Stochastic Models for Dynamic Resource Allocation

Amin Yarahmadi

Management Science Department

Submitted for the degree of Doctor of Philosophy at Lancaster University.

Advisors:

Dr Peter Jacko

Prof Kevin Glazebrook

September 2022

Abstract

Determining the efficacy of a novel intervention is vital before making it available to the public. The standard equal fixed randomisation procedure in the design of (static) experiments leads to an unbiased Maximum Likelihood Estimator (MLE) for each intervention. However, this approach results in a heavily suboptimal cumulative reward. On the other hand, it imposes limitations in some situations, especially for rare diseases, when it is desirable to design a clinical trial on a small number of subjects while treating them as well as possible. This motivates the use of response-adaptive procedures where the allocation ratios to each arm can be skewed toward the better-performing intervention as subject responses become available. Hence, we consider the Bayesian Beta-Bernoulli finite-horizon two-armed bandit problem with binary responses and the objective function of maximising the Bayes-expected total number of subject successes in the trial, which we call the subject benefit.

Using a memory-efficient implementation, dynamic programming is utilised as the solution method for the proposed model to derive the randomised designs. Despite the type of randomisation procedure, the MLE is estimated in a frequentist way using DP-based solutions at the end of the trial.

We first evaluate the bias of MLE and show that it is unacceptably high and variable due to the model's adaptiveness. We propose a new augmented estimator with the aim of mitigating the estimation bias whilst the DP actions are deterministic. Moreover, by modifying the allocation decision at every time step, we introduce two novel allocation procedures that mitigate the bias induced by the DP procedure: (i) DP using an augmented estimator, which adds a number of pseudo-successes to the worse-performing intervention, and (ii) randomised DP procedure, which perturbs the Bayes-optimal allocation decision with a given probability.

Lastly, another DP design is proposed based upon setting an interim analysis, in which some novel and non-trivial stopping criteria have been developed, in the middle of the trial. The interim analysis look can be implemented in the simulation step or both the DP procedure and the simulation step, identically.

We evaluated the proposed designs via extensive simulation studies in a broad range of scenarios.

This thesis addresses some key issues in the trade-off between reducing the bias in the estimation and improving the subject benefit in the bandit models, which can be considered as a limitation preventing bandit models from being implemented in practice.

Acknowledgements

I would like to express my thanks to my supervisors, Dr Peter Jacko and Prof Kevin Glazebrook, for all their support, guidance, and patience throughout this journey. What I have learnt from them is not limited to the bandit models but it goes beyond a PhD study and is about being a kind, supportive and responsible human being. My sincere gratitude also goes to Prof Nigel Stallard from Warwick Medical School for his sympathetic help and patience with me in the last six months of this PhD.

I am greatly indebted to Lancaster University, particularly the Management Science department, for providing me with such a fantastic opportunity to nurture my growth during my doctorate studies. I am also grateful to Mathematics and Statistic Hub (MASH) and Dr Anna Karapiperi for enthusiastically embracing my experience and collaboration in teaching and leading several maths and stats workshops in different departments.

I eternally grateful to a few small things that made a huge difference in this journey for me: gym, whisky, water, music, in particular, Mr (ostad) Mohammad-Reza Shajarian.

Finally, I do not know how to thank my small family for their constant whole-hearted compassion and mentally support from a distance.

Declaration

I declare that the work presented in this thesis is based on my original research, which has been done by myself under the supervision of Dr Peter Jacko and Prof Kevin Glazebrook. It also has not been previously submitted elsewhere for the award of any other degree.

Chapter 4, which will be submitted as “On the Estimation Bias of the Bayesian Decision-Theoretic Response-Adaptive Randomization Procedure." was virtually presented in *2021 INFORMS Annual Meeting* conference.

Chapter 5, which will be submitted as “Addressing the Trade-off between Optimal Cumulative Reward and Unbiased Estimation in Sequential Experiments." was presented in *EURO 2022* conference.

Amin Yarahmadi

List of Recurring Abbreviations

DP	Dynamic Programming
EFR	Equal Fixed Randomisation
HT	Horvitz-Thompson
IPW	Inverse probability weighted
MAB(P)	Multi-Armed Bandit (Problem)
MLE	Maximum Likelihood Estimator
OIDP	Optimistic on Inferior Dynamic Programming
RAR	Response-Adaptive Randomisation
RAnR	Response-Adaptive non Randomisation
RCT	Randomised Controlled Trial
RDP	Randomised Dynamic Programming
RMSE	Root Mean Squared Error

Contents

List of Figures **x**

1 Introduction and Motivation **1**

1.1 Outline of Thesis 4

2 Methodology and Background **6**

2.1 Randomisation 6

2.2 Response-Adaptive Randomisation (RAR) 8

2.3 Bandit Models 11

2.3.1 The Multi-Armed Bandit Problem (MABP) 11

2.3.2 Markov Decision Processes (MDPs) 12

2.3.3 Dynamic Programming (DP) Approach 14

2.4 Estimation 19

2.4.1 Estimation Bias 21

2.5 Simulation Set-up 22

2.6 Memory in Use 25

3 Model and After-trial Studies **26**

3.1 The two-armed Bayesian Beta-Bernoulli model 26

3.2 Maximum Likelihood Estimator (MLE) and its Bias 29
3.3 After-Trial Studies 32
3.3.1 Average Estimation Bias 33
3.3.2 Subject Benefit 36

4 Estimation Bias Mitigation – post trial 44

4.1 Introduction 44
4.1.1 Estimation Bias under RAR Procedures 47
4.1.2 Our Contributions 49
4.2 Estimation Bias: Theoretical Characterization 51
4.2.1 Maximum Likelihood Estimator and its Bias 51
4.2.2 Bayesian Estimators and their Bias 54
4.2.3 HT and IPW Estimators 57
4.3 Augmented Estimator and its Bias 59
4.4 Estimation Bias: Numerical Characterization for Decision-Theoretic Designs 62
4.4.1 Modification criterion and estimation functionality 64
4.4.2 Augmented Estimator Performance 66
4.4.3 Augmented HT in a non-Randomised RAR procedure 68
4.5 Discussion 75
4.6 Appendix 77
4.6.1 DP satisfies *Exploit* 77
4.6.2 Proofs of the main results 86
4.6.3 Both-arms modification and overshooting 91
4.6.4 Results with Jeffreys prior 93

5 Estimation Bias Mitigation – during trial 98

5.1 Introduction 98

5.1.1 Our contribution 102

5.2 Response-Adaptive non-Randomised (RAnR) procedure 104

5.2.1 Optimistic on Inferior Dynamic Programming (OIDP) . . . 105

5.2.2 Subject Benefit 106

5.2.3 Average Estimation Bias 108

5.3 Randomised Dynamic Programming (RDP) 111

5.3.1 Bi-level Randomisation- Alternative Interpretation of RDP . 113

5.3.2 Subject Benefit 115

5.3.3 Average Estimation Bias 116

5.4 An Unbiased Estimator 118

5.5 Root Mean Square Error (RMSE) and Data Visualisation 120

5.5.1 Root Mean Square Error (RMSE) 120

5.5.2 Data visualisation- Box plot 121

5.6 Discussion 123

6 Extension to Trials with Early Stopping 137

6.1 Introduction 137

6.2 Model 139

6.3 Simulation Set-up 142

6.4 Interim Analysis Characteristics 142

6.4.1 z-test 142

6.4.2 Stopping Criteria: Primary Set-up 144

6.4.3 Stopping Criteria: Developmental Process 144

6.4.4 Stopping Criteria: Final Necessary Conditions 147

6.5 Average Estimation Bias 148

6.6 Percentage of False Early Stopping 150

6.7 Subject benefit 160

6.8 Discussion 160

7 Conclusion and Further Work 164

7.1 Summary and Contributions 164

7.1.1 Chapter 3, Model and After-trial Studies 165

7.1.2 Chapter 4, Augmented Estimator 166

7.1.3 Chapter 5, OIDP and RDP 166

7.1.4 Chapter 6, DPWOI, DPWI 167

7.2 Future Work 168

References 171

List of Figures

3.1 After-trial population: 1000. The precision of value functions in the backward induction algorithm is set to be up to about 7 digits in the left-hand-side column whilst it goes up to about 16 digits in the right-hand-side counterparts. (a) $T = 60$ (b) $T = 120$ (c) $T = 180$ (d) $T = 240$. x-axis: Bias of MLE, y-axis: Covariance (MLE, Sample Size). 38

