difference for a fixed value of k at all levels of 2, 5, 10, and 20.

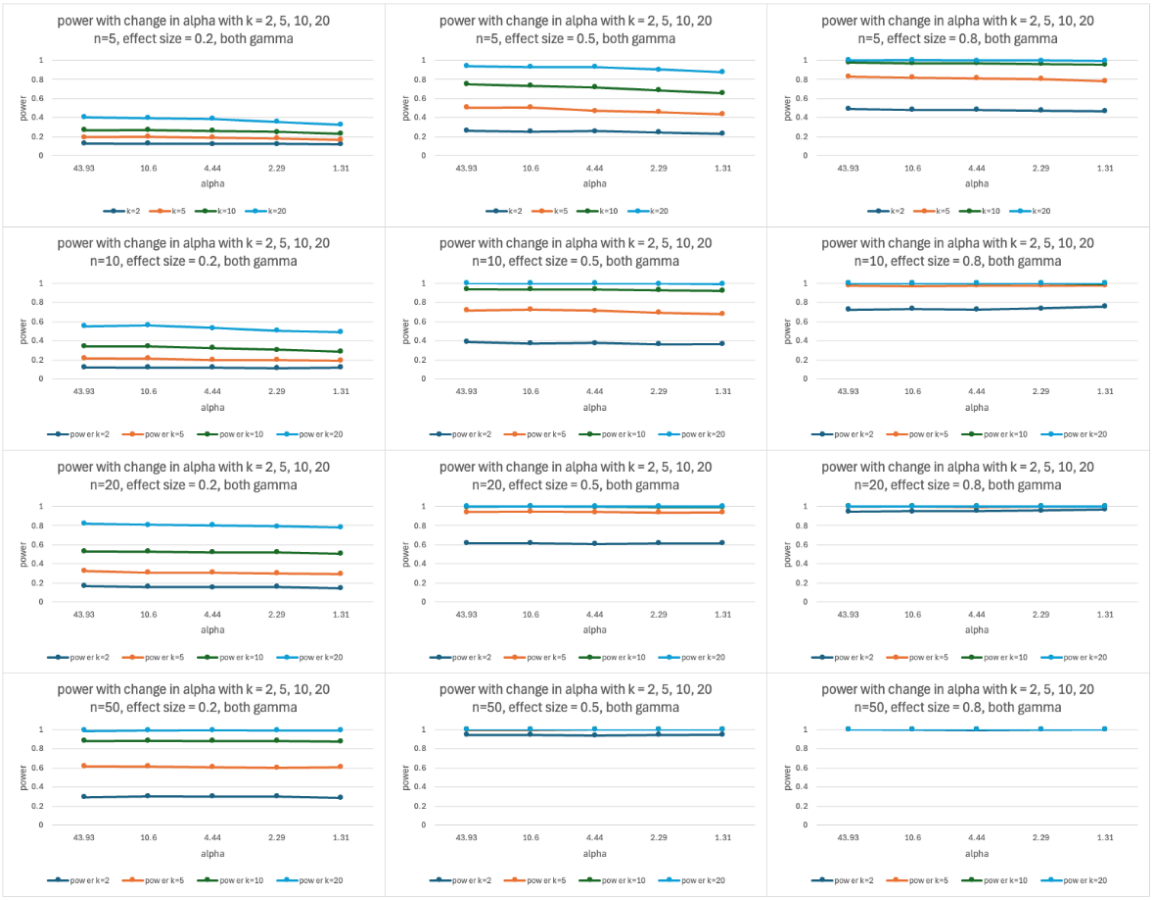

Figure 4.3.17 The power in the test, as a function of α and k across different effect sizes, with simulating two gamma distributions with different parameters using the mean difference for a fixed value of n at all levels of 5, 10, 20, and 50.

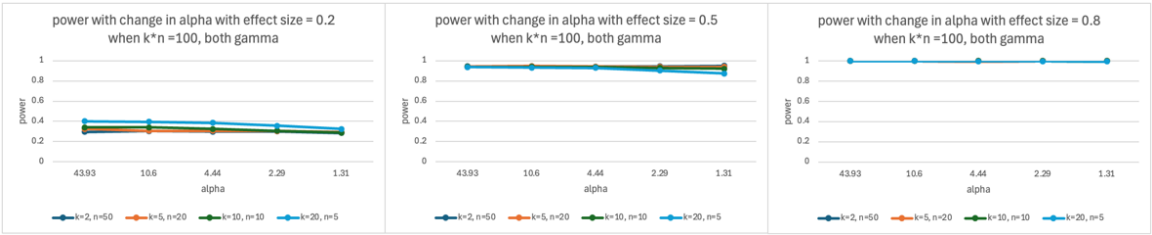

Figure 4.3.18 The power in the test, as a function of α and different combinations of k and n across different effect sizes, with simulating two gamma distributions with different parameters using the mean difference for a fixed total sample size of 200.

4.3.2 Standardized Mean Difference

4.3.2.1 Adding a Constant

From Figure 4.3.19, with a fixed value of k at all levels, the relative bias approaches zero as n and skewness increase. As shown in Figure 4.3.20, when n is fixed at 5 and 10, the relative bias approaches zero with a smaller k. There is no noticeable trend in the relative bias with effect sizes. When n = 20 and 50, the absolute relative bias is below 5% with different values of k. n=5 whenever the absolute relative bias is greater than 10%. When the skewness is greater than 1, the relative bias is positive, indicating an overestimation of the relative bias with a greater skewness. With a fixed total sample size of 200, the relative bias is the farthest from 0, with the most studies and the least number of patients per study, as presented in Figure 4.3.21.

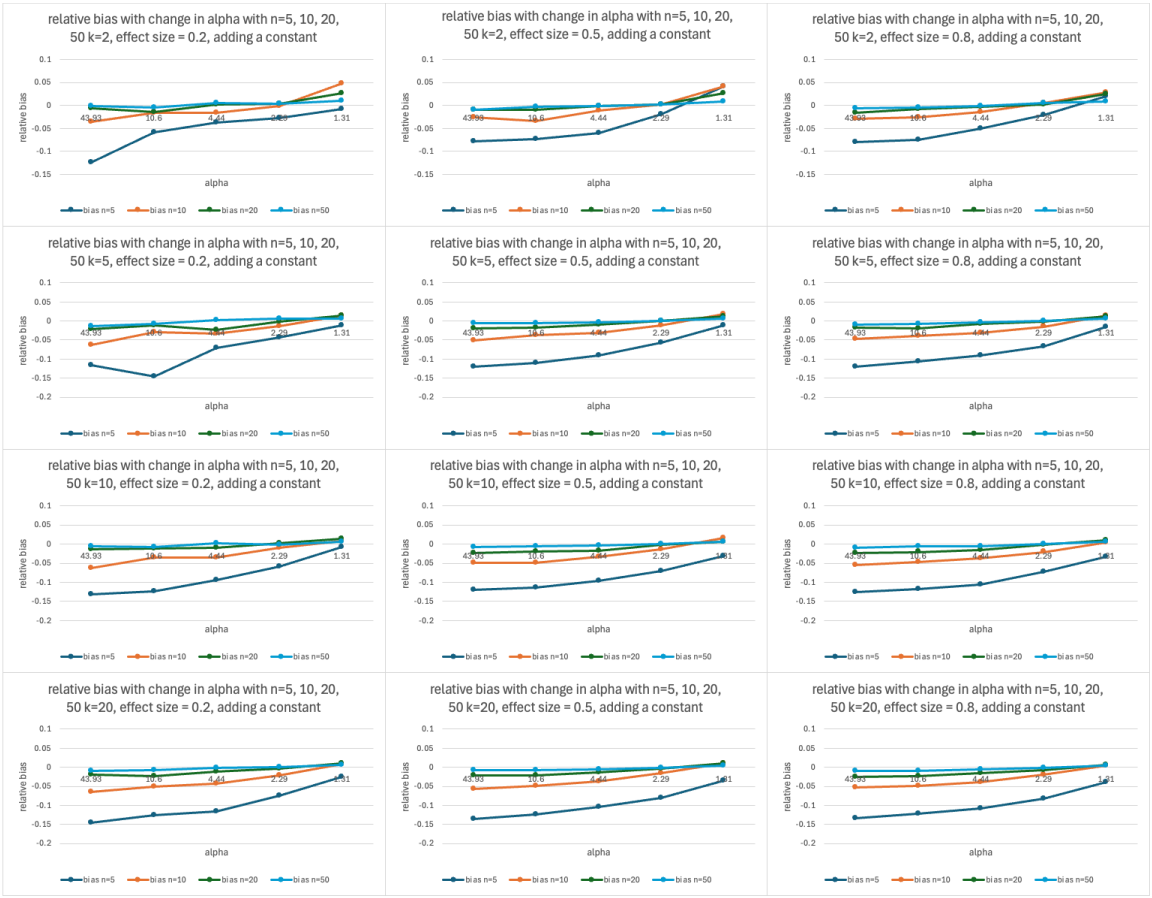

Figure 4.3.19 The relative bias in the estimated treatment effect, as a function of α and n across different effect sizes, with adding a constant to gamma data using the standardized mean difference for a fixed value of k at all levels of 2, 5, 10, and 20.

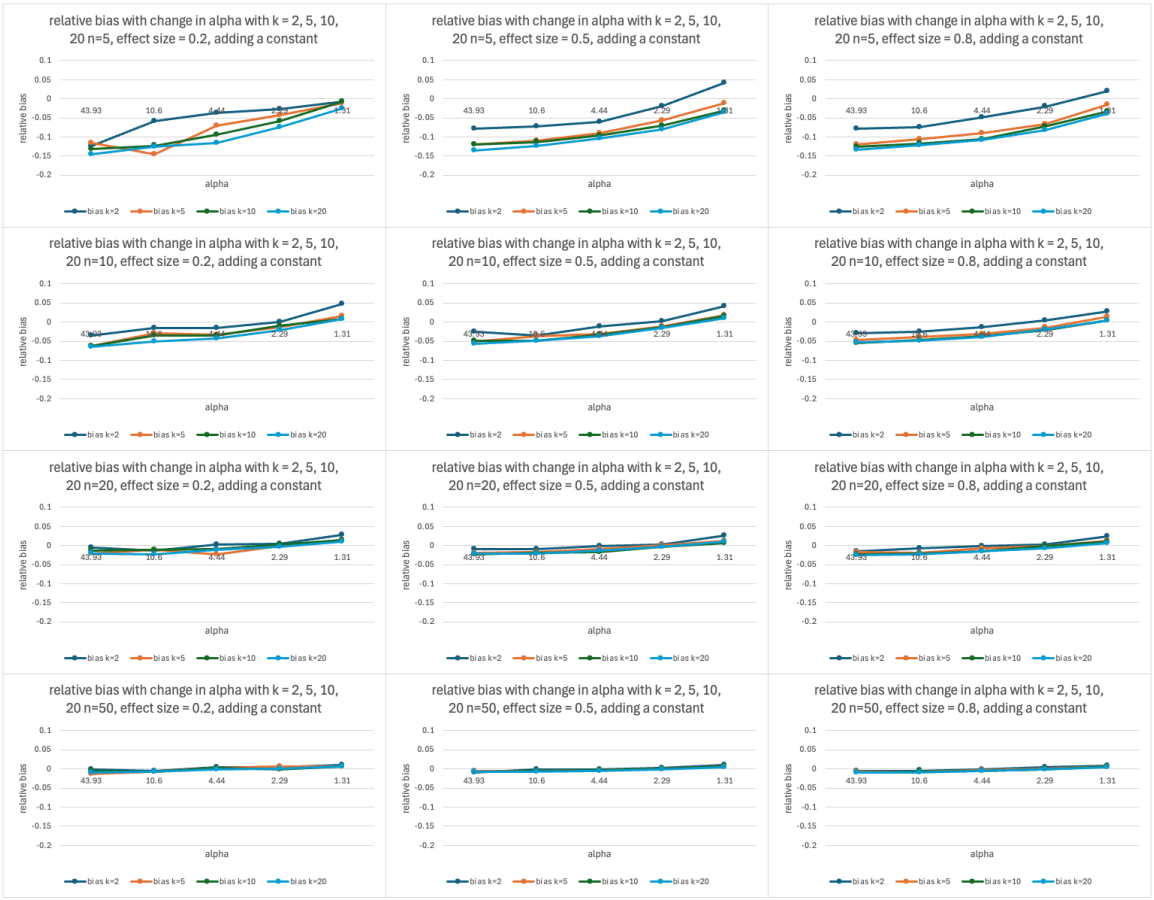

Figure 4.3.20 The relative bias in the estimated treatment effect, as a function of α and k across different effect sizes, with adding a constant to gamma data using the standardized mean difference for a fixed value of n at all levels of 5, 10, 20, and 50.

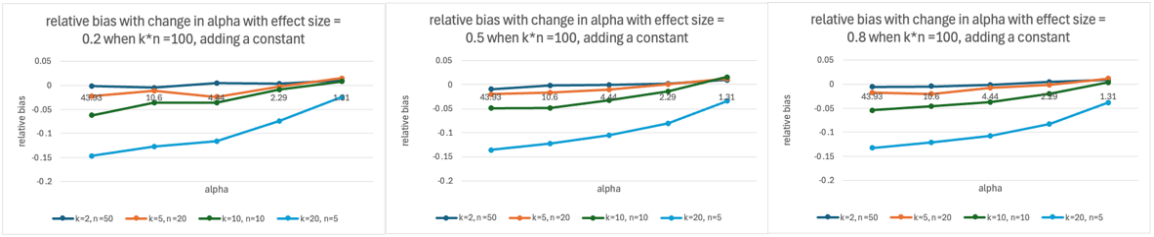

Figure 4.3.21 The relative bias in the estimated treatment effect, as a function of α and different combinations of k and n across different effect sizes, with adding a constant to gamma data using the standardized mean difference for a fixed total sample size of 200.

With a fixed value of k for all levels, the coverage probability increases as n decreases for all effect sizes, as illustrated in Figure 4.3.22. When n=5, 10, the coverage probability is above 0.95 and closest to 0.95 when n=20 and 50. With moderate and large effect sizes,

the coverage probability inflates when the skewness increases with n=5 and decreases as skewness increases with other n. In Figure 4.3.23, with a fixed value of n for all levels, the coverage probability increases when k decreases with n=5 for a large effect size. The skewness does not noticeably impact the coverage probability for a small effect size. When n=5, the coverage probability falls outside of 0.925 and 0.975; Moreover, they are all above 0.975. In Figure 4.3.24, with a fixed total sample size of 200, the coverage probability is not robust with the changing skewness when k=20, n=5, and is far from 0.95 compared with other combinations of k and n.

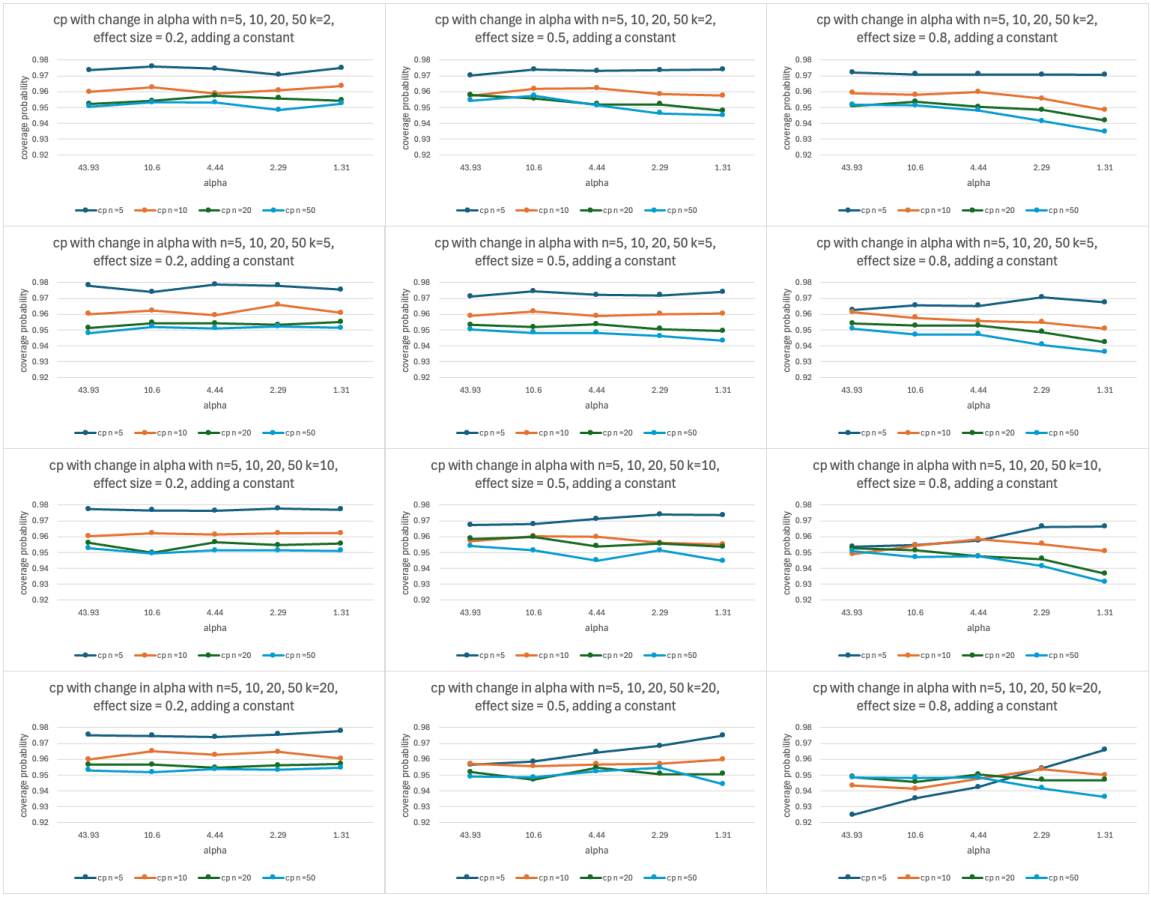

Figure 4.3.22 The coverage probability in the estimated treatment effect, as a function of α and n across different effect sizes, with adding a constant to gamma data using the standardized mean difference for a fixed value of k at all levels of 2, 5, 10, and 20.