3.2 After-trial population: 1 million. The precision of value functions in the backward induction algorithm is set to be up to about 7 digits in the left-hand-side column whilst it goes up to about 16 digits in the right-hand-side counterparts. (a) $T = 60$ (b) $T = 120$ (c) $T = 180$ (d) $T = 240$. x-axis: Bias of MLE, y-axis: Covariance (MLE, Sample Size). 39

3.3 After-trial population: 1 billion. The precision of value functions in the backward induction algorithm is set to be up to about 7 digits in the left-hand-side column whilst it goes up to about 16 digits in the right-hand-side counterparts. (a) $T = 60$ (b) $T = 120$ (c) $T = 180$ (d) $T = 240$. x-axis: Bias of MLE, y-axis: Covariance (MLE, Sample Size). 40

4.1 $\Big(\dot{s}_k(T) = 0, \dot{f}_k(T) = 0, \widetilde{s}_k(0) = 0, \widetilde{f}_k(0) = 0\Big)$ vs. $\Big(\dot{s}_k(T) = 0, \dot{f}_k(T) = 0, \widetilde{s}_k(0) = 1, \widetilde{f}_k(0) = 1\Big)$ i.e., $\Big($Frequentist MLE eq. (4.1)$\Big)$ vs. $\Big($Bayesian estimator eq. (4.3)$\Big)$: (a) T=60 (b) T=120 (c) T=180 (d) T=240. x-axis: Bias of estimator, y-axis: Covariance (Estimator, Sample Size). 69

4.2 $\Big(\dot{s}_k(T) = \widehat{\theta}_k(T), \dot{f}_k(T) = 0, \widetilde{s}_k(0) = 0, \widetilde{f}_k(0) = 0\Big)$ vs. $\Big(\dot{s}_k(T) = \widehat{\theta}_k(T), \dot{f}_k(T) = 0, \widetilde{s}_k(0) = 1, \widetilde{f}_k(0) = 1\Big)$: (a) T=60 (b) T=120 (c) T=180 (d) T=240. x-axis: Bias of estimator, y-axis: Covariance (Estimator, Sample Size). 70

4.3 $\Big(\dot{s}_k(T) = \ln(T+1)\widehat{\theta}_k(T), \dot{f}_k(T) = 0, \widetilde{s}_k(0) = 0, \widetilde{f}_k(0) = 0\Big)$ vs. $\Big(\dot{s}_k(T) = 4\ln(T+1)(\widehat{\theta}_k(T))^2, \dot{f}_k(T) = 0, \widetilde{s}_k(0) = 0, \widetilde{f}_k(0) = 0\Big)$: (a) T=60 (b) T=120 (c) T=180 (d) T=240. x-axis: Bias of estimator, y-axis: Covariance (Estimator, Sample Size). 71

4.4 $\Big(\dot{s}_k(T) = 9\ln(T+1)(\widehat{\theta}_k(T))^3, \dot{f}_k(T) = 0, \widetilde{s}_k(0) = 0, \widetilde{f}_k(0) = 0\Big)$ vs. $\Big(\dot{s}_k(T) = 27\sqrt{\ln(T+1)}(\widehat{\theta}_k(T))^3, \dot{f}_k(T) = 0, \widetilde{s}_k(0) = 0, \widetilde{f}_k(0) = 0\Big)$: (a) T=60 (b) T=120 (c) T=180 (d) T=240. x-axis: Bias of estimator, y-axis: Covariance (Estimator, Sample Size). 72

4.5 HT (IPW) estimator vs augmented equivalent: (a) T=60 (b) T=120 (c) T=180 (d) T=240. x-axis: Bias of estimator, y-axis: Covariance (Estimator, Sample Size). 74

4.6 Left-hand side: Applying augmentations on both arms $\left(\dot{s}_k(T) = 27\sqrt{\ln(T+1)}(\widehat{\theta}_k(T))^3, \dot{f}_k(T) = 0, \widetilde{s}_k(0) = 0, \widetilde{f}_k(0) = 0\right)$, and right-hand side: Overshooting by too large constant $\left(\dot{s}_k(T) = 243\sqrt{\ln(T+1)}(\widehat{\theta}_k(T))^3, \dot{f}_k(T) = 0, \widetilde{s}_k(0) = 0, \widetilde{f}_k(0) = 0\right)$: (a) T=60 (b) T=120 (c) T=180 (d) T=240. x-axis: Bias of estimator, y-axis: Covariance (Estimator, Sample Size). 92

4.7 $\left(\dot{s}_k(T) = 0, \dot{f}_k(T) = 0, \widetilde{s}_k(0) = 0, \widetilde{f}_k(0) = 0\right)$ vs. $\left(\dot{s}_k(T) = 0, \dot{f}_k(T) = 0, \widetilde{s}_k(0) = 1, \widetilde{f}_k(0) = 1\right)$ i.e., (Frequentist MLE eq. (4.1)) vs. (Bayesian estimator eq. (4.3)): (a) T=60 (b) T=120 (c) T=180 (d) T=240. x-axis: Bias of estimator, y-axis: Covariance (Estimator, Sample Size). 94

4.8 $\left(\dot{s}_k(T) = \widehat{\theta}_k(T), \dot{f}_k(T) = 0, \widetilde{s}_k(0) = 0, \widetilde{f}_k(0) = 0\right)$ vs. $\left(\dot{s}_k(T) = \widehat{\theta}_k(T), \dot{f}_k(T) = 0, \widetilde{s}_k(0) = 1/2, \widetilde{f}_k(0) = 1/2\right)$: (a) T=60 (b) T=120 (c) T=180 (d) T=240. x-axis: Bias of estimator, y-axis: Covariance (Estimator, Sample Size). 95

4.9 $\left(\dot{s}_k(T) = \ln(T+1)\widehat{\theta}_k(T), \dot{f}_k(T) = 0, \widetilde{s}_k(0) = 0, \widetilde{f}_k(0) = 0\right)$ vs. $\left(\dot{s}_k(T) = 4\ln(T+1)(\widehat{\theta}_k(T))^2, \dot{f}_k(T) = 0, \widetilde{s}_k(0) = 0, \widetilde{f}_k(0) = 0\right)$: (a) T=60 (b) T=120 (c) T=180 (d) T=240. x-axis: Bias of estimator, y-axis: Covariance (Estimator, Sample Size). 96

4.10 $\Big(\dot{s}_k(T) = 9\ln(T+1)(\widehat{\theta}_k(T))^3, \dot{f}_k(T) = 0, \widetilde{s}_k(0) = 0, \widetilde{f}_k(0) = 0\Big)$ vs. $\Big(\dot{s}_k(T) = 27\sqrt{\ln(T+1)}(\widehat{\theta}_k(T))^3, \dot{f}_k(T) = 0, \widetilde{s}_k(0) = 0, \widetilde{f}_k(0) = 0\Big)$: (a) T=60 (b) T=120 (c) T=180 (d) T=240. x-axis: Bias of estimator, y-axis: Covariance (Estimator, Sample Size). 97

5.1 Alternative interpretation of the RDP randomisation procedure . . 114

5.2 The estimation bias and covariance reduction comparison in the design with classical DP (left-hand side column) vs OIDP in which ($\dot{s} = 1$) (right-hand side column) for varying trial sizes: (a) $T = 60$ (b) $T = 120$ (c) $T = 180$ (d) $T = 240$. x-axis: Bias of MLE, y-axis: Covariance (MLE, Sample Size). 127