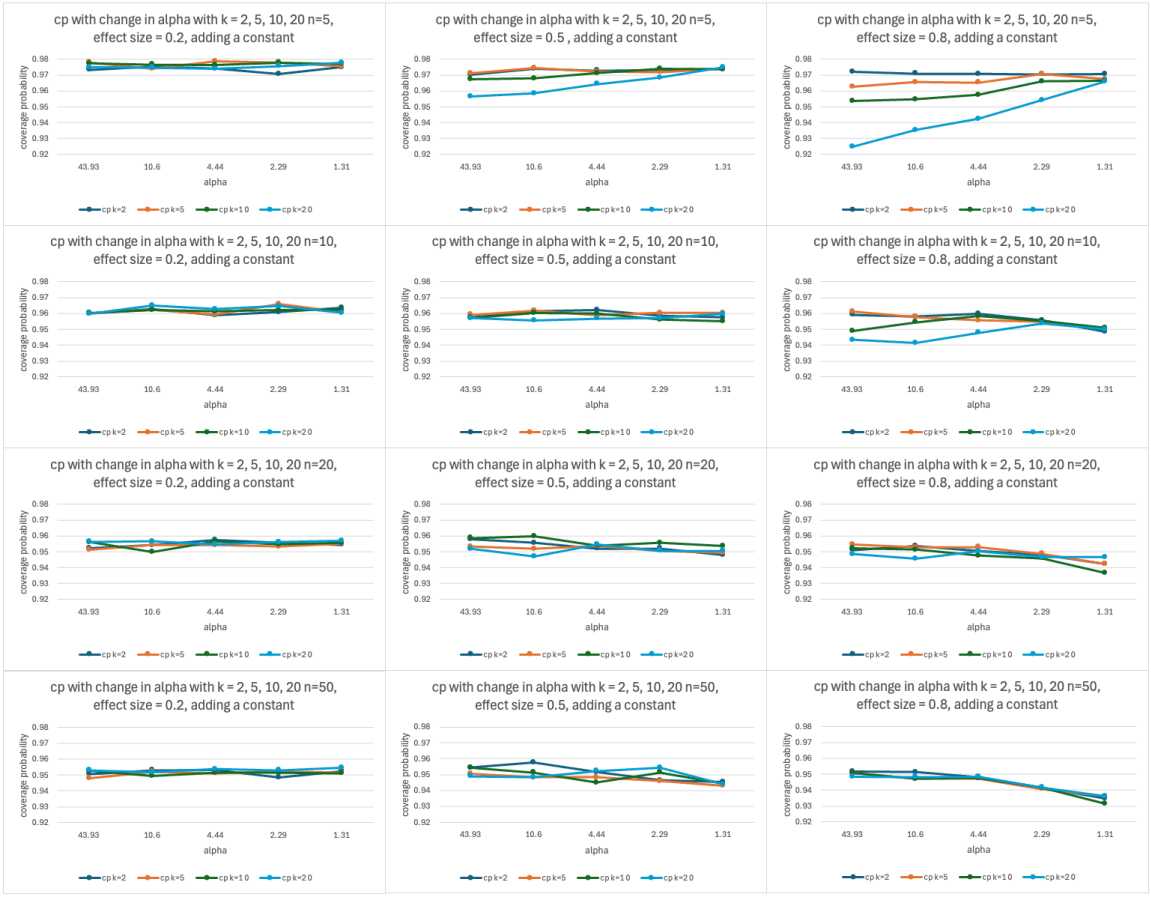

Figure 4.3.23 The coverage probability in the estimated treatment effect, as a function of α and k across different effect sizes, with adding a constant to gamma data using the standardized mean difference for a fixed value of n at all levels of 5, 10, 20, and 50.

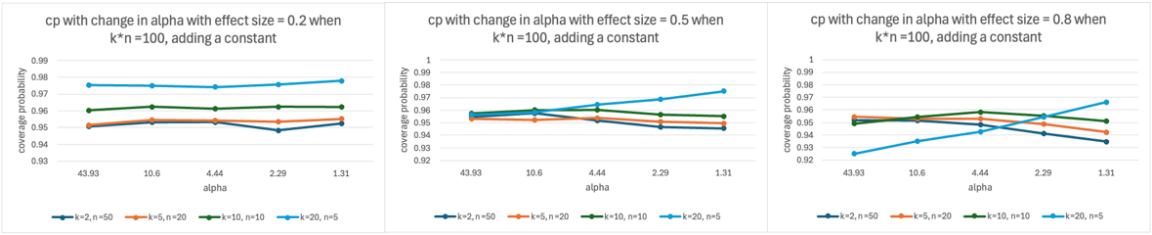

Figure 4.3.24 The coverage probability in the estimated treatment effect, as a function of α and different combinations of k and n across different effect sizes, with adding a constant to gamma data using the standardized mean difference for a fixed total sample size of 200.

The power increases as the number of studies, patients per study and effect sizes increase.

There is an unimportant impact of skewness on power. However, in Figure 4.3.25, with a

fixed value of k, there is a slight increase of power with the skewness when n=5 for moderate and large effect sizes. With a small effect size, the power exceeds 0.8 when the total sample size equals or exceeds 1000. With a moderate effect size, the power exceeds 0.8 when the total sample size equals or exceeds 200. With a large effect size, the power exceeds 0.8 when the total sample size exceeds 80.

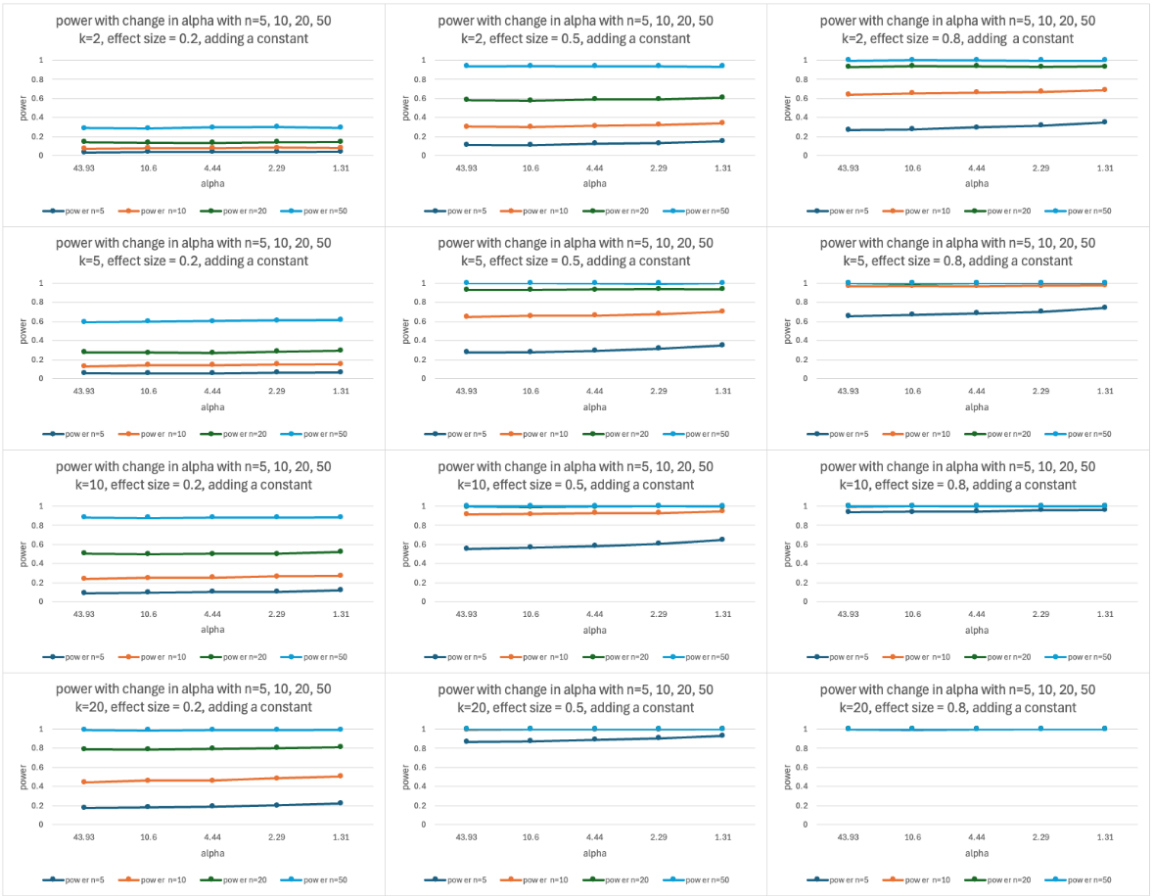

Figure 4.3.25 The power in the test, as a function of α and n across different effect sizes, with adding a constant to gamma data using the standardized mean difference for a fixed value of k at all levels of 2, 5, 10, and 20.

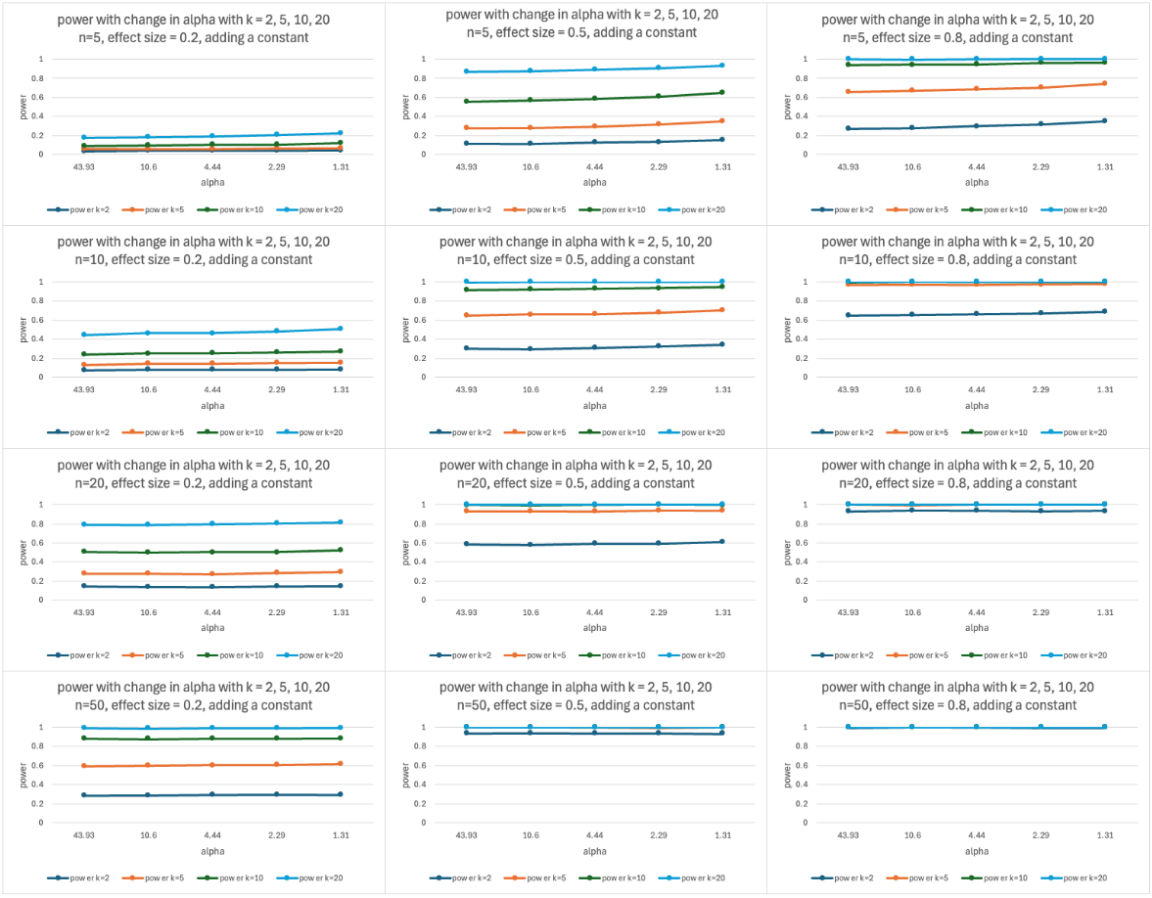

Figure 4.3.26 The power in the test, as a function of α and k across different effect sizes, with adding a constant to gamma data using the standardized mean difference for a fixed value of n at all levels of 5, 10, 20, and 50.

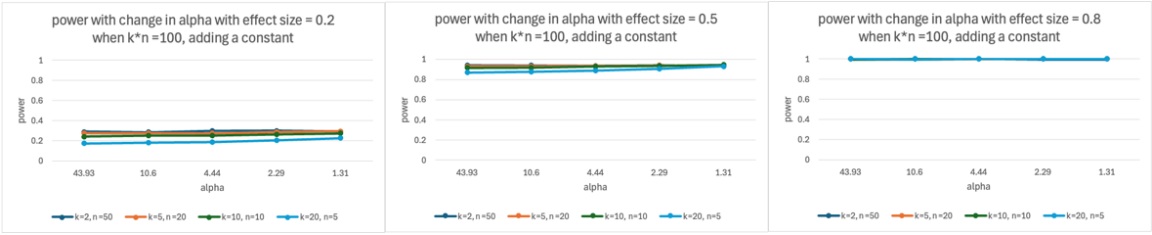

Figure 4.3.27 The power in the test, as a function of α and different combinations of k and n across different effect sizes, with adding a constant to gamma data using the standardized mean difference for a fixed total sample size of 200.

When using SMD, when k=2 and n=5 and 10, the type I error rate is unacceptable.

Skewness has no undetectable effect on type I error rate.

4.3.2.2 Both Gammas with Different βs

With a fixed value of k at all levels, the relative bias approaches zero as n increases, as illustrated in Figure 4.3.28. The relative bias moves farther from zero as skewness increases when n=5, 10 for small and moderate effect sizes with k= 10, 20. The relative bias approaches zero for a large effect size as the skewness increases when n=5 for a large k. With a fixed value of small n, small k results in better relative bias, as shown in Figure 4.3.29. The absolute relative bias is below 5% when n equals 20 and 50. Whenever the relative bias is more than 10%, n = 5. The relative bias for all the situations is negative except for one case where k=2, n=50, d=0.8, with the largest skewness. In Figure 4.3.30, with a fixed total sample size of 200, the relative bias is the worst, with the greatest number of studies and the least patients per study.

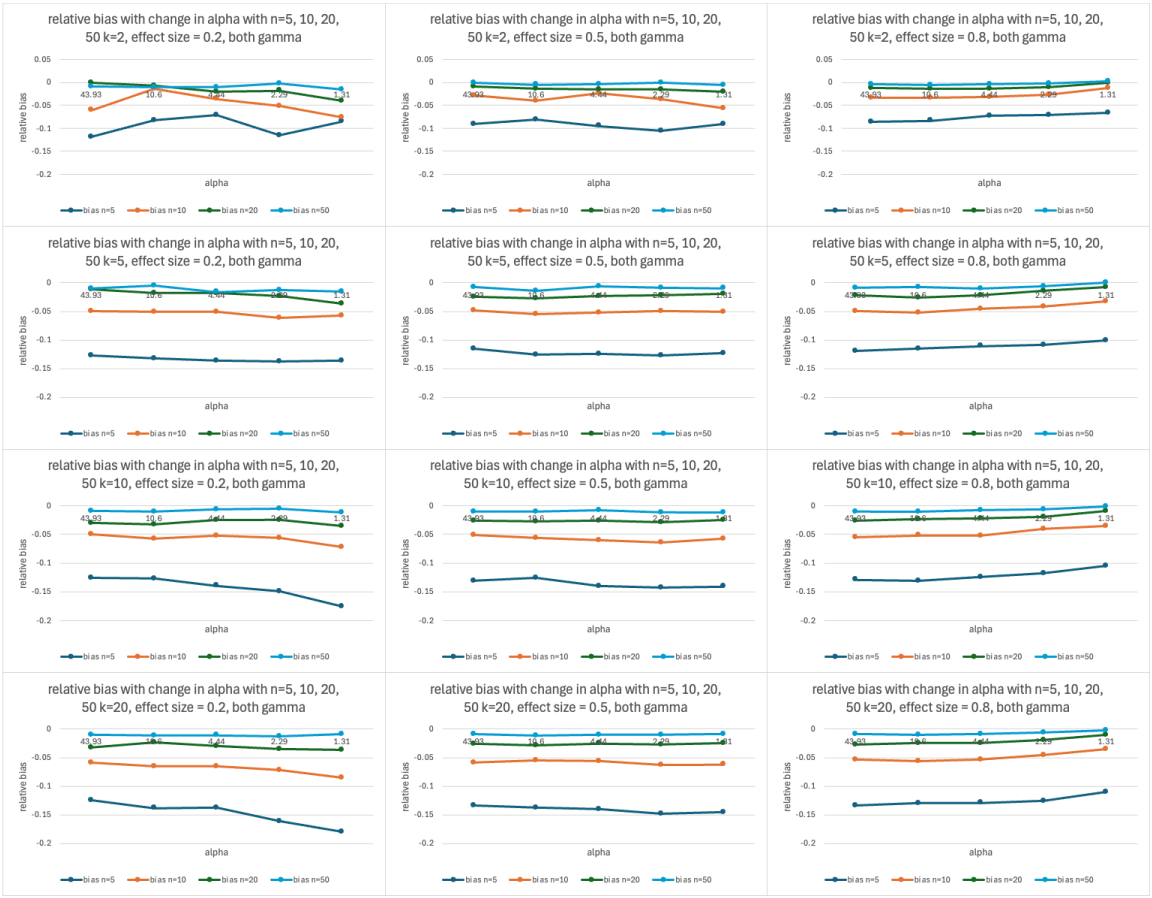

Figure 4.3.28 The relative bias in the estimated treatment effect, as a function of α and n across different effect sizes, with simulating two gamma distributions with different parameters using the standardized mean difference for a fixed value of k at all levels of 2, 5, 10, and 20.

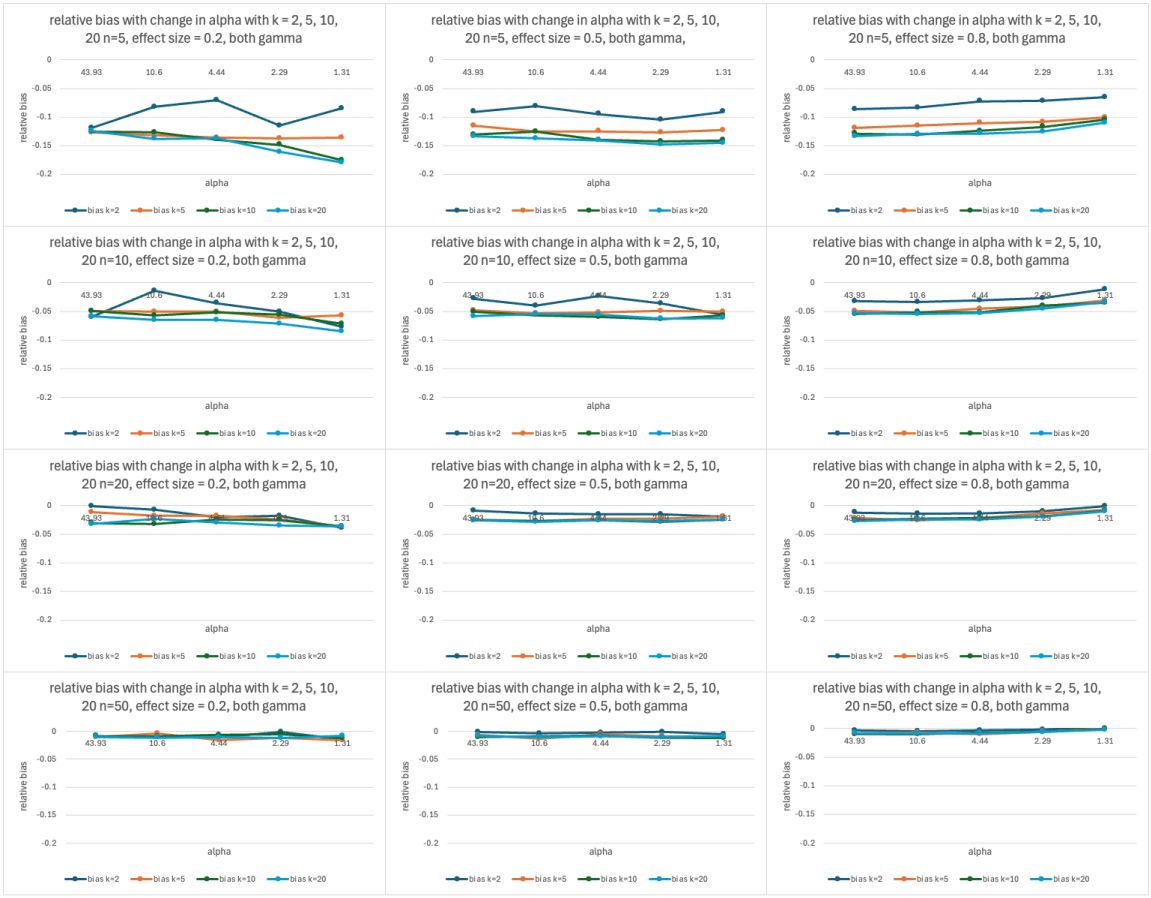

Figure 4.3.29 The relative bias in the estimated treatment effect, as a function of α and k across different effect sizes, with simulating two gamma distributions with different parameters using the standardized mean difference for a fixed value of n at all levels of 5, 10, 20, and 50.