5.3 The estimation bias and covariance reduction comparison in the design with OIDP in which ($\dot{s} = 2$) (left-hand side column) vs ($\dot{s} = \ln(t)$) (right-hand side column) for varying trial sizes: (a) $T = 60$ (b) $T = 120$ (c) $T = 180$ (d) $T = 240$. x-axis: Bias of MLE, y-axis: Covariance (MLE, Sample Size). 128

5.4 RDP procedure (bi-level randomisation): adaptive branch vs. pooled data: (a) $T = 60$ (b) $T = 120$ (c) $T = 180$ (d) $T = 240$. x-axis: Bias of MLE, y-axis: Covariance (MLE, Sample Size). 129

5.5 RDP procedure (bi-level randomisation) with different degrees of randomisation for $T = 60$: adaptive branch vs. pooled data: (a) $P = 0.5$ (b) $P = 0.6$ (c) $P = 0.7$ (d) $P = 0.8$. x-axis: Bias of MLE, y-axis: Covariance (MLE, Sample Size). 130

5.6 IPW vs. nIPW estimated in the designs using RDP procedure where $p = 0.9$: (a) $T = 60$ (b) $T = 120$ (c) $T = 180$ (d) $T = 240$. x-axis: Bias of the estimator, y-axis: Covariance (Estimator, Sample Size). 131

5.7 RMSE comparison for $T = 60$: OIDP designs with modifications vs. RDP designs with different degrees of randomisations. (a) $\dot{s} = 0$ vs $P = 0.6$ (b) $\dot{s} = 1$ vs $P = 0.7$ (c) $\dot{s} = 2$ vs $P = 0.8$ (d) $\dot{s} = ln(\tau + 1)$ vs $P = 0.9$. x-axis: Scenario, y-axis: RSME. 132

5.8 Box Plots comparison for $T = 240$: (a) Classical DP (b) OIDP with $\dot{s} = 2$ (c) RDP with $p = 0.9$ (d) EFR. x-axis: Scenario. 133

6.1 The estimation bias and covariance reduction comparison in the model case (a) *DPWOI*, and (b) *DPWI*. Note that $T = 120$ and $\alpha = 0.01$, and completed trials (left-hand side) are compared with those stopped at the interim (right-hand side). x-axis: Bias of the estimator, y-axis: Covariance (Estimator, Sample Size). 147

6.2 The estimation bias and covariance reduction comparison in the model case *DPWOI*, trial category (i), $T = 120$ vs $T = 240$: (a) $\alpha = 0.0025$ (b) $\alpha = 0.005$ (c) $\alpha = 0.01$ (d) $\alpha = 0.02$. x-axis: Bias of the estimator, y-axis: Covariance (Estimator, Sample Size). . . . 151

6.3 The estimation bias and covariance reduction comparison in the model case *DPWOI*, trial category (ii), $T = 120$ vs $T = 240$: (a) $\alpha = 0.0025$ (b) $\alpha = 0.005$ (c) $\alpha = 0.01$ (d) $\alpha = 0.02$. x-axis: Bias of the estimator, y-axis: Covariance (Estimator, Sample Size). . . . 152

6.4 The estimation bias and covariance reduction comparison in the model case *DPWOI*, trial category (iii), $T = 120$ vs $T = 240$: (a) $\alpha = 0.0025$ (b) $\alpha = 0.005$ (c) $\alpha = 0.01$ (d) $\alpha = 0.02$. x-axis: Bias of the estimator, y-axis: Covariance (Estimator, Sample Size). . . . 153

6.5 The estimation bias and covariance reduction comparison in the model case *DPWI*, trial category (i), $T = 120$ vs $T = 240$: (a) $\alpha = 0.0025$ (b) $\alpha = 0.005$ (c) $\alpha = 0.01$ (d) $\alpha = 0.02$. x-axis: Bias of the estimator, y-axis: Covariance (Estimator, Sample Size). . . . 154

6.6 The estimation bias and covariance reduction comparison in the model case *DPWI*, trial category (ii), $T = 120$ vs $T = 240$: (a) $\alpha = 0.0025$ (b) $\alpha = 0.005$ (c) $\alpha = 0.01$ (d) $\alpha = 0.02$. x-axis: Bias of the estimator, y-axis: Covariance (Estimator, Sample Size). . . . 155

6.7 The estimation bias and covariance reduction comparison in the model case *DPWI*, trial category (iii), $T = 120$ vs $T = 240$: (a) $\alpha = 0.0025$ (b) $\alpha = 0.005$ (c) $\alpha = 0.01$ (d) $\alpha = 0.02$. x-axis: Bias of the estimator, y-axis: Covariance (Estimator, Sample Size). . . . 156

6.8 The percentages of false early stopping for various significance levels: left-hand side $T = 120$ vs right-hand side $T = 240$. Part (a): *DPWOI*, and part (b): *DPWI*. x-axis: Scenario, y-axis: Percentage. 158

Chapter 1

Introduction and Motivation

Before a novel intervention is widely publicised, clinical trials aiming to determine the corresponding efficacy are among the potent tools that need to be undertaken (Pocock, 2013). Such trials, particularly in medical science, are typically composed of four phases (Jennison and Turnbull, 1999). Phase I trials dealing with toxicity and pharmacology are known as the exploratory stage, in which the participating population is small in size and healthy in condition. The primary objective of phase I is to establish a suitable and tolerable dose level for a new treatment. Phase II pilot studies are of moderate size with a few hundred diseased patients. Evaluating the initial efficacy and safety is of interest in this phase of the trials. For instance, the frequency of a successful dose intake for properly treating patients is investigated in the second phase (Peace and Chen, 2010). Promising treatments proceed to phase III, where more than a thousand patient volunteers are involved. Conventionally, the definitive evaluation of the new proposed treatment is conducted by comparing it with a control (standard therapy or placebo) in terms of effectiveness and absence of any long-term adverse side effects. This comparative phase III is pivotal as the statistical designs and analyses receive the most attention and scrutiny. Finally, phase IV, known as post-marketing surveillance, deals with additional testing and monitoring for long-term effects in the wider population.

The drug development process is expensive as it may cost one to two billion dollars and is time-consuming for patients as it could take 10 to 15 years to receive the final efficacy and safety approval. Hence, we need to pay more attention to patient-centric trials in which patients are treated as effectively as possible within the trial. In 2022, the United States Food and Drug Administration (FDA) through the Center for Drug Evaluation and Research (CDER) introduces a new Accelerating Rare disease Cures (ARC) Program, known generally as CDER's ARC Program U.S. Food and Drug Administration (2022), with the vision of enhancing the development of effective and safe treatment options for rare diseases. CDER's ARC Program aims to accelerate the availability of rare disease treatments by driving scientific and regulatory innovation and engagement.

On the other hand, drug development for the approximately 7000 rare diseases affecting over 400 million people worldwide is potentially complex owing to the limited understanding of the natural history of the diseases and patient populations. Hence, the primary objective of CDER is to overcome obstacles associated with rare disease trials and facilitate the development of therapies as much as possible in the U.S. However, there is still a great deal of room for improvement in the unmet need for FDA-approved therapies for rare diseases. Accordingly, this is where CDER's ARC Program mission can be accomplished by utilising scientific and regulatory innovation and engagement.

Focusing on phase III of clinical trials, where the effectiveness of a new treatment is compared with the standard therapy, we deal with those trials designed for rare diseases. Therefore, they are associated with all the above-mentioned complexities and hurdles.

Under the assumption of equal variances and normally distributed data, the current gold standard design used in clinical trials is the *randomised controlled trial* (RCT), in which the proportion of patients allocated to each participating therapy is pre-fixed and typically equal. Although RCT amplifies the chance of detecting any clinically and statistically meaningful differences, i.e. it maximises the statistical power, it suffers from a lack of flexibility in patient well-being in terms of randomising half of the participants to the inferior or control therapy. Whilst in the context of *rare* diseases a substantial proportion of all diseased patients may be involved in the trial, the primary ethical goal should be treating those patients *within* the trial as effectively as possible (Palmer and Rosenberger, 1999). In this circumstance, learning about treatment effectiveness with an assumption of having a large population outside available, as in the traditional fixed randomised, RCT, also known as *equal fixed randomisation* (EFR), may not be a plausible objective. This, in turn, motivates the use of response-adaptive designs in which clinicians can take advantage of the learning about treatment effectiveness using accruing data to skew the patients towards seemingly superior treatment as the trial progresses (Ahuja and Birge, 2016; Williamson et al., 2017).