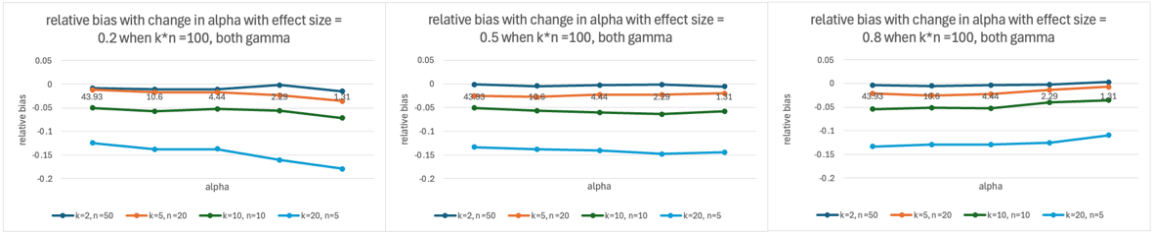

Figure 4.3.30 The relative bias in the estimated treatment effect, as a function of α and different combinations of k and n across different effect sizes, with simulating two gamma distributions with different parameters using the standardized mean difference for a fixed total sample size of 200.

In Figure 4.3.31, with a fixed value of k at all levels, the coverage probability is the worst when n=5 and is robust around 0.95 with other n for a small effect size. In Figure 4.3.32, with a fixed value of n at all levels, the coverage probability increases as k decreases and skewness increases for moderate and large effect sizes. The coverage probability is robust around 0.95 against the skewness with a small effect size and is robust when the skewness is less than 1 for moderate and large effect sizes. Most situations where the coverage probability is outside the acceptable range are either with small total sample sizes or large skewness. With a fixed total sample size of 200, the coverage probability is the worst when k=20 and n=5, as given in Figure 4.3.33.

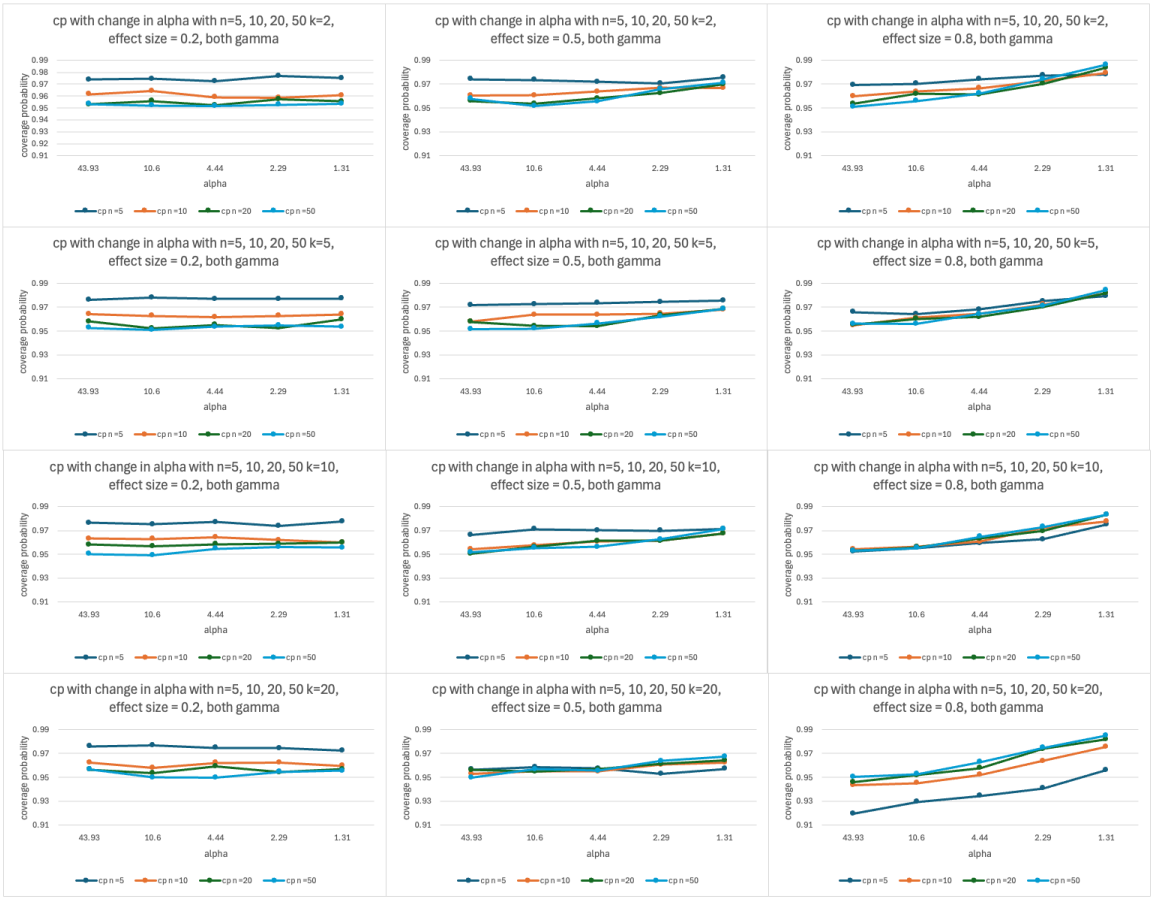

Figure 4.3.31 The coverage probability in the estimated treatment effect, as a function of α and n across different effect sizes, with two gamma distributions with different parameters using the standardized mean difference for a fixed value of k at all levels of 2, 5, 10, and 20.

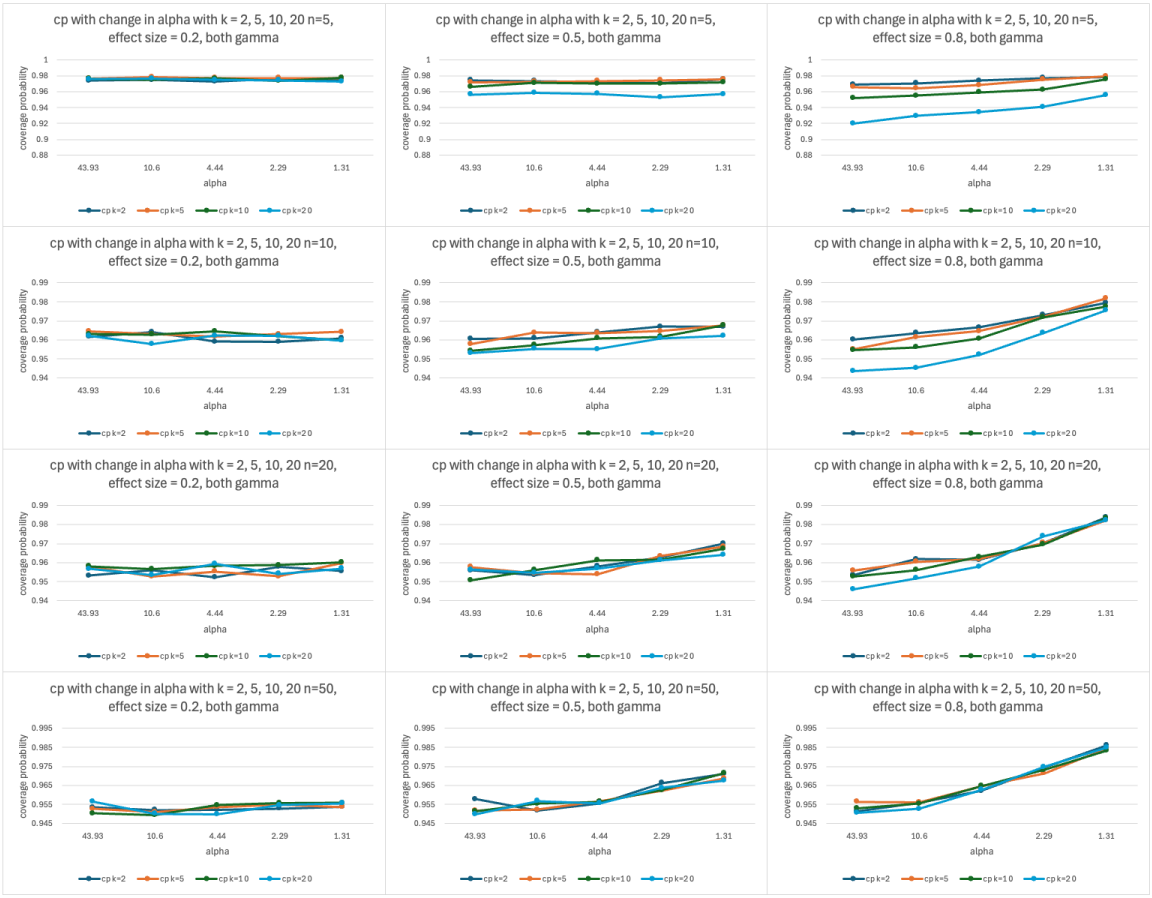

Figure 4.3.32 The coverage probability in the estimated treatment effect, as a function of α and k across different effect sizes, with simulating two gamma distributions with different parameters using the standardized mean difference for a fixed value of n at all levels of 5, 10, 20, and 50.

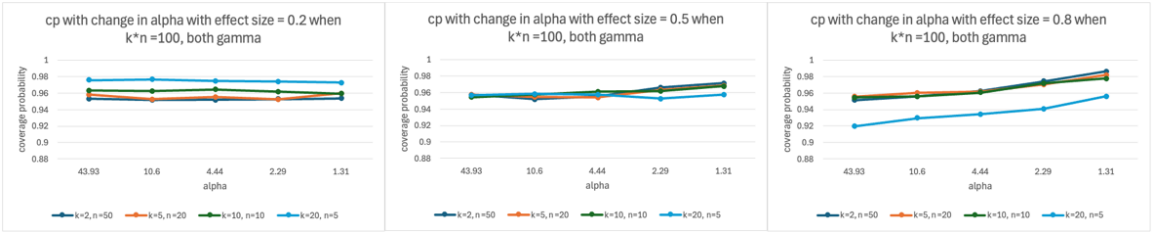

Figure 4.3.33 The coverage probability in the estimated treatment effect, as a function of α and different combinations of k and n across different effect sizes, with simulating two gamma distributions with different parameters using the standardized mean difference for a fixed total sample size of 200.

Figure 4.3.34 and Figure 4.3.35 show that the power increases as the number of studies, patients per study, and effect sizes increase. There is a negligible effect of skewness on

power. With a small effect size, the power exceeds 0.8 when the total sample size equals or exceeds 1000. With a moderate effect size, the power exceeds 0.8 when the total sample size equals or exceeds 200. With a large effect size, the power exceeds 0.8 when the total sample size exceeds 80.

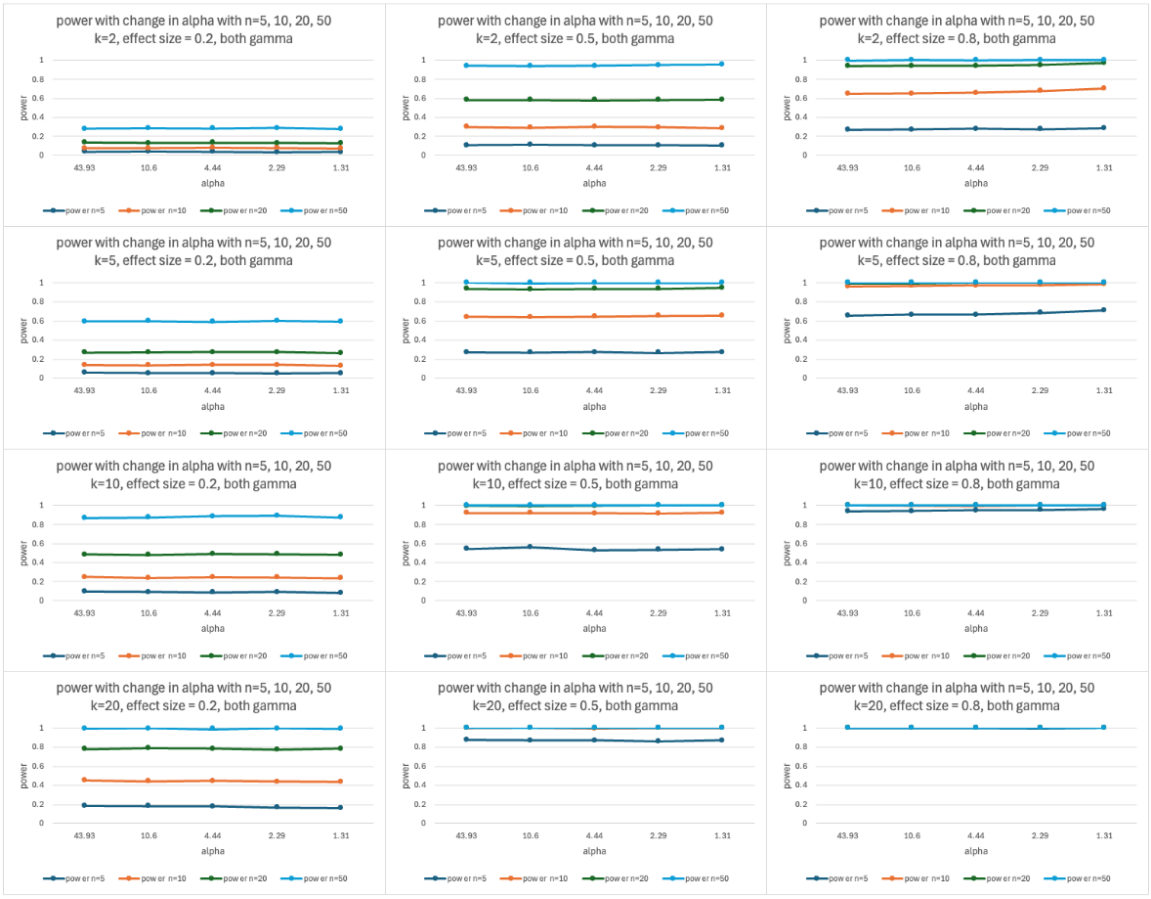

Figure 4.3.34 The power in the test, as a function of α and n across different effect sizes, with simulating two gamma distributions with different parameters using the standardized mean difference for a fixed value of k at all levels of 2, 5, 10, and 20.

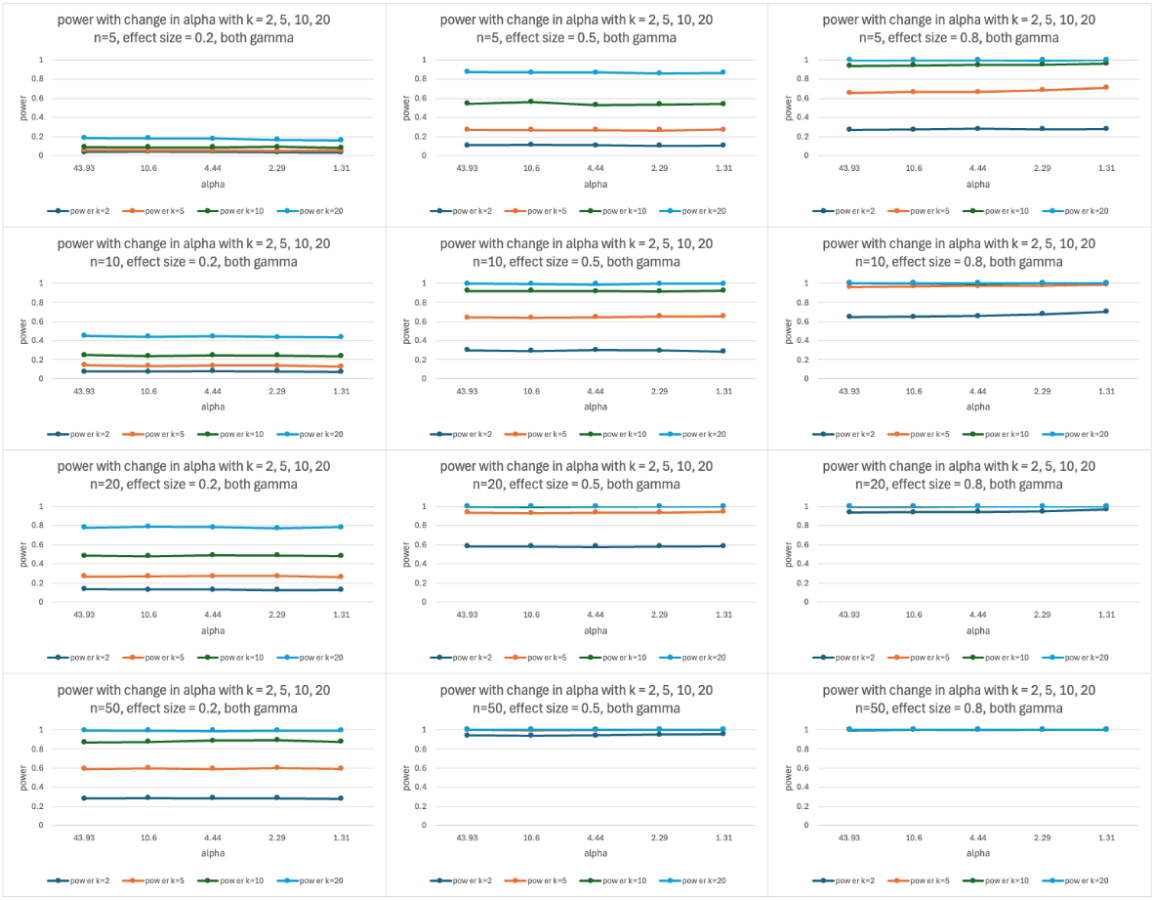

Figure 4.3.35 The power in the test, as a function of α and k across different effect sizes, with simulating two gamma distributions with different parameters using the standardized mean difference for a fixed value of n at all levels of 5, 10, 20, and 50.