In contrast to EFRs, response-adaptive designs favour individual ethics by keeping the balance between *exploration and exploitation*. In fact, this trade-off is typically between *individual ethics*, i.e. exploiting treatments with well enough up-to-date performance, and *collective ethics*, i.e. correctly exploring the better new treatment in case of existence. Hence, balancing this underlying exploration versus exploitation trade-off can be formulated by the *multi-armed bandit problem* (MABP) (Berry and Fristedt, 1985). *Dynamic programming* (DP) Bellman (1957) upon implementation can be a potential method for obtaining the optimal solution of the MABP, although it is expensive in terms of computational complexity.

Indeed, adaptive designs have become popular in many other fields, such as in sociology and education Rafferty et al. (2019), industry and regulatory bodies Lipsky and Lewis (2013), as well as in modern clinical trials. This increasing popularity and significance of adaptive designs, or to be more precise *response-adaptive randomisation* (RAR) procedures, leads us to go beyond the scope of clinical trials and consider the situation as the finite-horizon two-armed bandit problem. It is worth mentioning that this problem naturally appears per se or as a fundamental subproblem in the multi-armed bandit context (Jacko, 2019b). Hence, the majority of the novel findings in this thesis can be applied and implemented in some multi-armed generalisations.

1.1 Outline of Thesis

Chapter 2 provides the general background information and some key concepts that laid the groundwork for subsequent chapters of this thesis. The pertinent literature on randomisation and bandit theory, along with underlying principles of the *estimation* problem, will be outlined in this chapter.

The two-armed bandit model and maximum likelihood estimator and its theoretical aspects will be scrutinised in chapter 3. Also, to capture a general mindset of what we plan to do in this thesis, we touch upon the after-trial studies afterwards. Chapters 4 and 5 broadly discuss how to mitigate the MLE estimation bias using dynamic programming solutions of the MABP at the end and within the trial, respectively. In chapter 4, a novel family of estimators will be introduced, whilst in chapter 5, some novel allocation procedures will be the centre of the attention. The performance of the novel optimistic on inferior dynamic programming (OIDP) and randomised dynamic programming (RDP) procedures will be evaluated in the proposed model in chapter 5. However, in chapter 4, classical dynamic programming (DP) in which allocation rules are optimal and deterministic in solving the proposed two-armed model will be of interest. In chapter 6, the focus moves to the designs with an early stopping possibility. We first define some stopping criteria for running hypothesis testing in the middle of the trial and then compare the outcomes to those without interim analysis inspection.

Chapter 7 concludes this thesis by summarising the main contributions and addressing potential perspectives for further research.

Chapter 2

Methodology and Background

2.1 Randomisation

In the design of experiments, randomisation refers to a random assignment process where the experimental units are allocated across the treatment groups. Randomisation was initially popularised in an agricultural study by Fisher (1926) and soon after began receiving well-deserved attention in clinical studies, as discussed in a review paper by Amberson Jr et al. (1931). The first *randomised controlled trial* (RCTs) came later by Craft et al. (1998). From the probability theory point of view, in an experiment, subjects are randomised to interventions/arms [1], as the patients are randomly assigned to treatments in the clinical trial context.

Randomisation is now a fundamental component of the design of experiments as it (i) provides a comprehensive framework by which one can compare treatment groups distinctly, (ii) eliminates selection bias [2] giving rise to a valid treatment

[1] In this thesis, we use the word subject instead of patient and intervention/arm instead of treatment to follow Jacko (2019b) advice on unifying the terminology across the literature.

[2] Also known as treatment allocation bias in the literature.

group's efficacy estimation, and (iii) lays the groundwork for statistical inference, as discussed in (Rosenberger et al., 2012), (Rosenberger et al., 2019).

The randomisation strategies can be categorised into two classes: (i) *equal fixed randomisation* (EFR), where the allocation ratios (or, equivalently, “allocation probabilities" in e.g. Robertson et al. (2020) because ratios are e.g., 1:1 while probabilities would be 0.5 and 0.5) remain constant during the course of the trial, and (ii) *adaptive randomisation*, where allocation ratios can change within the trial (Chow and Chang, 2008). In fact, adaptive randomisation utilises the accrued data to update the randomisation ratios to achieve the experimental goals. Hu and Rosenberger (2006) classify adaptive randomisation into four main categories: (i) restricted randomisation, the procedure where the number of subjects balances across intervention groups, (ii) covariate adaptive randomisation procedure, where future allocation probabilities are determined based on observed allocation assignments and subject covariate values (iii) response-adaptive randomisation (RAR), a randomisation procedure in which allocation probabilities are adjusted based on accumulating subject responses to minimise the number of failure responses, and (iv) covariate-adjusted response-adaptive (CARA) randomisation procedure which is a combination of covariate adaptive and RAR procedures. Note that in this thesis, we solely focus on *response-adaptive randomisation* (RAR) procedures. To find the details of RAR methods, the readers are referred to (Hu and Rosenberger, 2006), (Rosenberger and Lachin, 2015).

2.2 Response-Adaptive Randomisation (RAR)

Response-adaptive randomisation[3] (RAR) procedures were developed to increase the allocation ratio of assigning subjects to more promising interventions as more information becomes available (Merrell et al., 2021). Apart from the fact that an RAR procedure can be considered either fully *randomised* or *deterministic*, there are some proponents against such approaches in the modern literature. For instance, Thall et al. (2015) addresses some ethical issues, such as the possibility of subject allocation imbalance in the wrong direction and loss of subject benefit enrolled in the trial due to inferential problems, as undesirable properties of the RAR procedures. On the other hand, Rosenberger and Lachin (2015) argue that it is undeniable that RAR mitigates the chance of randomising subjects to inferior interventions to a significant extent. However, it fails to eliminate the ethical problem completely. It is worth mentioning that the aim of utilising an RAR procedure as a general design of the experiment can be classified into three main objectives.

Statistical purposes, such as maximising the power of the trial, are among the first objective. In the RAR literature, “power" is mainly perceived as a frequentist property, i.e. the probability of rejecting a null hypothesis when the alternative is valid. Bear in mind that the power of a trial may have multiple definitions needing to be clearly stated when results are reporded (Robertson et al., 2020). Stallard and Rosenberger (2002) and Villar et al. (2015a) discuss the challenges of overcoming a low power in different RAR methods, whilst Williamson et al. (2017) propose an RAR procedure giving rise to a significant improvement in the power

[3] *outcome-adaptive* or *data-dependent* randomisation is usually used interchangeably in the literature.

of the test in comparison with other RAR counterparts in the literature.

The second common objective of the RAR procedure as a general design of the experiment is minimising *regret number of successes*[4](the difference between the cumulative number of success responses and that of the best fixed decision in hindsight, see e.g. (Kaufmann et al., 2012)) or equivalently maximising *subject benefit* (the number of subjects that are on the superior intervention, see e.g. (Rosenberger et al., 2001), (Villar et al., 2015a)). Setting up minimising the *regret number of successes* in the objective function of a trial is conventional in the machine learning community, Bubeck et al. (2012), whilst maximising the *subject benefit* is often of interest in the Operational Research literature.

The last objective of adopting an RAR procedure is the efficacy estimation. Despite some technical hindrances and difficulties in using the frequentist inference framework on results obtained from an RAR procedure (see, e.g. Proschan and Evans (2020); Rosenberger and Lachin (2015)), standard statistical tests and estimators can be simply used without adjustment in a straightforward manner. However, the estimation, particularly efficacy estimation, is often biased due to statistical dependencies on the responses in an RAR procedure. Hence, a considerable part of the literature has been devoted to addressing this problem by proposing bias-corrected estimators, see (Coad and Ivanova, 2001; Bowden and Trippa, 2017; Hadad et al., 2021).