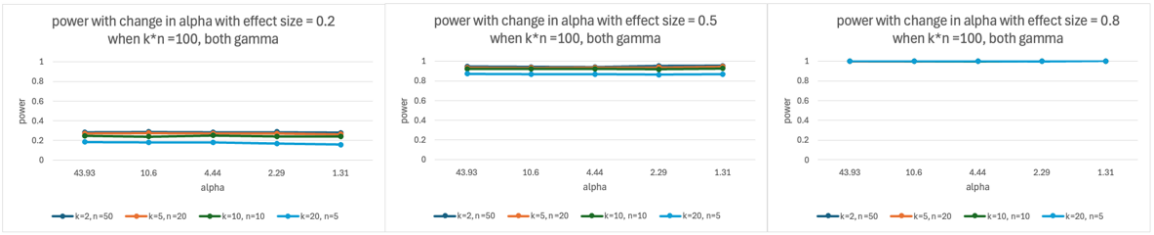

Figure 4.3.36 The power in the test, as a function of α and different combinations of k and n across different effect sizes, with simulating two gamma distributions with different parameters using the standardized mean difference for a fixed total sample size of 200.

4.4 χ^2 Distribution

4.4.1 Mean Difference

As shown in Figure 4.4.1, with a fixed value of k at all levels, the relative bias fluctuates more with a small effect size (d= 0.2) and a smaller sample size per study (n=5, 10). The relative bias is more stable with larger effect sizes (d=0.5, 0.8) and a larger sample size per study (n=20, 50). Similarly, as shown in Figure 4.4.2, with a fixed n of all levels, the relative bias fluctuates more with a small effect size (0.2) and fewer studies (k=2, 5). For both fixed k and n, the relative bias fluctuates less with smaller skewness ($d.f. = 88, 21$). Fewer relative biases of less than 1% are observed with more sample sizes per study and more studies. When the $d.f.$ is 3,5, and 9 and the number of studies is 2, the absolute relative bias is greater than 1% in most cases. With a fixed total sample size of 200, the relative bias fluctuates more with a small effect size, as shown in Figure 4.4.3. Still, there is no noticeable difference in relative bias with different combinations of k and n. When the total sample size is fixed at 200, the relative bias is less than 1.5%.

Figure 4.4.1 The relative bias in the estimated treatment effect, as a function of $d.f.$ and n across different effect sizes, with adding a constant to χ^2 data using the mean difference for a fixed value of k at all levels of 2, 5, 10, and 20.

Figure 4.4.2 The relative bias in the estimated treatment effect, as a function of $d.f.$ and k across different effect sizes, with adding a constant to χ^2 data using the mean difference for a fixed value of n at all levels.

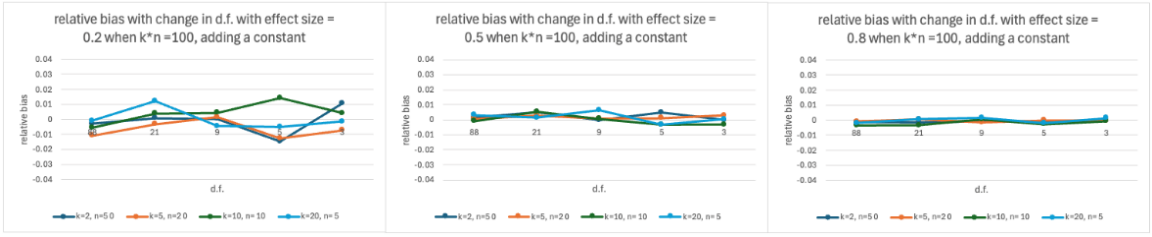

Figure 4.4.3 The relative bias in the estimated treatment effect, as a function of $d.f.$ and different combinations of k and n across different effect sizes, with adding a constant to χ^2 data using the mean difference for a fixed total sample size of 200.

In Figure 4.4.4, with a fixed value of k at all levels, the coverage probability is less than 92.5% when n=5 for all effect sizes and skewness. The coverage probability increases to

the acceptable range as sample sizes per study increase. There is an unimportant trend in coverage probability with increasing skewness. In Figure 4.4.5, with a fixed n of all levels, there is a slight increase in coverage probability with increasing skewness when n=5, 10, 20 for moderate and large effect size. The coverage probability is more stable when skewness increases at n=50. Whenever the coverage probability is outside the acceptable range, n=5 and 10. All coverage probability is less than 0.95 except when k=5, n=10 with a large effect size and the smallest skewness. There is a negligible trend in coverage probability with increasing skewness. With a fixed total sample size of 200, the coverage probability is the lowest, less than 90%, with the smallest sample size per study and the largest number of studies, as presented in Figure 4.4.6. When k=5, n=20, and k=2, n=50, the coverage probability is between 92.5% and 97.5%.

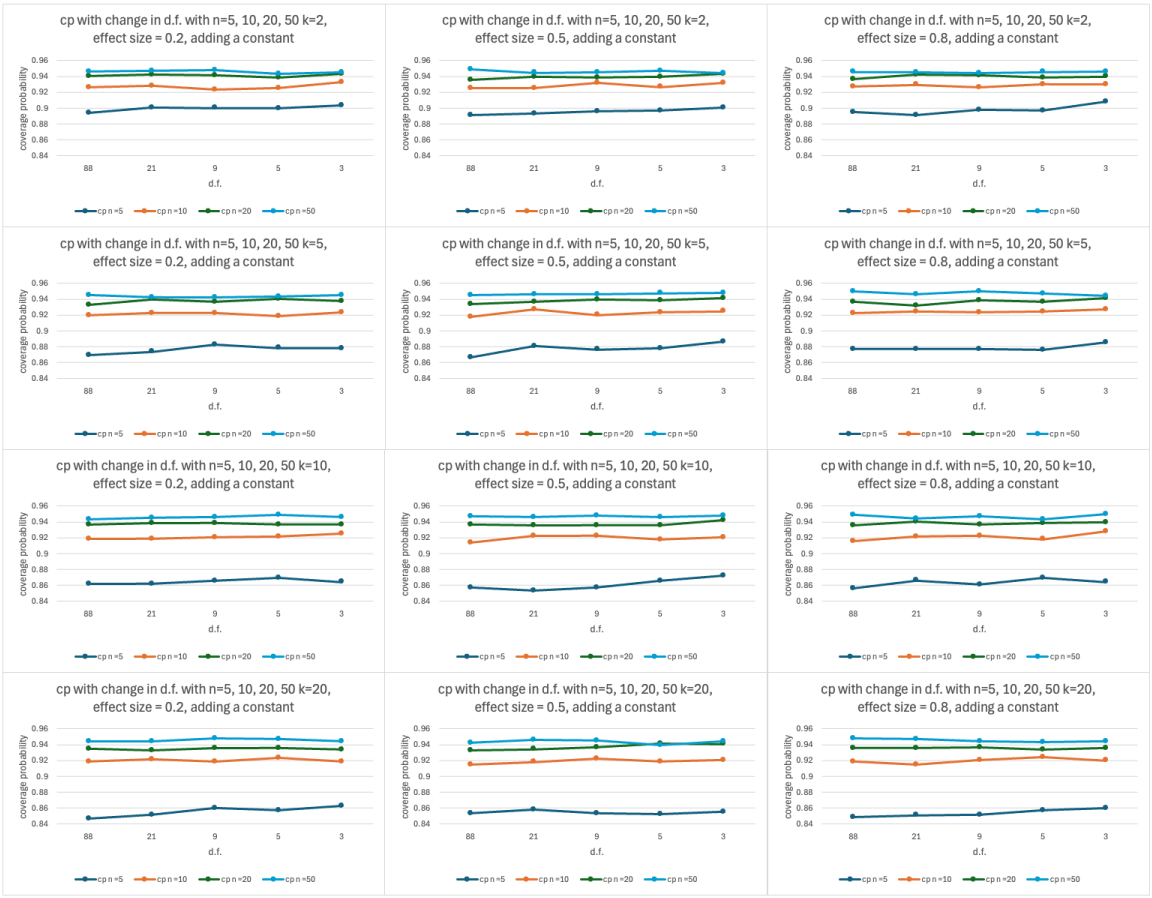

Figure 4.4.4 The coverage probability in the estimated treatment effect, as a function of $d.f.$ and n across different effect sizes, with adding a constant to χ^2 data using the mean difference for a fixed value of k at all levels of 2, 5, 10, and 20.

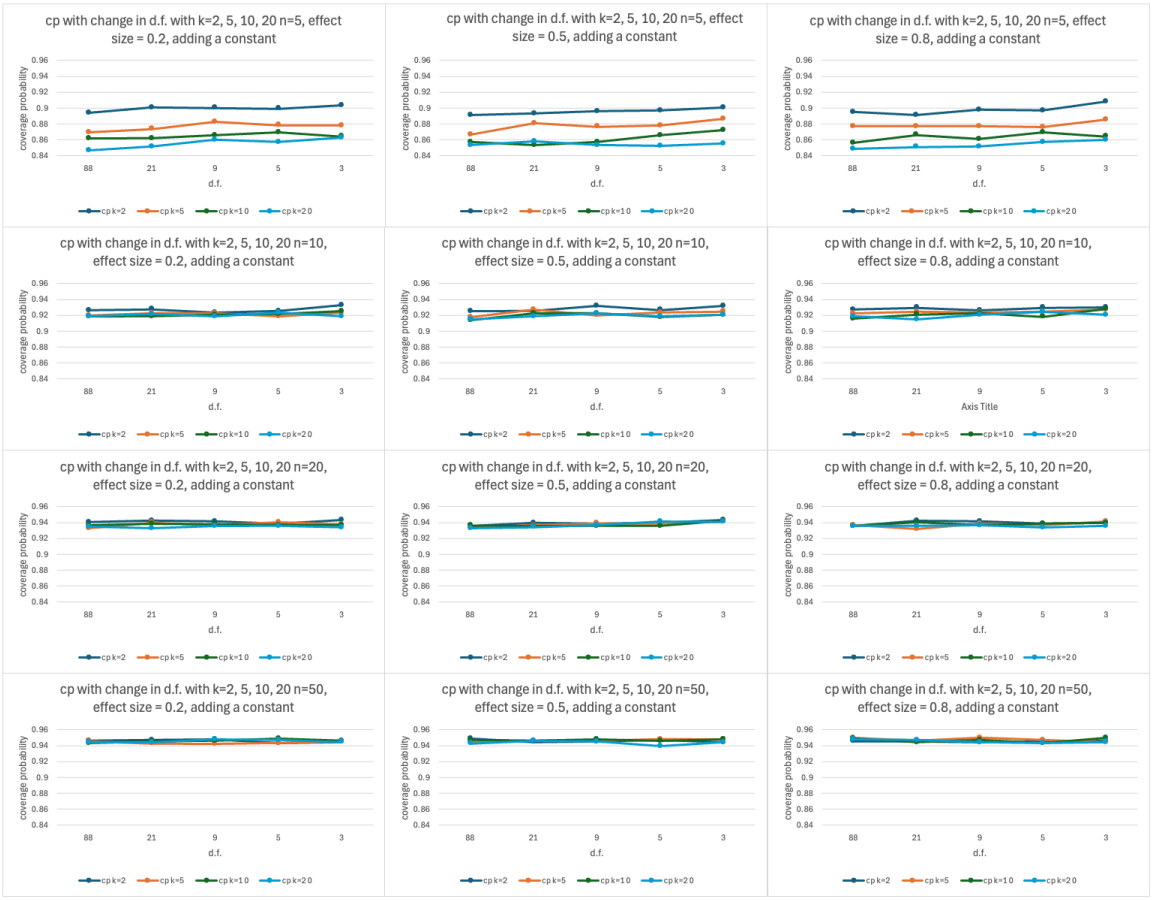

Figure 4.4.5 The coverage probability in the estimated treatment effect, as a function of $d.f.$ and k across different effect sizes, with adding a constant to χ^2 data using the mean difference for a fixed value of n at all levels of 5, 10, 20, and 50.

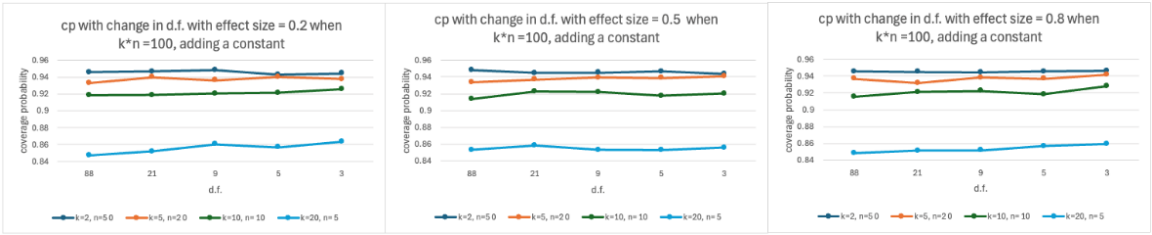

Figure 4.4.6 The coverage probability in the estimated treatment effect, as a function of $d.f.$ and different combinations of k and n across different effect sizes, with adding a constant to χ^2 data using the mean difference for a fixed total sample size of 200.

Figure 4.4.7 and Figure 4.4.8 show that the power increases as the sample size per study, the number of studies and the effect sizes increase. The skewness does not have an important impact on power. With a small effect size, the power is greater than 0.8 when

the total sample size exceeds 800. The power is greater than 0.8 with a moderate effect size when the total sample size exceeds 100. With a large effect size, the power is greater than 0.8 in most cases except when k=2, n=5 and k=2, n=10.

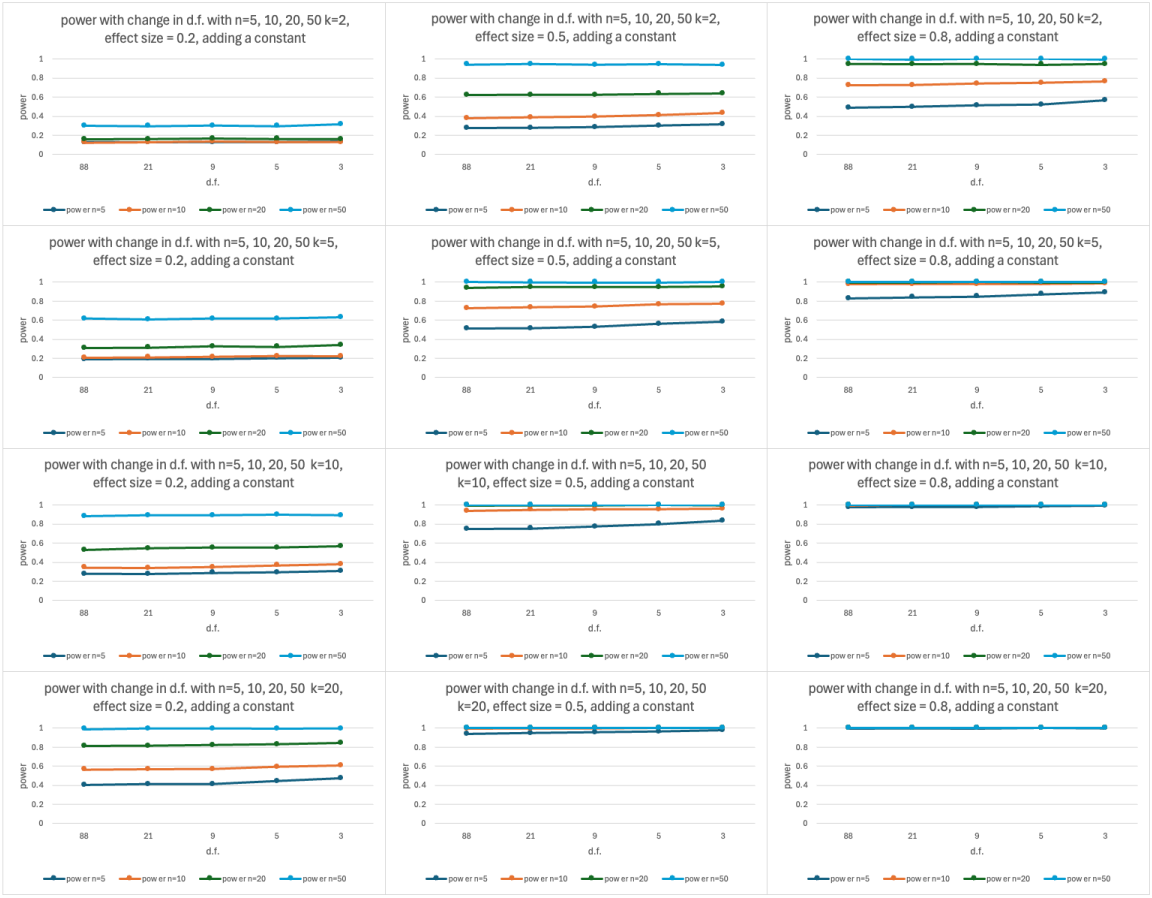

Figure 4.4.7 The power in the test, as a function of $d.f.$ *and n across different effect sizes, with adding a constant to* χ^2 *data using the mean difference for a fixed value of k at all levels of 2, 5, 10, and 20.*

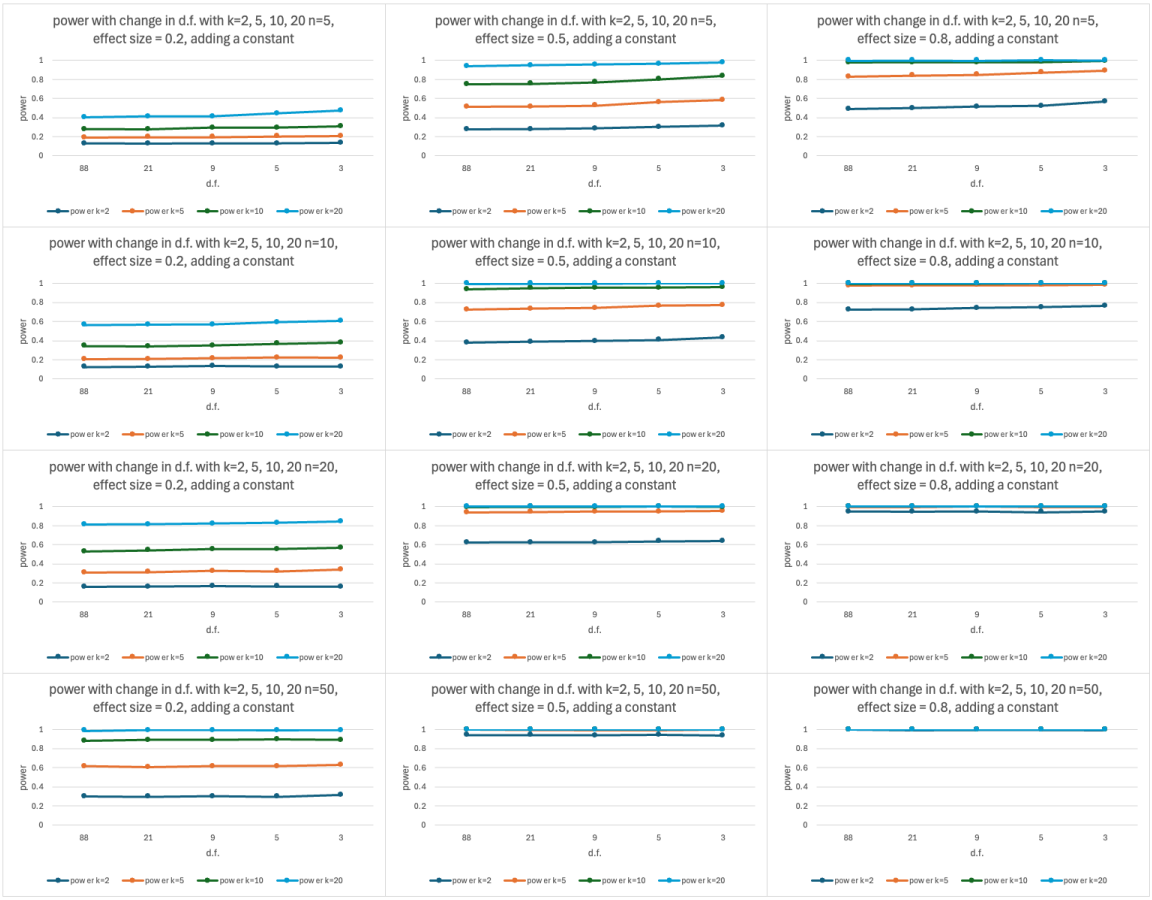

Figure 4.4.8 The power in the test, as a function of $d.f.$ and k across different effect sizes, with adding a constant to χ^2 data using the mean difference for a fixed value of n at all levels of 5, 10, 20, and 50.