A common feature of a substantial number of papers in the RAR literature is that they consider more than a single objective for the RAR procedure. Hence, they examine the trade-off between these mutually exclusive objectives in their

[4] Depending on the operating characteristics in a sequential experiment design, one may define *Bayes regret number of successes* or *frequentist regret number of successes*, see (Jacko, 2019b).

proposed designs. For instance, Merrell et al. (2021); Williamson et al. (2017, 2022) suggest an RAR procedure to maximise a utility function whilst balancing statistical power and controlling the type I error rate, i.e. keeping the balance between the first and the second objectives. Hadad et al. (2021), on the other hand, focus on providing an unbiased efficacy estimation whilst optimising the power, i.e. keeping the balance between the first and the last objectives. In this thesis, we mainly focus on the second and third objectives. In other words, we try to mitigate the efficacy estimation bias whilst maximising the subject benefits at the end of the proposed experiments.

Last but not least, responses in an RAR procedure can be discrete or continuous. Although choosing the type of response depends on the context of the experiment, it can be classified into *continuous* and *discrete* types of response. For continuous responses, see, e.g. Williamson and Villar (2020); Hadad et al. (2021), and for discrete ones, see, e.g. Bowden and Trippa (2017); Villar et al. (2015a), and also for composite cases, see, e.g. (Stallard et al., 2020), (Xu et al., 2022). In this thesis, we assume responses are binary as the resultant response-adaptive model is not only motivated by its widespread applicability, but it can serve as a fundamental framework for introducing additional problem features as well (Jacko, 2019b).

2.3 Bandit Models

2.3.1 The Multi-Armed Bandit Problem (MABP)

The Multi-armed bandit problems (MABPs), are a special class of optimal control problems. They define a situation in which a fixed limited set of resources should be allocated between competing intervention choices to maximise the ultimate expected reward over the time horizon. The allocation should be done sequentially and evolve randomly over time. Owing to the exceptional practical potential that MABP offers, it has received well-deserved attention from the Operational Research and Machine Learning communities as well as that in Medical Statistics. The first work on MABP can be attributed to Thompson (1933), which was later continued and developed by Robbins (1952), Bellman (1956), Gittins et al. (2011) and Villar et al. (2015a) in health applications, in particular.

The MABP provides a mathematical framework by which the tension between *learning* (or *exploration*) and *earning* (or *exploitation*) can be formalised. In fact, when one thinks about decision-making under uncertainty and being informed by data, MABP can ideally come into play and maximises any desired objectives, e.g. the overall reward, to achieve an *optimal policy.* Considering the fact that the MABP applications' scope is broad, the most common context motivated by this methodology is clinical trials, Villar et al. (2015a), where the trade-off between two objectives in tension is sought:

- To correctly identify the best treatment, i.e. *exploration* or *learning.*
- To treat patients as effectively as possible within the trial, i.e. *exploitation* or *earning.*

Although deployment of an MABP model can potentially result in optimal solutions and allocation policies, it has yet to be applied to a real-world clinical trial. Hence, we centre our attention around a general usage of MABP, where subjects[5] are allocated to interventions[6] (or arms) with the aim of optimising some predetermined criteria. In fact, whilst we try to maximise the number of successful responses within the trial, we aim for an accurate efficacy estimation at the end of the trial.

2.3.2 Markov Decision Processes (MDPs)

Without a shadow of a doubt, one of the most popular methods of modelling a MABP is as a *Markov decision process* (MDP), known as a particular class of processes which extend Markov processes by the addition of actions. In the MDP framework, a decision-maker encounters a set of decisions (or actions), associated with exclusive rewards, that must be taken at each stage. Hence, to formulate an MDP, it is required to introduce a range of features, namely decision epochs, states, actions, transitional probabilities and rewards (see Puterman (2014) for details).

Decision epochs t, also referred to as "time epoch t" are the points in the time horizon on which a decision must be taken, where we shall take $t \in \mathcal{T} := \Big\{0, 1, 2, ..., T\Big\}$, $T < +\infty$. All information required to choose an action from action set $\mathcal{A}$ at time epoch t is summarised the state $\boldsymbol{x}_t$. In this thesis, the state represents the up-to-date available knowledge about the effectiveness of an intervention. States can be updated once a new subject's response has been observed. In turn,

[5] Known as "patient" in Biometrics and Biostatistics, or "resource" in Operation and Management, as well as Economics.

[6] Equivalent to "treatments", or "bandits", depending on the context.

an action is taken based on the state and corresponds to allocating a subject to an intervention. Note that actions can be deterministic if the probability of an action being selected is 1 or randomised if each is selected with some probability. For a randomised-action case, see e.g. (Cheng and Berry, 2007). The main focus of this thesis is on deterministic actions, although we briefly touch upon randomised cases in chapter 5. After taking action a_t at time epoch t, the system transfers to a new state at time epoch $t+1$, $\boldsymbol{x}_{t+1}$, based on the transitional probability $P(\boldsymbol{x}_{t+1}|\boldsymbol{x}_t, a_t)$, and an associated reward $r^a_{\boldsymbol{x}}$ accrues to evaluate the chosen action. The transitional probabilities and rewards at each time epoch t depend only on the current state and action chosen in that state, which leads us to a Markovian system.

In formulating an MDP, the time horizon, T, can be considered finite or infinite depending on the problem circumstances. For the latter case, a discount factor $d \in (0, 1)$ is often introduced to ensure that *total* reward is finite. Additionally, the objective function of an MDP is typically considered the *total reward*, which is maximised as *expected total reward* if the time horizon is finite, see Puterman (2014), chapter 8, and as *expected total discounted reward* if the time horizon is infinite, see (Bellman, 1956), and ((Puterman, 2014), chapter 7).

It is worth mentioning that the *undiscounted* reward case is equivalent to setting the discount factor $d = 1$. In some works such as Wang (1991); Hardwick et al. (1991) the latter is referred to as *uniform discounting*. Hence, as the assumed horizon in this thesis is *finite*, we presume $d = 1$ and set the objective function of interest to maximise the *expected total reward*. Let $\boldsymbol{X}(t) = \boldsymbol{x}$ be the state of the system at time epoch t, and let $\mathbb{E}^{\pi}_{t}[.]$ refer to the expectation under policy[7]

[7] A *policy* is a mapping from the states set to the actions set. In other words, it is a rule

$\pi \in \Pi$ conditioned on information available up to the time epoch $t \in \mathcal{T}$. Then the expected total reward over the remainder of the time horizon $[t, T]$ is

$$\mathbb{N}_t^{\pi}(\boldsymbol{x}) := \mathbb{E}_t^{\pi}\left[\sum_{u=t}^{T} r_{\boldsymbol{X}(u)}^{A(u)} \,\middle|\, \boldsymbol{X}(t) = \boldsymbol{x}\right] \tag{2.1}$$

where $A(u)$ represents the action that is chosen at time epoch $u = t, ..., T$ under policy π and $r_{\boldsymbol{X}(u)}^{A(u)}$ is the reward received upon taking action $A(u)$. Finally, by maximising the expected total reward over the pre-determined time horizon, T, the *optimal policy* can be ultimately obtained. Note that the existence of an optimal policy for an MDP with a finite horizon is proved in (Berry and Fristedt, 1985).

2.3.3 Dynamic Programming (DP) Approach

A MABP, in principle, can be solved by two possible methods: the Dynamic Programming (DP) technique and an index-based approach. The latter gives rise to sub-optimal solutions, whilst the DP results in a Bayes-optimal solution, which is an exact one. There is also another important distinction here: the work due to Gittins (1979); Villar et al. (2015b) show that there are index-based (optimal) solutions to certain classes of infinite horizon MABPs with discounting, although our focus is on the finite horizon case in which index based optimal solutions do not exist necessarily, and therefore the deployment of DP is always available instead. Since MABPs can be formulated as MDPs, in this thesis, we utilise the

which results in a sequence of actions, each of which is taken in light of the information available when the action is chosen in the state $\boldsymbol{x}$ at time epoch t. In this thesis, we assume policies are *past-measurable*, or *history-dependent* in Puterman (2014), or *non-anticipating* in Jacko (2019b) at any time epoch t. All variations mean that the policy does not depend on what happens after time epoch t.

standard DP approach introduced and popularised by Bellman (1957) and solve the proposed bandit models in an exact manner.

Informally, the fundamental idea of DP is based on enumeration. It breaks the problem down into a series of sub-problems and then solves and stores each separately. To complete the solution of the original problem, it re-uses and puts all sub-problems together, which in turn, leads to a significant reduction in computational burden.

Put technically, DP iterative procedure is based on *value function* calculations aimed at maximising the expected total rewards over the set of all policies π, from each time epoch t. The value function, F_t, stands for the best possible value of the objective in (2.1). Starting at time epoch t with state $\boldsymbol{x}$ the value function is defined as below

$$F_t(\boldsymbol{x}) = \max_{\pi\in\Pi} \mathbb{N}_t^{\pi}(\boldsymbol{x}) = \max_{\pi\in\Pi} \mathbb{E}_t^{\pi}\left[\sum_{u=t}^{T} d^{(u-t)} r_{\boldsymbol{X}(u)}^{A(u)} \,\middle|\, \boldsymbol{X}(t) = \boldsymbol{x}\right] \tag{2.2}$$

where d is the discount factor. Following on the fundamental property of the value function satisfying the below recursive formula, known as the *Bellman equation*[8], we have

$$F_t(\boldsymbol{x}) = \max_{a\in\mathcal{A}}\left\{ r_{\boldsymbol{x}}^{a} + d\sum_{\boldsymbol{x}'} P(\boldsymbol{x}'|\boldsymbol{x},a)F_{t+1}(\boldsymbol{x}')\right\} \quad for \quad 0 \leq t \leq T-1 \tag{2.3}$$

where $P(\boldsymbol{x}'|\boldsymbol{x},a)$ is the transitional probability of transferring from state $\boldsymbol{x}$ at time epoch t to some new state $\boldsymbol{x}'$ at time epoch $t+1$ under the action a. Moreover,

[8] Known initially as the *functional equation* by Bellman (1957). Other name alternatives are also the *fundamental equation of dynamic programming* by Berry and Fristedt (1985); the *dynamic programming equation* by Gittins et al. (2011); the *optimality equation* by Puterman (2014), to name but a few.

it is worth mentioning that (2.3) is composed of (i) the *immediate reward*, $r^a_{\boldsymbol{x}}$, and (ii) the expected (discounted) *future reward*, i.e. the second term. The latter is realised by implementing an optimal policy from time $t+1$ onwards, whilst the former is earned instantly from action a chosen at state $\boldsymbol{x}$. In other words, the future reward is the multiplication of the expected total reward obtained by following the policy π from time epoch $t+1$, where the following state, $\boldsymbol{x}'$, is achieved if action a is taken at time epoch t, to T, and the probability of being in the state $\boldsymbol{x}'$ itself. Hence, at each time epoch, i.e. in every state, the action that maximises the aggregation of both rewards is chosen.

The final optimisation problem is to find the maximum expected total reward over the whole time horizon given an initial state $\boldsymbol{x}_0 = \boldsymbol{x}$, that is, $F_0(\boldsymbol{x})$, at the beginning of the trial, $t = 0$. The optimal policy, π^*, which determines the sequence of optimal actions giving rise to the ultimate expected total reward, also can be expressed as

$$\pi^*(\boldsymbol{x}_0) := \arg\max_{\pi\in\Pi}\mathbb{N}^{\pi}_0(\boldsymbol{x}) \tag{2.4}$$

Note that *arg max* is an operation that finds the argument that gives the maximum value from a target function. It can also be defined as a set of points for which the target function attains the largest value. Hence, $\pi^*(\boldsymbol{x}_0)$ is defined as the set of all optimal policies.

Backward Induction

We utilise *backward induction*, which is the process of evaluating the value functions recursively backwards in time, from the *final* time epoch and for all possible states until the start of the problem at $t = 0$, to determine a sequence of optimal actions, and subsequently solve the problem 2.4 in an exact manner.

The idea is that no action is taken at the last time epoch, $t = T$, assuming a finite horizon situation. In fact, the final decision should be taken at $t = T - 1$, which, in turn, implies that the terminal reward is an immediate reward corresponding to the state only, i.e. $F_T(\boldsymbol{x}) = r_{\boldsymbol{x}}$. In the Operational Research literature, final value function, usually considered 0, is referred to as the *salvage*, or *scrap value* (Puterman, 2014). In chapter 3, we will discuss after-trial studies where the terminal reward is not set to 0. A similar approach which avoids some specific states by setting up penalties, known as *constrained randomised dynamic programming* (CRDP), is investigated by (Williamson et al., 2017). The maximum expected reward for every possible state is calculated for the rest of the time epochs in the decision process. Note that this calculation also incorporates the information obtained from the subsequent time epoch. The algorithm continues determining the optimal reward and the corresponding optimal policy until the beginning of the problem at time epoch $t = 0$. The backward induction algorithm can be summarised as follows:

1. Let $t = T$ and $F_T(\boldsymbol{x}) = r_{\boldsymbol{x}}$ for all $\boldsymbol{x} = \boldsymbol{x}_T$,
2. For $t = T - 1, T - 2, ..., 0$ and for each $\boldsymbol{x} = \boldsymbol{x}_t$, calculate:

(I)

$$F_t(\boldsymbol{x}) = \max_{a \in \mathcal{A}} \left\{ r_{\boldsymbol{x}}^a + d \sum_{\boldsymbol{x}'} P(\boldsymbol{x}'|\boldsymbol{x}, a) F_{t+1}(\boldsymbol{x}') \right\}$$

(II)

$$\pi_t^*(\boldsymbol{x}) = \arg\max_{a \in \mathcal{A}} \left\{ r_{\boldsymbol{x}}^a + d \sum_{\boldsymbol{x}'} P(\boldsymbol{x}'|\boldsymbol{x}, a) F_{t+1}(\boldsymbol{x}') \right\}$$

3. If $t = 0$, stop. If not, repeat step 2.

The algorithm terminates with returning an optimal policy (or policies), which is determined stage by stage form step II in above algorithm, $\pi^*(\boldsymbol{x})$, from state $\boldsymbol{x} = \boldsymbol{x}_0$, together with $F_0(\boldsymbol{x})$. The latter contains the maximum expected (discounted) sum of rewards achieved by following the former. Hence, two multi-dimensional arrays are required to be defined to implement the backward induction algorithm: (i) an array recording the F values corresponding to each state, and (ii) one for optimal policies π^* containing the sequence of actions leading to these F values. It is noteworthy that this algorithm in detail can be found in work by Williamson et al. (2017), and also in Appendix 4.6.1 of chapter 4.

Computational Complexity

The practicality of DP is often limited, since its computational complexity, also known as *curse of dimensionality*, is an inseparable feature of the DP approach. Bellman (1961) states that as the number of interventions increases, the size of the problem grows exponentially. As an example demonstrating the computational complexity associated with the DP, consider the problem of allocating subjects in a trial of size $T = n$ with two available arms (C and D) and binary response (success or failure). This results in 4^n possible paths that need to be enumerated, as four possible outcomes will exist at each time epoch t. Also, for the relationship between the DP's computational requirements and the problem's horizon, T, when a small number of arms are considered, see (Villar et al., 2015a). Moreover, Zhang et al. (2019) show that solving the DP becomes infeasible if one considers three arms for a horizon of 100. On the other hand, owing to the advancement in computers' memory and configuration, Sutton and Barto (2018), and Jacko (2019b) show that the DP procedure can be used to solve MDPs for much larger problem sizes despite taking time for the final solution. For instance, Jacko (2019a) develops a computer package by which MABPs with binary responses can be solved by DP in a few minutes for $T \approx 1000$, a few hours for $T \approx 2000$ or a few days for $T \approx 4000$.