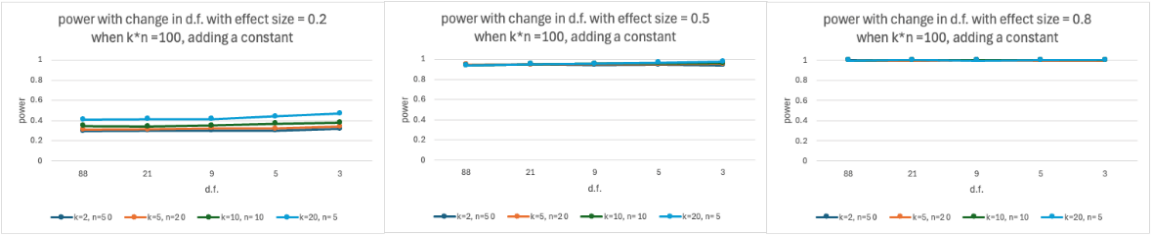

Figure 4.4.9 The power in the test, as a function of $d.f.$ and different combinations of k and n across different effect sizes, with adding a constant to χ^2 data using the mean difference for a fixed total sample size of 200.

The type I error rate is acceptable when k exceeds 2 and n>10. The skewness does not have an important impact on the type I error rate.

4.4.2 Standardized Mean Difference

In Figure 4.4.10, with a fixed k at all levels, the relative bias approaches zero as n increases and skewness decreases. There is no noticeable relationship between the relative bias and the effect size. In Figure 4.4.11, with a fixed n at all levels, the relative bias is closer to zero with a larger k, and when $d.f. = 5$. More fluctuation in relative bias is observed with a small effect size. Minimal fluctuation of the absolute relative bias, within 2%, is found when n = 50. With a fixed total sample size of 200, the relative bias is the closest to zero when k=2, n=50, and farthest from zero when k=20, n=5, as given in Figure 4.4.12. When the skewness is large ($d.f. = 5, 3$), the relative bias is likely to be positive. However, in most situations, the relative bias is negative, showing an underestimation of the treatment effect using SMD.

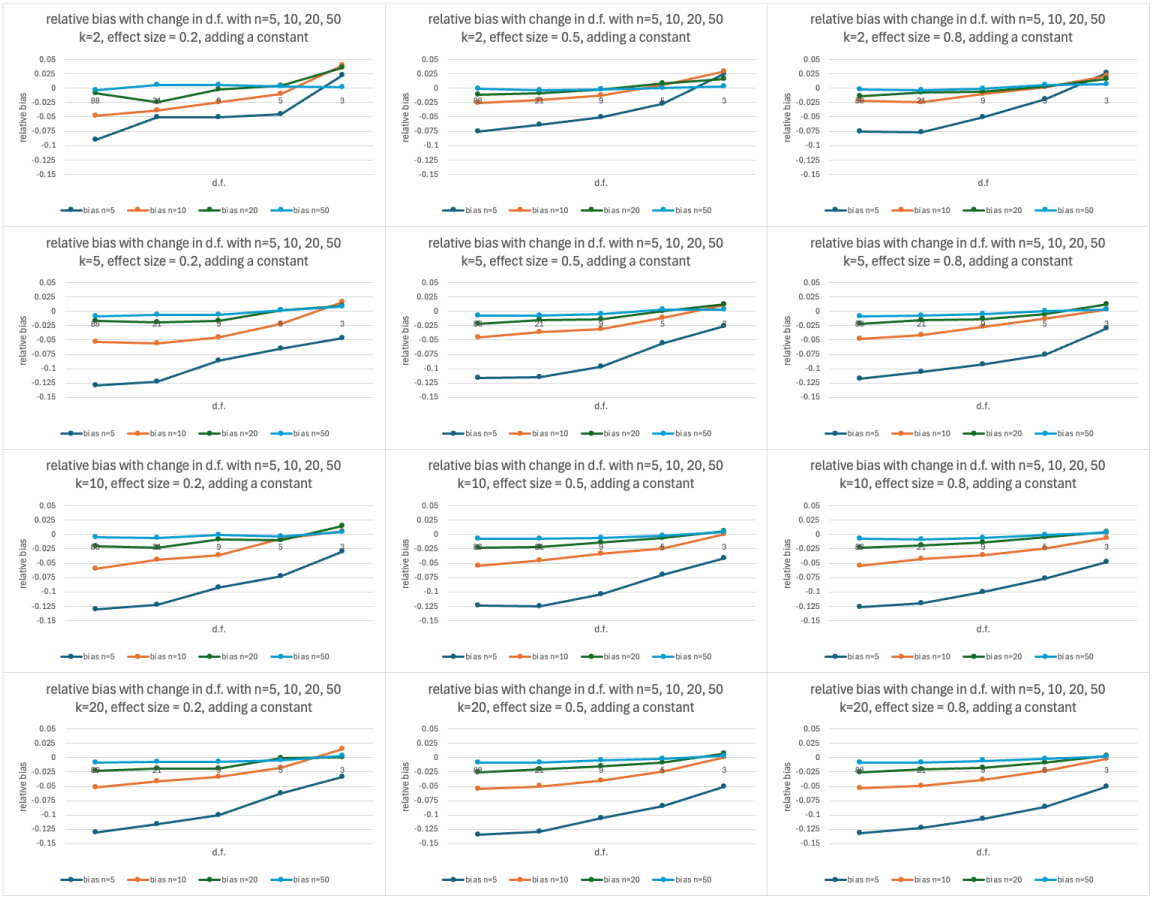

Figure 4.4.10 The relative bias in the estimated treatment effect, as a function of $d.f.$ *and n across different effect sizes, with adding a constant to* χ^2 *data using the standardized mean difference for a fixed value of k at all levels of 2, 5, 10, and 20.*

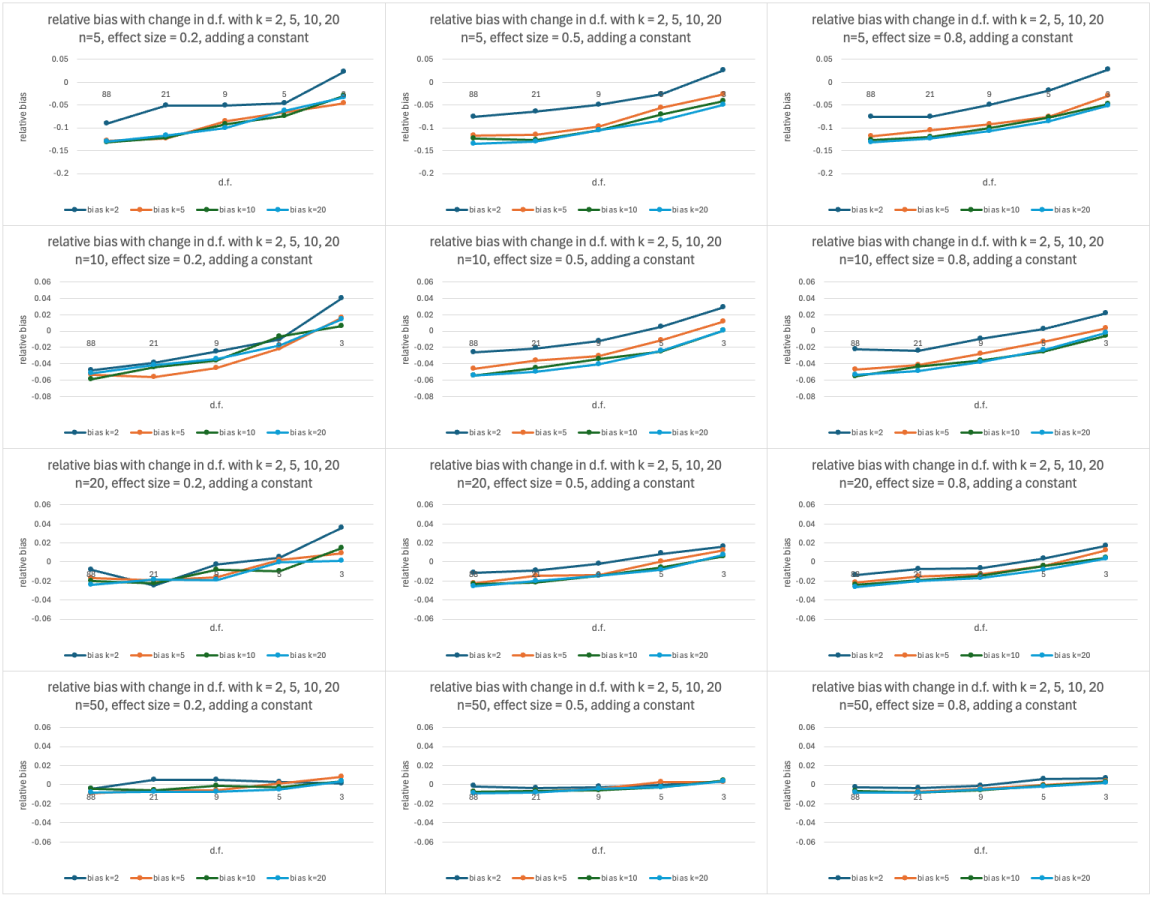

Figure 4.4.11 The relative bias in the estimated treatment effect, as a function of $d.f.$ *and k across different effect sizes, with adding a constant to* χ^2 *data using the standardized mean difference for a fixed value of n at all levels of 5, 10, 20, and 50.*

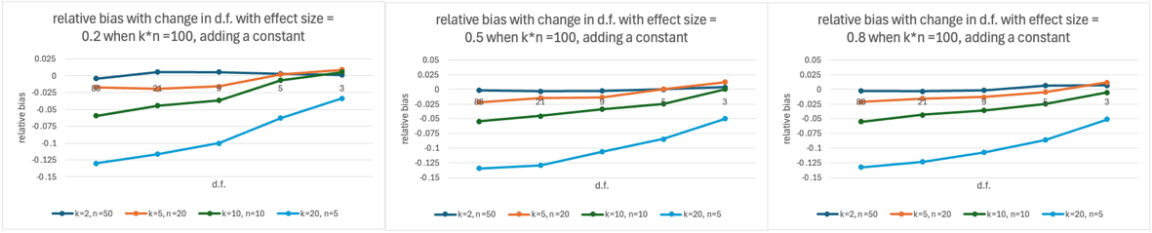

Figure 4.4.12 The relative in the estimated treatment effect, as a function of $d.f.$ *and different combinations of k and n across different effect sizes, with adding a constant to* χ^2 *data using the standardized mean difference for a fixed total sample size of 200.*

Figure 4.4.13 indicates that when k is fixed at all levels, the coverage probability

increases with skewness for n=5. However, for n=20 and n=50, the coverage probability

decreases with skewness when k=10 and k=20, particularly for moderate to large effect sizes. There is a worse coverage probability with smaller n, as the coverage probability exceeds 0.97, while improved coverage is noted with larger n, where the coverage probability is around 0.95. As presented in Figure 4.4.14, the coverage probability rises with increased skewness when k=20 and n=5 for both moderate and large effect sizes and when k=20 and n=10 with a large effect size. Conversely, the coverage probability decreases as skewness increases for n=20 and 50 with a large effect size. In addition, for a fixed value of n, the largest k results in a lower coverage probability for moderate and large effect sizes. Nonetheless, k does not have a noticeable impact on the coverage probability for a small effect size across all n values, nor does it affect coverage probability with moderate and large effect sizes as long as the sample size is efficiently large. There are no situations where the coverage probability is less than the lower bound of the acceptable range. In Figure 4.4.15, with a fixed total sample size, the coverage probability is between 0.94 and 0.97 for all effect sizes and combinations of n and k. More variability of coverage probability with more skewness ($d.f.= 9, 5, 3$). Effect sizes do not directly influence the coverage probability when the total sample size is fixed.

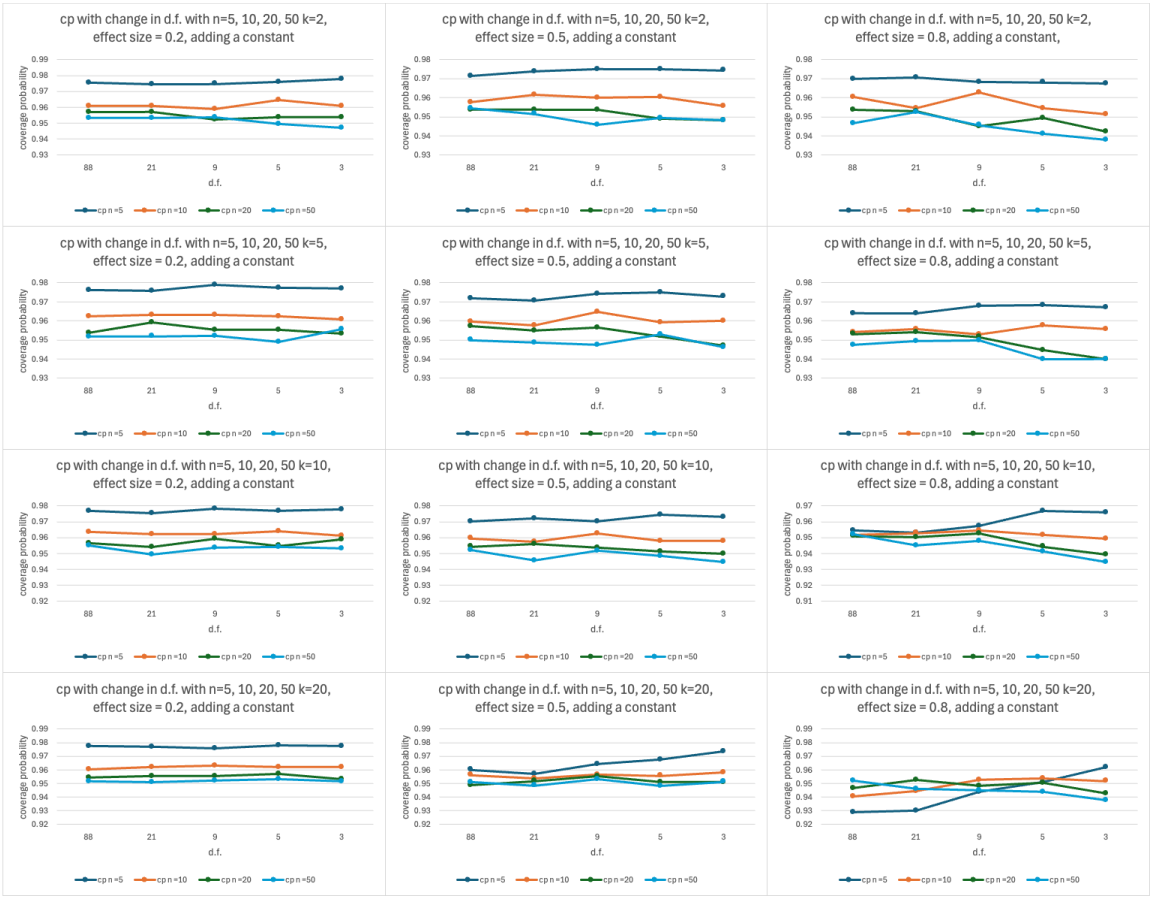

Figure 4.4.13 The coverage probability in the estimated treatment effect, as a function of $d.f.$ and n across different effect sizes, with adding a constant to χ^2 data using the standardized mean difference for a fixed value of k at all levels of 2, 5, 10, and 20.