Several methods have been proposed to overcome the computational complexity problem in the operational research and bandit literature. In particular, when the dimension of the state space is too large, the value function can be approximated heuristically instead of attempting exact calculation. This procedure introduces *approximate dynamic programming* (ADP), which is beyond the scope of this thesis. The interested reader is referred to (Powell, 2007).

2.4 Estimation

The estimation of the efficacy of each participating arm, as the third objective of the RAR procedures mentioned in section 2.2, is now one of the state-of-the-art focal points in many recent research works. Apart from the estimation method and the estimator type, providing an unbiased estimator is often the primary desired objective in the RAR literature; see, for example (Whitehead, 1986), (Luedtke and Van Der Laan, 2016), and (Nie et al., 2018). From a general point of view, an efficacy estimation can be done during or at the end of RAR procedures. Each of these possibilities is associated with a particular objective. For example, the latter can aim for providing a more accurate conclusion about interventions' efficacy for post-trial applications, whilst the former may serve as a parameter estimation tool to fit the best model with the aim of doing regression analysis (Liu and Chen, 2016).

Estimating a parameter (or parameters) can be done via different approaches. Considering a linear regression model, the *ordinary least squares estimation* (OLS) technique, which is a type of *linear least squares* (LLS) method, is often used for estimating the unknown parameters of interest. OLS, or LLS more generally, estimates parameters by minimizing the squared discrepancies between observed data and their expected values. This approach is prevalent in the MABP setting (mainly when the number of arms is more than two), where the parameters of interest to be estimated are the arm selection probabilities Zhang et al. (2020), or arms' probability distributions, see the survey by (Burtini et al., 2015). Another approach to estimating an unknown parameter (or parameters) is the *maximum likelihood estimation* (MLE), which is a method of estimating the parameters by maximising the likelihood function derived from a given probability model, given some observed data. Hence, this estimation approach is typically used at the end of the trial when some responses are observed (Bowden and Trippa, 2017; Marschner, 2021).

In the sequential decision-making process, or to be more specific, the clinical trials context, the estimation can be applied to various operating characteristics and studies such as success probabilities, survival rate, and dose-finding studies, to name but a few. For example, assessing the chance of detecting treatment effects in clinical trial designs are investigated by (Han et al., 2022). After deciding which are the parameter(s) of interest to be estimated, choosing an appropriate estimator will be the next step.

2.4.1 Estimation Bias

One of the principal characteristics of an estimator is bias. According to the definition, the bias of an estimator is the difference between the expectation of the estimator and the parameter's true value. Let θ be the actual parameter of interest and $\widehat{\theta}$ be the estimator; then the bias is defined by

$$\text{Bias}\left[\widehat{\theta}\right] := E\left[\widehat{\theta} - \theta\right] = E\left[\widehat{\theta}\right] - \theta \tag{2.5}$$

An estimator is unbiased if its bias is equal to zero for all values of parameter θ, or equivalently if its expected value matches the parameter. In practice, it is sometimes rare that an estimator is unbiased, as many sources can give rise to an estimation procedure being associated with bias. Although estimation with a small amount of bias is acceptable in practice, several techniques and novel estimators have been recently proposed to eliminate estimation bias[9] in various settings.

It has been proved that the sample means are biased estimators of the treatment success probabilities in an RAR settings (Villar et al., 2015a). In contrast to the *equal fixed randomisation* (EFR) approaches, in which the sample sizes are pre-fixed and also the proportions of allocation to participating arms can be determined beforehand, in an RAR context, sample sizes are random variables as allocation ratios can be altered as the trial goes on (Stallard et al., 2020). In fact, such adaptivity imposes complex correlations between the data collection procedure. Additionally, sample mean estimation bias can be either negative or positive, depending on data collection, i.e. the RAR circumstances, see, (Nie et al., 2018;

[9] Also known as sample mean bias or selection bias, estimation bias of efficacy, and *maximum likelihood estimator* (MLE) bias.

Shin et al., 2019a,b). The challenge of eliminating efficacy estimation bias has recently been extensively studied in the RAR context. This process of elimination has been investigated via two approaches: (i) proposing a novel debiasing algorithm, Nie et al. (2018), and (ii) introducing a novel estimator (or family of estimators), which is asymptotically unbiased (Hadad et al., 2021).

In this thesis, we mainly focus on the *maximum likelihood estimator* (MLE) as an estimator for the actual arms' efficacies (or success probabilities) in the proposed trial designs. We do so because MLE can be a decent estimator for deciding on the required sample size since, for instance, the calculation of the sample size in clinical trials requires the specification of the treatment effect for which the study is powered (Wassmer and Brannath, 2016). We also aim to improve the efficacy estimation to draw a correct conclusion about the future efficiency of participating interventions. This conclusion can be obtained from previous and similar clinical trials, pilot studies, or preceding trial phases. Hence, we are interested in MLEs being estimated with less bias and as accurately as possible.

2.5 Simulation Set-up

Simulation is one of the popular and straightforward tools for evaluation MABPs by which the accuracy vs runtime trade-off, using a decent number of simulation runs, can be addressed (Villar et al., 2015b; Jacko, 2019b). Since this thesis is mostly built around simulation results, we evaluate all proposed trial designs assuming a broad range of two-arm trial scenarios. What we do in the simulation stage is, for a given trial solved by DP, in each simulation replication, we run the trial using DP-based solutions and record the total number of success and failure observations obtained from each arm at the end of the trial. Actually, this process can be summarised as follows: after the first couple of deterministic allocations explained below, the next allocation is determined by flipping a coin in the third time epoch. From the third time epoch onward, the following allocation is determined according to the DP solutions, which, in turn, are chosen based on the up-to-date observed numbers of success and failure responses. The process is continued up to the last time epoch, where the ultimate total number of success and failure observations on both arms are saved in the memory. It is noteworthy that the simulation stage is more or less the same for all proposed designs. However, the DP procedures used to solve the two-armed model differ from design to design and are determined based on the trial design requirements. For instance, in chapter 4, all DP algorithms are amongst classical deterministic procedures, whilst in chapter 5, they are mainly amongst randomised or some novel procedures.

Let C stand for the control arm and D for the experimental counterpart, and subsequently, θ_C and θ_D denote the true corresponding success probabilities. Then, we take scenarios to be in a broad range of $\theta_C \in (0,\ 0.1, ..., 1)$ and $\theta_D \in (0,\ 0.1, ..., 1)$ where all symmetric combinations are removed. Note that scenarios $(0.1, 0.2)$ and $(0.2, 0.1)$, for example, return the same simulation results if we mirror the θ_C and θ_D. Hence, to save the running time and memory intake, we remove all scenarios in which $\theta_C > \theta_D$ and deal with those $\theta_C \leq \theta_D$. One may presume the general trend of assumed scenarios is similar to an upper-triangular matrix, which results in having 66 different scenarios, i.e. $(12 \times 11)/2$. Moreover, in all plots representing the simulation results in this thesis, the bias of the MLE is depicted by circles for control arm C and stars for experimental arm D.

It is worth mentioning that the frequentist estimator (and its bias) is affected by the researchers' choice in dealing with situations where there are no realised allocations to one of the arms. These situations can be classified into two circumstances: (i) removing these outcomes from the sample space or (ii) forcing to have at least one observation on each arm. In this study, we apply the latter case in which each arm is allocated once in the first two time epochs through all proposed trials regardless of estimator type. Note that the former may give rise to some other type of bias, such as selection bias[10].

Furthermore, we set the trial sizes as a multiple of 60: $T = 60, 120, 180$ and 240, to have the potential of participating up to 6 arms in the trials. In turn, having an equal number of observations can be achievable for each. We also make use of colours to show the different 66 scenarios consisting of a pair of success probabilities: $(\theta_C, \theta_D) = (0,0), (0,0.1), ..., (0.9,1), (1,1)$. To do so, we introduce *arm D effect* parameter which is the absolute difference between θ_D and θ_C, i.e. $|\theta_D - \theta_C|$. For example, colour *green* represents simulation results corresponding to the scenarios in which arm D effect is zero, i.e. $\theta_D = \theta_C$[11]. For all 11 chosen colours and their pertinent family of scenarios, please see table 2.1.