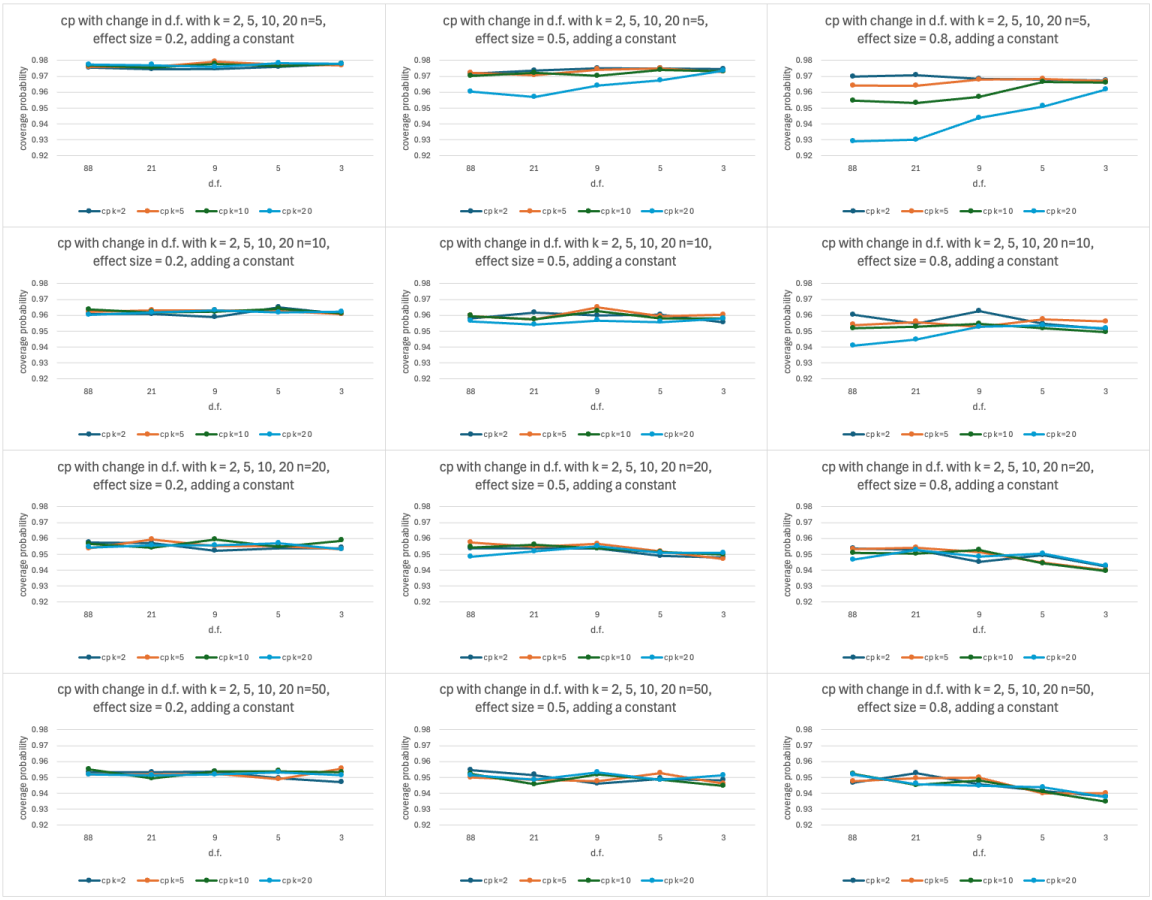

Figure 4.4.14 The coverage probability in the estimated treatment effect, as a function of $d.f.$ and k across different effect sizes, with adding a constant to χ^2 data using the standardized mean difference for a fixed value of n at all levels of 5, 10, 20, and 50.

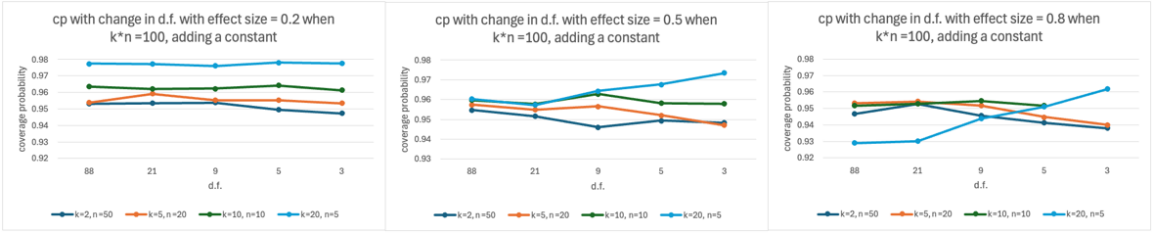

Figure 4.4.15 The coverage probability in the estimated treatment effect, as a function of $d.f.$ and different combinations of k and n across different effect sizes, with adding a constant to χ^2 data using the standardized mean difference for a fixed total sample size of 200.

More sample sizes per study, more studies and larger effect sizes lead to greater power, as shown in Figure 4.4.16 and Figure 4.4.17. There is no noticeable trend of power with

changing skewness. The power is greater than 0.8 when the total sample size exceeds 800 for a small effect size. The power is greater than 0.8 when the total sample size exceeds 200 for a moderate effect size. The power is greater than 0.8 when the total sample size exceeds 80 for a large effect size. In Figure 4.4.18, with a fixed total sample size of 200, the power is lowest when k is the largest and n is the lowest for small and moderate effect sizes.

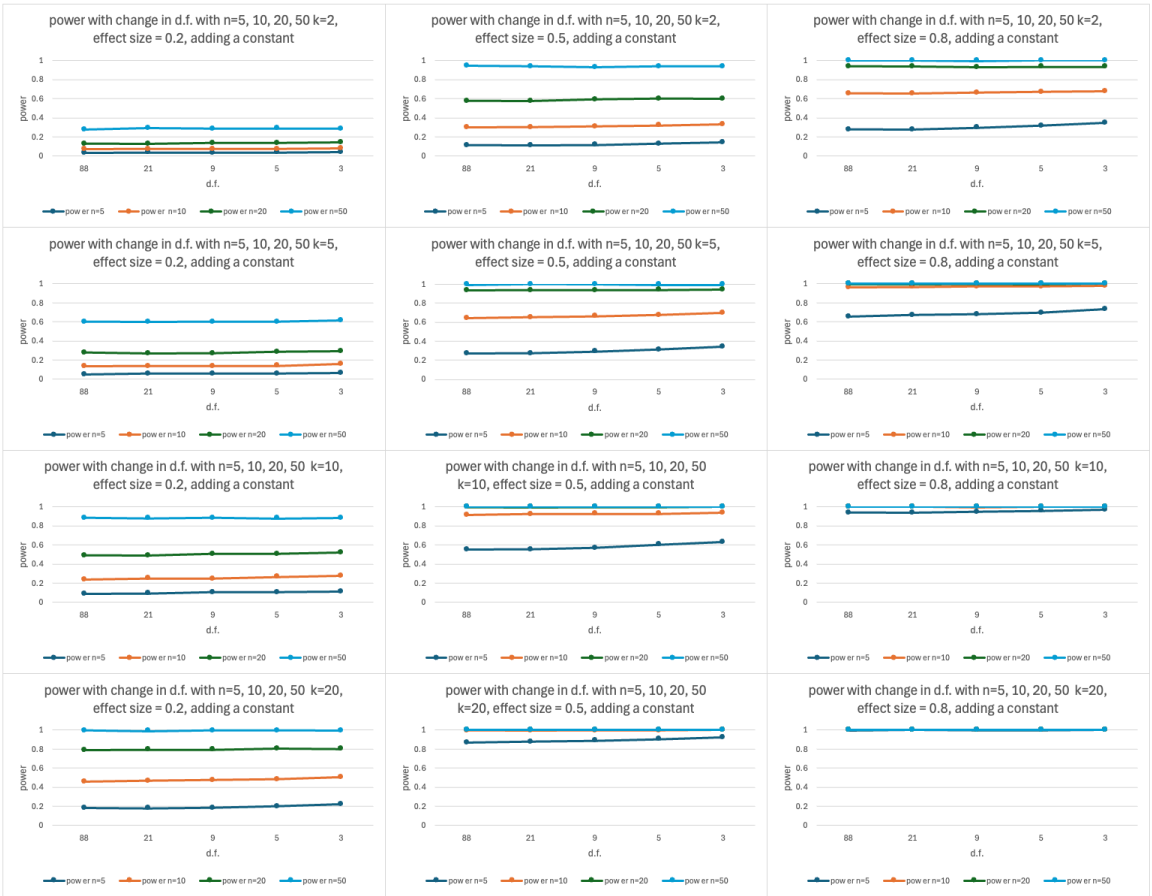

Figure 4.4.16 The power in the test, as a function of $d.f.$ and n across different effect sizes, with adding a constant to χ^2 data using the standardized mean difference for a fixed value of k at all levels of 2, 5, 10, and 20.

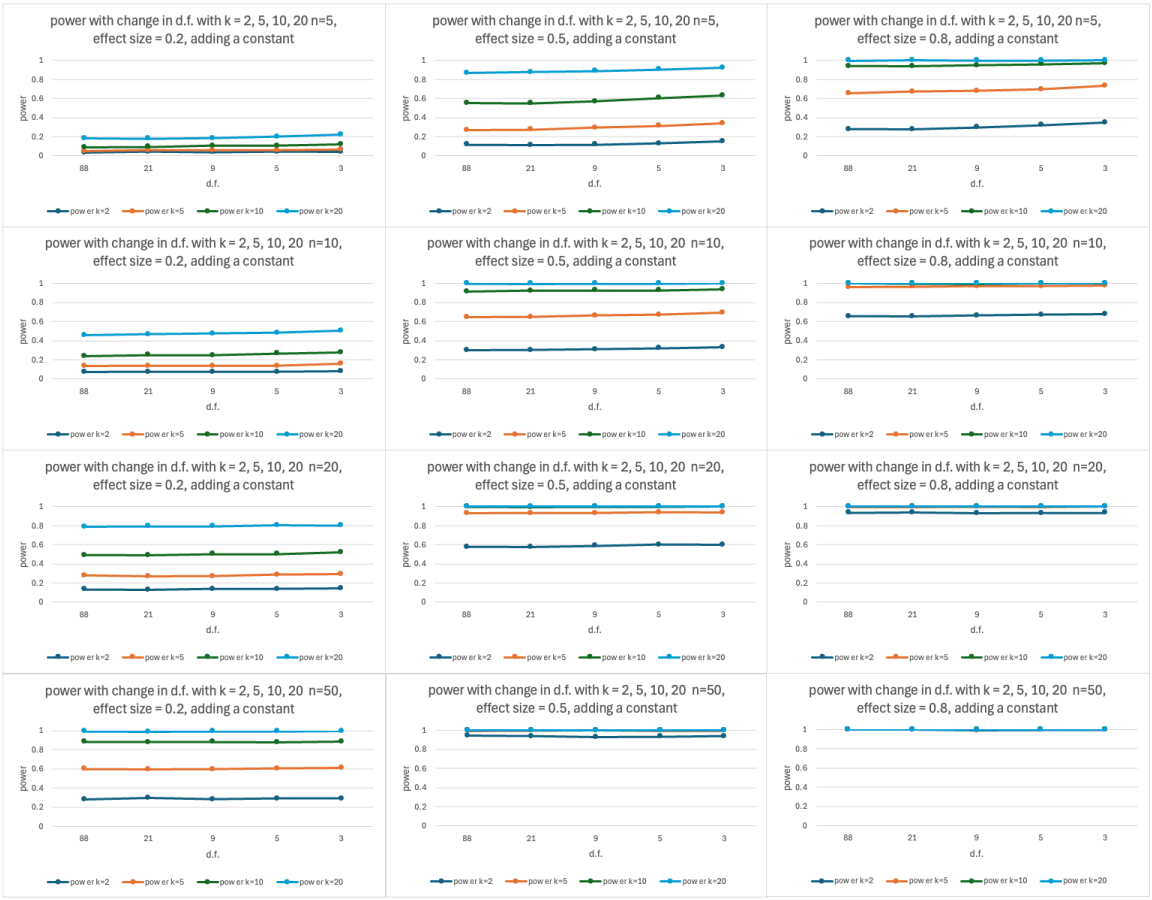

Figure 4.4.17 The power in the test, as a function of $d.f.$ *and k across different effect sizes, with adding a constant to* χ^2 *data using the standardized mean difference for a fixed value of n at all levels of 5, 10, 20, and 50.*

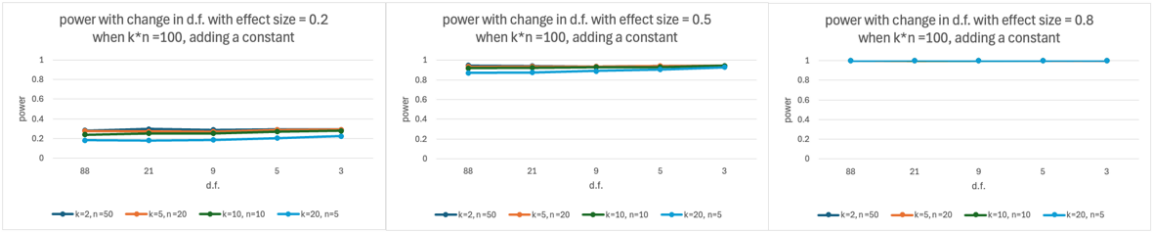

Figure 4.4.18 The power in the test, as a function of $d.f.$ *and different combinations of k and n across different effect sizes, with adding a constant to* χ^2 *data using the standardized mean difference for a fixed total sample size of 200.*

When k=2, n=5, 10, the type I error rate exceeds 0.075. When n=5, and k=5, 10, 20, the type I error rate is below 0.025.

Chapter 5 Conclusion and Discussion

5.1 Summary of the Results

A central finding from the simulation is that larger sample sizes per group lead to a more robust overall estimated effect against non-normality, even when only a few studies are included. This results in improved relative bias, coverage probability of the overall estimated effect, power, and type I error rate of the meta-analysis. Another important finding is that in most situations, more studies with a smaller sample size per group lead to worse relative bias, coverage probability and power if the total sample size is fixed. Furthermore, the effects of skewness and effect sizes vary among different distributions of the outcome variable and methods of considering the effect size. The relative bias of the overall estimated effect is mostly negative, particularly with smaller skewness when using the standardized mean difference. Moreover, when using the standardized mean difference, the absolute relative bias is bigger than the mean difference in most cases across all distributions, which is consistent with Lin's study (2018), and this suggests that one should use the mean difference when possible. In addition, a larger relative bias of SMD may result from the inappropriate choice of SDs (standard deviation), according to Hopkins and Rolands (2024). They also stated that using inappropriate SDs, for example, failing to remove technical errors, can introduce heterogeneity, which makes the fixed-effect model less appropriate. The coverage probability is below the nominal level (95%) when using the mean difference and exceeds the nominal level with small sample sizes

per group in most cases when using the standardized mean difference. Additionally, at least a total sample size of 800 is required to have a power greater than 0.8 for a small effect size ($d = 0.2$), and the total sample size can be decreased when the effect size increases. However, the estimation of the power is biased whenever the type I error rate is not at the nominal level, which is 0.05 in our study, as well as due to the bias from the overall effect estimation. The type I error rate of the meta-analysis is generally unaffected by skewness for most of the distributions we examined, except for the gamma distribution.

5.1.1 The Mixture of Normal Distributions

Non-robust results are observed for the mixture of normal distributions when a large number of studies are included with a sample size per group of less than 10, or with a large proportion of non-responders in the intervention group, or with a small effect size. With a fixed total sample size, fewer studies with more samples per group are preferred as they provide a better relative bias, coverage probability and power. When using the mean difference, the sample size per group has a greater effect than the number of studies on the coverage probability. When n < 10, coverage probability is less than 0.925. Power drops sharply when more than half of the sample in the intervention group are non-responders. The type I error rate is generally acceptable except when the sample size per study is 5 and 10. When the sample size per group is 5, a smaller k gives a better relative bias as it is the closest to zero. When the sample size per group is 5, the worst relative bias and unacceptable coverage probability are observed, given the acceptable range for the coverage probability is between 0.925 and 0.975. In addition, the type I error rate is

not between 0.025 and 0.075 when the sample size per group is 5 and the number of studies is greater than 2.

5.1.2 Lognormal Distribution

More samples per study and fewer studies provide better relative bias and coverage probability for lognormal distributions. In addition, the coverage probability and the type I error are unacceptable when the sample size per group is 5. When simulating two same distributions, adding a constant to one group and using the mean difference, a negligible effect of skewness is observed on the absolute relative bias, which remains below 3% and improves with larger effect sizes but moves closer to zero as the skewness increases with a small effect size for the standardized mean. The coverage probability increases as the number of studies decreases for small n. There is a slight increase in power as skewness grows with the sample size per group at 5; as mentioned before, this increase could result from the bias of the overall effect estimation and an incorrect type I error. The type I error rate is above 0.05 in all situations when using the mean difference and is less than 0.025 when the sample size per group is 5 and less than the number of studies with the standardized mean difference. When simulating two lognormal distributions with different μ_s, the relative bias moves away from zero as n decreases and skewness increases for the mean difference and moves towards zero as skewness increases for the standardized mean difference. Fewer studies give better relative bias for a fixed value of a small sample size per group. Coverage probability moves away from higher skewness and larger k for moderate and large effect sizes.

5.1.3 Gamma Distribution

For gamma distributions, larger sample sizes per group and fewer studies give better relative bias and coverage probability, and the coverage probability is outside the acceptable range with small samples per study. When simulating two same distributions and adding a constant to one group, the relative bias and the coverage probability increase as k decreases with a fixed sample size per group of 5. When using the mean difference, a small sample size (n=5,10, and k=2) shows greater fluctuation of the relative bias with the change in skewness. when using the mean difference. Power increases as skewness increases. In addition, the type I error rate decreases as the skewness increases. When using the standardized mean difference, the relative bias is only positive with a skewness greater than 1. The coverage inflates when the skewness increases for larger effect sizes with small samples per study. When only two studies with a sample size per group of 5 and 10, the type I error rate is unacceptable. When simulating two gamma distributions with different β_s and using the mean difference, more studies introduce a greater absolute relative bias with a fixed sample size per group, and it can reach 30% with highly skewed outcome variables. The coverage probability increases with skewness and different effect sizes. Moreover, when skewness is greater than 1, the coverage probability is unacceptable. There is a slight decrease in power as skewness increases with a small sample size per group. When using the standardized mean difference, the relative bias moves farther from zero as skewness increases with group size of 5 and 10 for small and moderate effect sizes with 10 and 20 studies. The coverage probability is

robust around 0.95 against the skewness with a small effect size and is robust when the skewness is less than 1 for moderate and large effect sizes.

5.1.4 χ^2 Distribution

For χ^2 distribution, when using the mean difference, the relative bias fluctuates more with increasing skewness with a small effect size, a smaller sample size per study and fewer studies. Whenever the coverage probability is outside the acceptable range, the sample size per study is 5 and 10. There is a negligible trend in coverage probability with increasing skewness when using the mean difference. However, when using the standardized mean difference, the coverage probability behaves oppositely when the sample size is small or large, respectively, and more studies result in a lower coverage probability for moderate and large effect sizes with a fixed sample size per group. The skewness does not have an important impact on power. The type I error rate is unacceptable when the sample size per group is less than 5.

The above findings suggest that investigators should preferably include large trials when conducting a meta-analysis, as our simulation indicates that fewer studies with larger sample sizes yield more robust results, even when the primary data are highly skewed. If large trials are uncommon in the research topic, the investigator should seek to obtain the raw data and assess the normality of the dataset, especially when dealing with data with a lower bound, such as volume distribution, blood concentration, and scale outcomes (Deeks, Higgins, & Group, 2022). If the data is skewed, transforming the primary data, such as through log transformation, is recommended. However, if investigators cannot

obtain the raw data, considering a nonparametric meta-analysis method (Michiels & Onghena, 2018) or converting data from a continuous to a binary outcome by setting a threshold can be an option.

5.2 Limitations

A few limitations of this simulation can be acknowledged. In this simulation, the sample size is the same across two groups in each study. However, in real life, there are situations where the ratio of sample sizes of the control and intervention can be 2:1 or 1:2, for instance. Moreover, we only investigate situations where the studies are of the same size in each meta-analysis, which may not be realistic. Also, using the standardized mean difference is not the most appropriate in cases where the variance of the two groups is not the same, such as simulating two lognormal and gamma distributions with different parameters, as the pooled standard deviation method used in the primary study assumes the variance of the two groups are the same. In addition, we assume the distribution of the outcome variables for two groups are the same, except when examining the mixture of normal distributions, in which case the control group follows a normal distribution, and the intervention group follows a mixture of normal distributions. Moreover, in practice, obtaining the raw data from each study may be challenging since they usually provide summary statistics, and assessing normality from the summary statistics can be difficult. Another limitation of this simulation is that the random-effects model is not included.

Therefore, these limitations suggest several directions for future research on the fixed-effect meta-analysis model. Firstly, we could examine more complex scenarios where the

sample size per group within the same study and the sample size across studies vary. Secondly, when using the standardized mean difference as the effect size, we could evaluate whether the variances of the two groups within the same study vary and then select the appropriate variance pooling methods accordingly. Thirdly, we could investigate more cases where the two groups in a study follow different distributions and studies with various distributions. Moreover, more distributions, such as multinomial and negative binomial distributions, could be assessed. Also, we could simulate the effect size for each study directly instead of using the raw data, which may more accurately reflect reality when conducting a meta-analysis. Lastly, different random-effects models can be investigated in future studies with this simulation framework.

5.3 Conclusion

To conclude, small studies with highly skewed data provide non-robust meta-analysis results for a fixed-effect model. Moreover, when conducting meta-analyses, larger sample sizes per study with fewer studies are preferred, compared to having a smaller sample size per study with more studies. Therefore, this simulation suggests that investigators need to be cautious with the distribution of the raw data when conducting meta-analysis using the fixed-effect model.

Reference

Bender, T. T., Leyens, J., Sellin, J., Kravchenoko, D., Conrad, R., Mucke, M., & Seidel, M. F. (2020). Therapeutic options for patients with rare rheumatic diseases: a systematic review and meta-analysis. *Orphanet Journal of Rare Diseases, 15*, https://doi.org/10.1186/s13023-020-01576-5.

Blanca, M. J., Arnau, J., López-Montiel, D., Bono, R., & Bendayan, R. (2013). Skewness and kurtosis in real data samples. *Methodology: European Journal of Research Methods for the Behavioral and Social Science, 9*(2), 78-84. https://doi.org/10.1027/1614-2241/a000057.

Bono, R., Blanca, M., Arnau, J., & Gómez-Benito, J. (2017). Non-normal Distributions Commonly Used in Health, Education, and Social Sciences: A Systematic Review. *Frontiers in Psychology, 8*, https://doi.org/10.3389/fpsyg.2017.01602 .

Borenstein, M. (2009). Effect Size for Continuous Data. In H. Cooper, L. V. Hedges, & J. C. Valentine, *The Handbook of Research Synthesis and Meta-Analysis. 2ND Edition.* New York, New York: Russell Sage Foundation.

Borenstein, M., Hedges, L. V., Higgins, J. P., & Rothstein, H. R. (2009). *Introduction to Meta-Analysis.* John Wiley & Sons, Ltd.

Bradley, J. V. (1978). Robustness? *British Journal of Mathematical and Statistical Psychology, 31*(2), 144-152. https://doi.org/10.1111/j.2044-8317.1978.tb00581.x.

Cohen, J. (1988). *Statistical Power Analysis for the Behavioral Sciences.* New York: Lawrence Erlbaum Associates.

Davey, J., Turner, R. M., Clarke, M. J., & Higgins, J. P. (2011). Characteristics of meta-analyses and their component studies in the Cochrane Database of Systematic Reviews: a cross-sectional, descriptive analysis. *BMC Medical Research Methodology, 11*, https://doi.org/10.1186/1471-2288-11-160.

Deeks, J. J., Higgins, J. P., & Group, C. S. (2022). *Statistical Algorithms in Review Manager.*

Hedges, L. V., & Vevea, J. L. (1998). Fixed- and random-effects models in meta-analysis. *Psychological Methods, 3*(4), 486-504. https://doi.org/10.1037/1082-989x.3.4.486. Retrieved from Psychological Methods: https://doi.org/10.1037/1082-989x.3.4.486

Higgins, J. P., Thomas, J., Chandler, J., Cumpston, M., Li, T., Page, M., & Welch, V. (2024). *Cochrane Handbook for Systematic Reviews of Interventions version 6.5 (updated August 2024).* Cochrane.

Higgins, J. P., White, I. R., & Anzures-Cabrera, J. (2008). Meta-analysis of skewed data: Combining results reported on log-transformed or raw scales. *Statistics in Medicine, 27*(29), 6072-6092. https://doi.org/10.1002/sim.3427.

Hopkins, W. G., & Rowlands, D. S. (2024). Standardization and other approaches to meta-analyze differences in means. *Statistics in Medicine, 43*(16), 3092–3108. https://doi.org/10.1002/sim.10114.

Jackson, D., & White, I. R. (2018). When should meta-analysis avoid making hidden normality assumptions? *Biometrical Journal*, 1040-1058. https://doi-org.libaccess.lib.mcmaster.ca/10.1002/bimj.201800071.

Johnson, N. L., Kotz, S., & Balakrishnan, N. (1994). *Continuous Univariate Distributions, Volume 1, 2nd Edition.* Wiley-Interscience.

Kelley, K. (2005). The effects of nonnormal distributions on confidence intervals around the standardized mean difference: bootstrap and parametric confidence intervals. *Educational and Psychological Measurement, 65*, 51-69. https://doi.org/10.1177/0013164404264850.

Kontopantelis, E., & Reeves, D. (2010). Performance of statistical methods for meta-analysis when true study effects are non-normally distributed: A simulation study. *Statistical Methods in Medical Research, 21*(4), 409-426. https://doi.org/10.1177/0962280210392.

Lin, L. (2018). Bias caused by sampling error in metaanalysis with small sample sizes. *PLoS ONE, 13*(9), https://doi.org/ 10.1371/journal.pone.0204056.

Mellen, P. B., & Herrington, D. M. (2005). Pharmacogenomics of blood pressure response to antihypertensive treatment. *Journal of Hypertension, 23*(7), 1311-1325. https://doi.org/10.1097/01.hjh.0000173510.52987.68.

Micceri, T. (1989). The Unicorn, The Normal Curve, and Other Improbable Creatures. *Psychology Bulletein, 105*(1), 156-166. https://doi.org/10.1037/0033-2909.105.1.156.

Michiels, B., & Onghena, P. (2018). Nonparametric meta-analysis for single-case research: Confidence intervals for combined effect sizes. *Behavior Research Methods, 51*(3), 1145-1160. https://doi.org/10.3758/s13428-018-1044-5.

Pequeno, N. P., De Araújo Cabral, N. L., Marchioni, D. M., Lima, S. C., & De Oliveira Lyra, C. (2020). Quality of life assessment instruments for adults: a systematic review of population-based studies. *Health and Quality of Life Outcomes, 18*, https://doi.org/10.1186/s12955-020-01347-7.

Plieninger, T., Hui, C., Gaertner, M., & Huntsinger, L. (2014, May 27). *The Impact of Land Abandonment on Species Richness and Abundance in the Mediterranean Basin: A Meta-Analysis.* Retrieved from PLoS ONE 9(5): https://doi.org/10.1371/journal.pone.0098355

Roberts, N. P., Kitchener, N. J., Kenardy, J., Robertson, L., Lewis, C., & Bisson, J. I. (2019). Multiple session early psychological interventions for the prevention of post-traumatic stress disorder.

Sun, R. W., & Cheung, S. F. (2020). The influence of nonnormality from primary studies on the standardized mean difference in meta-analysis. *Behavior Research Methods, 52*(4), 1552-1567. https://doi.org/10.3758/s13428-019-01334-x.

Takeshima, N., Sozu, T., Tajika, A., Ogawa, Y., Hayasaka, Y., & Furukawa, T. (2014). Which is more generalizable, powerful and interpretable in meta-analyses, mean difference or standardized mean difference? *BMC Medical Research Methodology, 14*, https://doi.org/10.1186/1471-2288-14-30.

Weeks, G., George, J., Maclure, K., & Stewart, D. (2016). Non-medical prescribing versus medical prescribing for acute and chronic disease management in primary and secondary care. *Cochrane Database of Systematic Reviews*(11). Retrieved from Cochrane Database of Systematic Reviews.

Wilcox, R. R. (2005). *Introduction to Robust Estimation and Hypothesis Testing. 3rd Edition.* Elsevier.

Wilcox, R. R., & Keselman, H. J. (2003). Modern Robust Data Analysis Methods: Measures of Central Tendency. *Psychological Methods, 8*(3), 254-274. https://doi.org/10.1037/1082-989x.8.3.254.

Yamada, S., Gotoh, T., Nakashima, Y., Kayaba, K., Ishikawa, S., Nago, N., . . . Kajii, E. (2001). Distribution of Serum C-Reactive Protein and Its Association with Atherosclerotic Risk Factors in a Japanese Population. *American Journal of Epidemiology, 153*(12), 1183-1190. https://doi.org/10.1093/aje/153.12.1183.

Appendix

The appendix provides a summary table of all the algebra for different cases within the same distribution of the outcome variables in terms of the skewness, parameters, and corresponding true effect for mean difference and the standardized mean difference. Also, the type I error rate results for different distributions of the outcome variable with different skewness, the number of studies, and the number of individuals per group are provided.

Distribution	Skewness	Effect Size	Parameters	Formulas of the True Effect for MD	True MD	True SMD
the mixture of normals	NA	0.2	p=0	$(1-p)*(\mu_2-\mu_1)$	0.2	0.2
the mixture of normals	NA	0.2	p=0.1	$(1-p)*(\mu_2-\mu_1)$	0.18	0.18
the mixture of normals	NA	0.2	p=0.2	$(1-p)*(\mu_2-\mu_1)$	0.16	0.16
the mixture of normals	NA	0.2	p=0.3	$(1-p)*(\mu_2-\mu_1)$	0.14	0.14
the mixture of normals	NA	0.2	p=0.4	$(1-p)*(\mu_2-\mu_1)$	0.12	0.12
the mixture of normals	NA	0.2	p=0.5	$(1-p)*(\mu_2-\mu_1)$	0.1	0.1
the mixture of normals	NA	0.2	p=0.6	$(1-p)*(\mu_2-\mu_1)$	0.08	0.08

the mixture of normals	NA	0.2	p=0.7	$(1-p)*(\mu_2-\mu_1)$	0.06	0.06
the mixture of normals	NA	0.2	p=0.8	$(1-p)*(\mu_2-\mu_1)$	0.04	0.04
the mixture of normals	NA	0.2	p=0.9	$(1-p)*(\mu_2-\mu_1)$	0.02	0.02
the mixture of normals	NA	0.2	p=1	$(1-p)*(\mu_2-\mu_1)$	0	0
the mixture of normals	NA	0.5	p=0	$(1-p)*(\mu_2-\mu_1)$	0.5	0.5
the mixture of normals	NA	0.5	p=0.1	$(1-p)*(\mu_2-\mu_1)$	0.45	0.45
the mixture of normals	NA	0.5	p=0.2	$(1-p)*(\mu_2-\mu_1)$	0.4	0.4
the mixture of normals	NA	0.5	p=0.3	$(1-p)*(\mu_2-\mu_1)$	0.35	0.35
the mixture of normals	NA	0.5	p=0.4	$(1-p)*(\mu_2-\mu_1)$	0.3	0.3
the mixture of normals	NA	0.5	p=0.5	$(1-p)*(\mu_2-\mu_1)$	0.25	0.25
the mixture of normals	NA	0.5	p=0.6	$(1-p)*(\mu_2-\mu_1)$	0.2	0.2
the mixture of normals	NA	0.5	p=0.7	$(1-p)*(\mu_2-\mu_1)$	0.15	0.15
the mixture of normals	NA	0.5	p=0.8	$(1-p)*(\mu_2-\mu_1)$	0.1	0.1
the mixture of normals	NA	0.5	p=0.9	$(1-p)*(\mu_2-\mu_1)$	0.05	0.05

the mixture of normals	NA	0.5	p=1	$(1-p)*(\mu_2-\mu_1)$	0	0
the mixture of normals	NA	0.8	p=0	$(1-p)*(\mu_2-\mu_1)$	0.8	0.8
the mixture of normals	NA	0.8	p=0.1	$(1-p)*(\mu_2-\mu_1)$	0.72	0.72
the mixture of normals	NA	0.8	p=0.2	$(1-p)*(\mu_2-\mu_1)$	0.64	0.64
the mixture of normals	NA	0.8	p=0.3	$(1-p)*(\mu_2-\mu_1)$	0.56	0.56
the mixture of normals	NA	0.8	p=0.4	$(1-p)*(\mu_2-\mu_1)$	0.48	0.48
the mixture of normals	NA	0.8	p=0.5	$(1-p)*(\mu_2-\mu_1)$	0.4	0.4
the mixture of normals	NA	0.8	p=0.6	$(1-p)*(\mu_2-\mu_1)$	0.32	0.32
the mixture of normals	NA	0.8	p=0.7	$(1-p)*(\mu_2-\mu_1)$	0.24	0.24
the mixture of normals	NA	0.8	p=0.8	$(1-p)*(\mu_2-\mu_1)$	0.16	0.16
the mixture of normals	NA	0.8	p=0.9	$(1-p)*(\mu_2-\mu_1)$	0.08	0.08
the mixture of normals	NA	0.8	p=1	$(1-p)*(\mu_2-\mu_1)$	0	0
the mixture of normals	NA	0.2	all p values	$\mu_2-\mu_1$	0.2	0.2
the mixture of normals	NA	0.5	all p values	$\mu_2-\mu_1$	0.5	0.5

the mixture of normals	NA	0.8	all p values	$\mu_2 - \mu_1$	0.8	0.8
lognormal	0.3	0.2	$\sigma = 0.1$	$d * \sqrt{\{(\exp(\sigma^2) - 1)\exp(2\mu + \sigma^2)\}}$	0.02	0.2
lognormal	0.61	0.2	$\sigma = 0.2$	$d * \sqrt{\{(\exp(\sigma^2) - 1)\exp(2\mu + \sigma^2)\}}$	0.04	0.2
lognormal	0.95	0.2	$\sigma = 0.3$	$d * \sqrt{\{(\exp(\sigma^2) - 1)\exp(2\mu + \sigma^2)\}}$	0.06	0.2
lognormal	1.32	0.2	$\sigma = 0.4$	$d * \sqrt{\{(\exp(\sigma^2) - 1)\exp(2\mu + \sigma^2)\}}$	0.09	0.2
lognormal	1.75	0.2	$\sigma = 0.5$	$d * \sqrt{\{(\exp(\sigma^2) - 1)\exp(2\mu + \sigma^2)\}}$	0.12	0.2
lognormal	0.3	0.5	$\sigma = 0.1$	$d * \sqrt{\{(\exp(\sigma^2) - 1)\exp(2\mu + \sigma^2)\}}$	0.05	0.5
lognormal	0.61	0.5	$\sigma = 0.2$	$d * \sqrt{\{(\exp(\sigma^2) - 1)\exp(2\mu + \sigma^2)\}}$	0.10	0.5
lognormal	0.95	0.5	$\sigma = 0.3$	$d * \sqrt{\{(\exp(\sigma^2) - 1)\exp(2\mu + \sigma^2)\}}$	0.16	0.5
lognormal	1.32	0.5	$\sigma = 0.4$	$d * \sqrt{\{(\exp(\sigma^2) - 1)\exp(2\mu + \sigma^2)\}}$	0.23	0.5
lognormal	1.75	0.5	$\sigma = 0.5$	$d * \sqrt{\{(\exp(\sigma^2) - 1)\exp(2\mu + \sigma^2)\}}$	0.30	0.5
lognormal	0.3	0.8	$\sigma = 0.1$	$d * \sqrt{\{(\exp(\sigma^2) - 1)\exp(2\mu + \sigma^2)\}}$	0.08	0.8
lognormal	0.61	0.8	$\sigma = 0.2$	$d * \sqrt{\{(\exp(\sigma^2) - 1)\exp(2\mu + \sigma^2)\}}$	0.16	0.8
lognormal	0.95	0.8	$\sigma = 0.3$	$d * \sqrt{\{(\exp(\sigma^2) - 1)\exp(2\mu + \sigma^2)\}}$	0.26	0.8

lognormal	1.32	0.8	$\sigma = 0.4$	$d * \sqrt{\{(\exp(\sigma^2) - 1)\exp(2\mu + \sigma^2)\}}$	0.36	0.8
lognormal	1.75	0.8	$\sigma = 0.5$	$d * \sqrt{\{(\exp(\sigma^2) - 1)\exp(2\mu + \sigma^2)\}}$	0.48	0.8
lognormal	0.3	0.2	$\mu = 0.02,$ $\sigma = 0.1$	$exp\left(\frac{\sigma^2}{2}\right)(\exp(\mu) - 1)$	0.02	0.2
lognormal	0.61	0.2	$\mu = 0.04,$ $\sigma = 0.2$	$exp\left(\frac{\sigma^2}{2}\right)(\exp(\mu) - 1)$	0.04	0.2
lognormal	0.95	0.2	$\mu = 0.061,$ $\sigma = 0.3$	$exp\left(\frac{\sigma^2}{2}\right)(\exp(\mu) - 1)$	0.07	0.2
lognormal	1.32	0.2	$\mu = 0.083,$ $\sigma = 0.4$	$exp\left(\frac{\sigma^2}{2}\right)(\exp(\mu) - 1)$	0.09	0.2
lognormal	1.75	0.2	$\mu = 0.107,$ $\sigma = 0.5$	$exp\left(\frac{\sigma^2}{2}\right)(\exp(\mu) - 1)$	0.13	0.2
lognormal	0.3	0.5	$\mu = 0.05,$ $\sigma = 0.1$	$exp\left(\frac{\sigma^2}{2}\right)(\exp(\mu) - 1)$	0.05	0.5
lognormal	0.61	0.5	$\mu = 0.101,$ $\sigma = 0.2$	$exp\left(\frac{\sigma^2}{2}\right)(\exp(\mu) - 1)$	0.11	0.5
lognormal	0.95	0.5	$\mu = 0.154,$ $\sigma = 0.3$	$exp\left(\frac{\sigma^2}{2}\right)(\exp(\mu) - 1)$	0.17	0.5
lognormal	1.32	0.5	$\mu = 0.210,$ $\sigma = 0.4$	$exp\left(\frac{\sigma^2}{2}\right)(\exp(\mu) - 1)$	0.25	0.5
lognormal	1.75	0.5	$\mu = 0.271,$ $\sigma = 0.5$	$exp\left(\frac{\sigma^2}{2}\right)(\exp(\mu) - 1)$	0.35	0.5
lognormal	0.3	0.8	$\mu = 0.08,$ $\sigma = 0.1$	$exp\left(\frac{\sigma^2}{2}\right)(\exp(\mu) - 1)$	0.08	0.8
lognormal	0.61	0.8	$\mu = 0.163,$	$exp\left(\frac{\sigma^2}{2}\right)(\exp(\mu) - 1)$	0.18	0.8

			$\sigma = 0.2$			
lognormal	0.95	0.8	$\mu = 0.249,$ $\sigma = 0.3$	$exp\left(\frac{\sigma^2}{2}\right)(\exp(\mu) - 1)$	0.30	0.8
lognormal	1.32	0.8	$\mu = 0.341,$ $\sigma = 0.4$	$exp\left(\frac{\sigma^2}{2}\right)(\exp(\mu) - 1)$	0.44	0.8
lognormal	1.75	0.8	$\mu = 0.444,$ $\sigma = 0.5$	$exp\left(\frac{\sigma^2}{2}\right)(\exp(\mu) - 1)$	0.63	0.8
gamma	0.3	0.2	$\alpha = 43.93$	$d * \sqrt{(\alpha\beta^2)}$	1.33	0.2
gamma	0.61	0.2	$\alpha = 10.60$	$d * \sqrt{(\alpha\beta^2)}$	0.65	0.2
gamma	0.95	0.2	$\alpha = 4.44$	$d * \sqrt{(\alpha\beta^2)}$	0.42	0.2
gamma	1.32	0.2	$\alpha = 2.29$	$d * \sqrt{(\alpha\beta^2)}$	0.30	0.2
gamma	1.75	0.2	$\alpha = 1.31$	$d * \sqrt{(\alpha\beta^2)}$	0.23	0.2
gamma	0.3	0.5	$\alpha = 43.93$	$d * \sqrt{(\alpha\beta^2)}$	3.31	0.5
gamma	0.61	0.5	$\alpha = 10.60$	$d * \sqrt{(\alpha\beta^2)}$	1.63	0.5
gamma	0.95	0.5	$\alpha = 4.44$	$d * \sqrt{(\alpha\beta^2)}$	1.05	0.5
gamma	1.32	0.5	$\alpha = 2.29$	$d * \sqrt{(\alpha\beta^2)}$	0.76	0.5
gamma	1.75	0.5	$\alpha = 1.31$	$d * \sqrt{(\alpha\beta^2)}$	0.57	0.5
gamma	0.3	0.8	$\alpha = 43.93$	$d * \sqrt{(\alpha\beta^2)}$	5.30	0.8
gamma	0.61	0.8	$\alpha = 10.60$	$d * \sqrt{(\alpha\beta^2)}$	2.60	0.8
gamma	0.95	0.8	$\alpha = 4.44$	$d * \sqrt{(\alpha\beta^2)}$	1.69	0.8
gamma	1.32	0.8	$\alpha = 2.29$	$d * \sqrt{(\alpha\beta^2)}$	1.21	0.8
gamma	1.75	0.8	$\alpha = 1.31$	$d * \sqrt{(\alpha\beta^2)}$	0.92	0.8
gamma	0.3	0.2	$\alpha = 43.93,$ $\beta_c = 1,$ $\beta_t = 1.031$	$\alpha(\beta_t - \beta_c)$	1.35	0.2
gamma	0.61	0.2	$\alpha = 10.60,$ $\beta_c = 1,$	$\alpha(\beta_t - \beta_c)$	0.67	0.2

			$\beta_t = 1.063$			
gamma	0.95	0.2	$\alpha = 4.44$, $\beta_c = 1$, $\beta_t = 1.100$	$\alpha(\beta_t - \beta_c)$	0.44	0.2
gamma	1.32	0.2	$\alpha = 2.29$, $\beta_c = 1$, $\beta_t = 1.142$	$\alpha(\beta_t - \beta_c)$	0.32	0.2
gamma	1.75	0.2	$\alpha = 1.31$, $\beta_c = 1$, $\beta_t = 1.193$	$\alpha(\beta_t - \beta_c)$	0.25	0.2
gamma	0.3	0.5	$\alpha = 43.93$, $\beta_c = 1$, $\beta_t = 1.078$	$\alpha(\beta_t - \beta_c)$	3.45	0.5
gamma	0.61	0.5	$\alpha = 10.60$, $\beta_c = 1$, $\beta_t = 1.167$	$\alpha(\beta_t - \beta_c)$	1.77	0.5
gamma	0.95	0.5	$\alpha = 4.44$, $\beta_c = 1$, $\beta_t = 1.273$	$\alpha(\beta_t - \beta_c)$	1.21	0.5
gamma	1.32	0.5	$\alpha = 2.29$, $\beta_c = 1$, $\beta_t = 1.403$	$\alpha(\beta_t - \beta_c)$	0.92	0.5
gamma	1.75	0.5	$\alpha = 1.31$, $\beta_c = 1$, $\beta_t = 1.578$	$\alpha(\beta_t - \beta_c)$	0.76	0.5
gamma	0.3	0.8	$\alpha = 43.93$, $\beta_c = 1$, $\beta_t = 1.129$	$\alpha(\beta_t - \beta_c)$	5.65	0.8

gamma	0.61	0.8	$\alpha = 10.60,$ $\beta_c = 1,$ $\beta_t = 1.283$	$\alpha(\beta_t - \beta_c)$	3.00	0.8
gamma	0.95	0.8	$\alpha = 4.44,$ $\beta_c = 1,$ $\beta_t = 1.480$	$\alpha(\beta_t - \beta_c)$	2.13	0.8
gamma	1.32	0.8	$\alpha = 2.29,$ $\beta_c = 1,$ $\beta_t = 1.755$	$\alpha(\beta_t - \beta_c)$	1.73	0.8
gamma	1.75	0.8	$\alpha = 1.31,$ $\beta_c = 1,$ $\beta_t = 2.193$	$\alpha(\beta_t - \beta_c)$	1.56	0.8
chi-square	0.3	0.2	$d.f. = 88$	$d * \sqrt{2 * d.f.}$	2.65	0.2
chi-square	0.61	0.2	$d.f. = 21$	$d * \sqrt{2 * d.f.}$	1.30	0.2
chi-square	0.95	0.2	$d.f. = 9$	$d * \sqrt{2 * d.f.}$	0.85	0.2
chi-square	1.32	0.2	$d.f. = 5$	$d * \sqrt{2 * d.f.}$	0.63	0.2
chi-square	1.75	0.2	$d.f. = 3$	$d * \sqrt{2 * d.f.}$	0.49	0.2
chi-square	0.3	0.5	$d.f. = 88$	$d * \sqrt{2 * d.f.}$	6.63	0.5
chi-square	0.61	0.5	$d.f. = 21$	$d * \sqrt{2 * d.f.}$	3.24	0.5
chi-square	0.95	0.5	$d.f. = 9$	$d * \sqrt{2 * d.f.}$	2.12	0.5
chi-square	1.32	0.5	$d.f. = 5$	$d * \sqrt{2 * d.f.}$	1.58	0.5
chi-square	1.75	0.5	$d.f. = 3$	$d * \sqrt{2 * d.f.}$	1.22	0.5
chi-square	0.3	0.8	$d.f. = 88$	$d * \sqrt{2 * d.f.}$	10.61	0.8
chi-square	0.61	0.8	$d.f. = 21$	$d * \sqrt{2 * d.f.}$	5.18	0.8
chi-square	0.95	0.8	$d.f. = 9$	$d * \sqrt{2 * d.f.}$	3.39	0.8
chi-square	1.32	0.8	$d.f. = 5$	$d * \sqrt{2 * d.f.}$	2.53	0.8
chi-square	1.75	0.8	$d.f. = 3$	$d * \sqrt{2 * d.f.}$	1.96	0.8

Table A.1 a summary of all the algebra for different cases within the same distribution of the outcome variables in terms of the skewness, parameters, and corresponding true effect for mean difference and the standardized mean difference

The Number of Studies	The Number of Individuals per Group	Mean Difference	Standardized Mean Difference
2	5	0.1031	0.0264
2	10	0.0764	0.0403
2	20	0.0591	0.0468
2	50	0.0554	0.0511
5	5	0.1285	0.0192
5	10	0.0809	0.037
5	20	0.0619	0.0465
5	50	0.0529	0.0455
10	5	0.1398	0.023
10	10	0.081	0.0346
10	20	0.0613	0.046
10	50	0.0542	0.0488
20	5	0.1475	0.0212
20	10	0.0854	0.0385
20	20	0.0614	0.0427
20	50	0.0578	0.0489
2	5	0.1084	0.0275
2	10	0.0759	0.0393
2	20	0.0573	0.0486
2	50	0.0568	0.0468
5	5	0.1177	0.0226
5	10	0.0738	0.0345
5	20	0.0594	0.0455
5	50	0.0576	0.0467
10	5	0.1452	0.0225
10	10	0.0838	0.0397
10	20	0.0619	0.047

10	50	0.0539	0.0486
20	5	0.1473	0.0221
20	10	0.0777	0.0409
20	20	0.0638	0.0464
20	50	0.0523	0.0479
2	5	0.1061	0.0301
2	10	0.0736	0.0383
2	20	0.0552	0.0425
2	50	0.0517	0.0496
5	5	0.1247	0.0217
5	10	0.078	0.0393
5	20	0.0628	0.0414
5	50	0.0578	0.0492
10	5	0.139	0.021
10	10	0.0807	0.0355
10	20	0.0661	0.044
10	50	0.0568	0.0457
20	5	0.1474	0.0246
20	10	0.0817	0.0352
20	20	0.0631	0.044
20	50	0.0557	0.0449

Table A.2 the type I error rate for the mixture of normal distributions of the outcome variable with the different number of studies, and the number of individuals per group when using the mean difference and the standardized mean difference as the effect sizes

Skewness	The Number of Studies	The Number of Individuals per Group	lognormal	Gamma	χ^2
0.3	2	5	0.1019	0.949	0.9954
0.61	2	5	0.1042	0.6547	0.945
0.95	2	5	0.1122	0.3982	0.8048
1.32	2	5	0.1111	0.2617	0.613
1.75	2	5	0.1186	0.1786	0.4414
0.3	2	10	0.0774	0.91	0.9982
0.61	2	10	0.0783	0.4634	0.8969
0.95	2	10	0.0796	0.2511	0.6449
1.32	2	10	0.078	0.1672	0.4282
1.75	2	10	0.0835	0.1204	0.2982
0.3	2	20	0.0585	0.7421	0.9932
0.61	2	20	0.0567	0.2764	0.715
0.95	2	20	0.0575	0.1567	0.4239
1.32	2	20	0.0613	0.1057	0.266
1.75	2	20	0.0696	0.0875	0.1906
0.3	2	50	0.0593	0.4125	0.8976
0.61	2	50	0.0527	0.1513	0.389
0.95	2	50	0.0558	0.0977	0.2065
1.32	2	50	0.0535	0.0699	0.1436
1.75	2	50	0.0536	0.0647	0.1101
0.3	5	5	0.1263	0.125	0.1231
0.61	5	5	0.1297	0.1289	0.1299
0.95	5	5	0.1275	0.1232	0.1219
1.32	5	5	0.1245	0.1164	0.1224
1.75	5	5	0.1155	0.1213	0.1125
0.3	5	10	0.0753	0.0785	0.0797

0.61	5	10	0.0785	0.0727	0.0773
0.95	5	10	0.0785	0.0739	0.0762
1.32	5	10	0.0757	0.0749	0.0781
1.75	5	10	0.0725	0.0721	0.076
0.3	5	20	0.0579	0.06	0.062
0.61	5	20	0.0576	0.0582	0.0659
0.95	5	20	0.0598	0.0587	0.0611
1.32	5	20	0.055	0.0625	0.0637
1.75	5	20	0.0582	0.0642	0.0602
0.3	5	50	0.0544	0.0546	0.0543
0.61	5	50	0.0538	0.0527	0.0544
0.95	5	50	0.0545	0.0565	0.0522
1.32	5	50	0.0515	0.0531	0.05
1.75	5	50	0.053	0.0517	0.0543
0.3	10	5	0.1372	0.1375	0.1401
0.61	10	5	0.1373	0.1381	0.1396
0.95	10	5	0.1386	0.1366	0.1356
1.32	10	5	0.1299	0.1324	0.1354
1.75	10	5	0.1294	0.1262	0.1327
0.3	10	10	0.079	0.0856	0.0824
0.61	10	10	0.0806	0.0797	0.0802
0.95	10	10	0.0825	0.082	0.0758
1.32	10	10	0.077	0.0739	0.0777
1.75	10	10	0.0812	0.0762	0.0767
0.3	10	20	0.0617	0.0605	0.0607
0.61	10	20	0.0604	0.0619	0.0633
0.95	10	20	0.0635	0.0654	0.064
1.32	10	20	0.0627	0.0607	0.0633
1.75	10	20	0.0628	0.0593	0.0599

0.3	10	50	0.0536	0.057	0.0541
0.61	10	50	0.0573	0.0569	0.0565
0.95	10	50	0.0564	0.0546	0.0537
1.32	10	50	0.0529	0.0575	0.0493
1.75	10	50	0.0547	0.0566	0.0543
0.3	20	5	0.1474	0.1467	0.1449
0.61	20	5	0.149	0.1454	0.1399
0.95	20	5	0.1507	0.1459	0.148
1.32	20	5	0.143	0.1409	0.1493
1.75	20	5	0.1458	0.1345	0.1439
0.3	20	10	0.0776	0.0793	0.0843
0.61	20	10	0.0794	0.0809	0.0813
0.95	20	10	0.0797	0.0823	0.0851
1.32	20	10	0.0806	0.0774	0.0826
1.75	20	10	0.0764	0.0802	0.0794
0.3	20	20	0.0628	0.0632	0.0649
0.61	20	20	0.066	0.0613	0.0622
0.95	20	20	0.0614	0.0606	0.0648
1.32	20	20	0.0631	0.0595	0.0608
1.75	20	20	0.0625	0.0615	0.0625
0.3	20	50	0.0577	0.0547	0.0602
0.61	20	50	0.0578	0.0514	0.0543
0.95	20	50	0.054	0.0547	0.0567
1.32	20	50	0.0543	0.0579	0.0538
1.75	20	50	0.0555	0.0551	0.0581

Table A.3 the type I error rate for the lognormal, gamma, and χ^2 distributions of the outcome variable with different skewness, the number of studies, and the number of individuals per group when using the mean difference as the effect sizes

Skewness	The Number of Studies	The Number of Individuals per Group	lognormal	Gamma	χ^2
0.3	2	5	0.1183	0.1186	0.1127
0.61	2	5	0.1189	0.111	0.117
0.95	2	5	0.1135	0.1173	0.1147
1.32	2	5	0.117	0.1117	0.1207
1.75	2	5	0.1142	0.1143	0.119
0.3	2	10	0.081	0.0831	0.0896
0.61	2	10	0.0808	0.084	0.0854
0.95	2	10	0.084	0.0854	0.0859
1.32	2	10	0.0841	0.0865	0.0898
1.75	2	10	0.082	0.0862	0.088
0.3	2	20	0.0663	0.0672	0.0684
0.61	2	20	0.063	0.0669	0.062
0.95	2	20	0.0691	0.0654	0.0638
1.32	2	20	0.0656	0.0689	0.0677
1.75	2	20	0.0721	0.0625	0.0663
0.3	2	50	0.0611	0.0573	0.0543
0.61	2	50	0.0566	0.0583	0.0563
0.95	2	50	0.0558	0.0532	0.0633
1.32	2	50	0.0543	0.0553	0.0537
1.75	2	50	0.0564	0.0574	0.0593
0.3	5	5	0.0237	0.0198	0.0228
0.61	5	5	0.0243	0.0221	0.0217
0.95	5	5	0.0236	0.0238	0.0223
1.32	5	5	0.0219	0.0218	0.0232
1.75	5	5	0.022	0.0224	0.0212
0.3	5	10	0.04	0.038	0.0368

0.61	5	10	0.0356	0.0357	0.0388
0.95	5	10	0.0371	0.036	0.039
1.32	5	10	0.0417	0.0365	0.0358
1.75	5	10	0.0364	0.0365	0.0408
0.3	5	20	0.043	0.044	0.0409
0.61	5	20	0.0426	0.0453	0.0402
0.95	5	20	0.0461	0.0424	0.044
1.32	5	20	0.0428	0.0438	0.0459
1.75	5	20	0.043	0.0464	0.0454
0.3	5	50	0.0467	0.0456	0.0536
0.61	5	50	0.0471	0.0492	0.0481
0.95	5	50	0.0455	0.044	0.0486
1.32	5	50	0.049	0.0487	0.0425
1.75	5	50	0.0457	0.0488	0.0501
0.3	10	5	0.0201	0.022	0.0216
0.61	10	5	0.0218	0.0217	0.0203
0.95	10	5	0.0216	0.0235	0.0205
1.32	10	5	0.0239	0.0224	0.0211
1.75	10	5	0.022	0.023	0.022
0.3	10	10	0.0353	0.0403	0.0354
0.61	10	10	0.0366	0.0358	0.0373
0.95	10	10	0.0368	0.0362	0.0358
1.32	10	10	0.0369	0.0349	0.0351
1.75	10	10	0.0368	0.0372	0.0369
0.3	10	20	0.0437	0.0406	0.0452
0.61	10	20	0.0433	0.0469	0.0474
0.95	10	20	0.0459	0.0428	0.0467
1.32	10	20	0.0434	0.0445	0.0421
1.75	10	20	0.0451	0.0423	0.0453

0.3	10	50	0.0475	0.045	0.0413
0.61	10	50	0.0477	0.048	0.0498
0.95	10	50	0.0503	0.0469	0.0478
1.32	10	50	0.0489	0.0436	0.0494
1.75	10	50	0.0436	0.0481	0.0479
0.3	20	5	0.0193	0.0216	0.0215
0.61	20	5	0.0219	0.019	0.0239
0.95	20	5	0.019	0.0232	0.0225
1.32	20	5	0.0219	0.0197	0.0216
1.75	20	5	0.0239	0.0223	0.0226
0.3	20	10	0.0342	0.0357	0.0407
0.61	20	10	0.0372	0.0362	0.0387
0.95	20	10	0.0396	0.036	0.0379
1.32	20	10	0.0396	0.0405	0.0424
1.75	20	10	0.0381	0.0371	0.0341
0.3	20	20	0.0436	0.046	0.0407
0.61	20	20	0.041	0.0405	0.0413
0.95	20	20	0.0448	0.0432	0.0391
1.32	20	20	0.0451	0.0418	0.0439
1.75	20	20	0.0468	0.0446	0.0445
0.3	20	50	0.0488	0.0457	0.0444
0.61	20	50	0.0427	0.0426	0.0464
0.95	20	50	0.0451	0.0496	0.0508
1.32	20	50	0.0464	0.0454	0.0488
1.75	20	50	0.0479	0.0472	0.0463

Table A.4 the type I error rate for the lognormal, gamma, and χ^2 distributions of the outcome variable with different skewness, the number of studies, and the number of individuals per group when using the standardized mean difference as the effect sizes