Finally, the number of simulation replications is set to one million for each scenario and therefore, the parameter of interest is calculated by averaging out all obtained simulation values at the total number of simulation replications, which is one million. This number of simulation replications makes our conclusions and calculations more accurate and convincing.

[10] Selection bias is the bias introduced by the selection of individuals, groups, or data for analysis in such a way that proper randomization is not achieved, thereby failing to ensure that the sample obtained is representative of the population intended to be analysed.

[11] This family of circumstances can also be named the *null scenarios*.

$\lvert\theta_D - \theta_C\rvert$	0	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9	1
Colour	**green**	**blue**	**black**	**purple**	**red**	**yellow**	**orange**	**cyan**	**pink**	**violet**	**brown**

Table 2.1: Colour-scenario information

2.6 Memory in Use

This section addresses a limitation on the computational capacity of the backwards induction algorithm we encountered for all proposed trial designs throughout this thesis. Note that the allocation policy of the DP procedure is computed on a standard laptop with 16 GB of RAM using Julia (programming language). The maximum trial size we have been able to compute via the mentioned configuration is 240. In addition, upon implementing the backwards induction algorithm, two multi-dimensional arrays are required to be defined: (i) an array recording the F values corresponding to each combination of states, and (ii) one for optimal policies π^* containing the sequence of actions leading to these F values. The limitation is imposed on the former when the trial size is $T = 240$ and the accuracy of rounding (number of digits) matters. Hence, we consider two different cases for setting up the value function array: (i) elements in the value function matrices are rounded by a single-precision floating-point format (float32) whilst the bound of the error is set to $\epsilon = 10^{-7}$ for trial sizes $T = 60, 120$, and 180, (ii) elements in the value function matrices are rounded by a double-precision floating-point format (float64) whilst the bound of the error is set to $\epsilon = 10^{-16}$ for trial size $T = 240$, solely.

Chapter 3

Model and After-trial Studies

3.1 The two-armed Bayesian Beta-Bernoulli model

In this section, we formulate the two-armed Bayesian Beta-Bernoulli model with binary responses as a Markov decision process using the terminology of (Jacko, 2019b). We consider arms labelled by $k \in \mathcal{K} := \Big\{C, D\Big\}$, and the response set by $\in \mathcal{O} := \Big\{0, 1\Big\}$ where 0 and 1 stand for a failure and success, respectively. The responses are uncertain, and the success probability of arm k is modelled as Bernoulli-distribution [1] with parameter $0 \leq \theta_k \leq 1$. Subjects arrive (i.e. are recruited) at discrete time epochs $t \in \mathcal{T} := \Big\{0, 1, 2, ..., T-1\Big\}$, where $T < +\infty$ is the time horizon or experiment/trial size. In addition, we assume that responses are observable immediately, meaning that a response of the current subject is observed before taking the next subject's allocation decision. To represent states in this model, we use a four element vector $\boldsymbol{x} := \Big(s_C, f_C, s_D, f_D\Big)$, where s_k and

[1] In probability theory and statistics, the Bernoulli distribution is a discrete probability distribution of a random variable which takes the value 1 with probability θ and the value 0 with probability $1 - \theta$.

f_k represent the number of observed successes and failures on arm k respectively. Moreover, the probability of observing response $o \in \mathcal{O}$ in state $\boldsymbol{x}$ if the current subject is allocated to arm $k \in \mathcal{K}$ is denoted by $q_{k,x,o}$, with $\sum_{o \in \mathcal{O}} q_{k,\boldsymbol{x},o} = 1$. Assuming that at every time epoch $t \in \mathcal{T}$ a randomised actions a identified by a pair (p_C^a, p_D^a) in which the arrived subject is allocated to arm $C(D)$ with probability $p_C^a(p_D^a)$, the *action set* $\mathcal{A}$ can formally be taken as $\mathcal{A} = \Big\{a;\ p_C^a + p_D^a = 1\Big\}$. Since we consider the allocation procedure to be deterministic, and randomisation (with equal probabilities) occurs if only there is no difference in allocating either arm, we have: $p_C^a,\ p_D^a \in \{0, 1, 1/2\}$. In turn, the expected value of observing one success at a time epoch $t < T$ can be defined as $r_{\boldsymbol{x}}^a = p_C^a . q_{C,\boldsymbol{x},1} + p_D^a . q_{D,\boldsymbol{x},1}$, and $r_{\boldsymbol{x}} = 0$ for the final epoch $t = T$.

Since success probabilities are unknown we take the Bayesian approach and assume that θ_k is a random variable drawn from the Beta distribution[2]. That is, we use the Bayesian Beta-Bernoulli model for each arm $k \in \mathcal{K}$ and consider that at the beginning of the trial, $t = 0$, each arm is given a prior Beta distribution with parameters $(\widetilde{s_k}(0), \widetilde{f_k}(0))$. Three uninformative prior Beta distributions are typically considered in the literature. The most conventional one is the Bayes prior with parameters $\Big(\widetilde{s_k}(0), \widetilde{f_k}(0)\Big) = (1, 1)$ while the others include the Jeffreys prior $\Big(\widetilde{s_k}(0), \widetilde{f_k}(0)\Big) = (1/2, 1/2)$ and the Haldane prior $\Big(\widetilde{s_k}(0), \widetilde{f_k}(0)\Big) = (0, 0)$.

Owing to the fact that the Beta distribution is a conjugate prior with respect to the Bernoulli likelihood, the posterior distribution follows another Beta distribution with parameters $\Big(\widetilde{s_k}(t), \widetilde{f_k}(t)\Big)$ containing initial prior (interpreted as pseudo-

[2] In probability theory and statistics, the Beta distribution is a family of continuous probability distributions defined on the interval $[0, 1]$ parametrised by two positive shape parameters, denoted by α and β (in this thesis, they stand for the number of success, s, and failure, f, observations, respectively) that appear as exponents of the random variable and control the shape of the distribution.

observations) as well as up-to-date observed data, i.e $\widetilde{s_k}(t) = \widetilde{s_k}(0) + s_k(t), \widetilde{f_k}(t) = \widetilde{f_k}(0) + f_k(t)$, where $s_k(t)$ and $f_k(t)$ stand for the numbers of success and failure responses observed up to time epoch t (prior excluded) on arm k, respectively. Consequently, the probability of observing response $o \in \mathcal{O}$ can be defined as follows:

$$q_{k,\boldsymbol{x},o} = \begin{cases} \frac{\widetilde{s_k}}{\widetilde{s_k}+\widetilde{f_k}} & \text{if } o = 1 \\ \frac{\widetilde{f_k}}{\widetilde{s_k}+\widetilde{f_k}} & \text{if } o = 0 \end{cases} \tag{3.1}$$

Next, we define an objective function, in which we are looking for the policy maximising the Bayes-expected number of successes, i.e. the subject benefit in our terminology and the patient benefit in clinical trials:

$$\pi^* := \text{argmax}_{\pi \in \Pi} \mathbb{N}_0^\pi \Big(\mathbf{0}\Big) \tag{3.2}$$

where

$$\mathbb{N}_t^\pi(\boldsymbol{x}) := \mathbb{E}_t^\pi \left[\sum_{u=t}^{T} r_{\boldsymbol{X}(u)}^{A(u)} \,\middle|\, \boldsymbol{X}(t) = \boldsymbol{x} \right] \tag{3.3}$$

where $\mathbb{E}_t^\pi[.]$ refers to the expectation under policy $\pi \in \Pi$ conditioned on information available up to the time epoch $t \in \mathcal{T}$, $A(u)$ is the action prescribed by policy π at time u and $\boldsymbol{X}(u)$ is the state at time u. The state process is such that at every time epoch t, the state $\boldsymbol{x}(t)$ satisfies $s_C(t) + f_C(t) + s_D(t) + f_D(t) = t$. The two-armed Bayesian Beta-Bernoulli model described above can be solved by DP via backward induction based on the calculation of optimal value functions $\mathbb{N}_t^{\pi^*}(\boldsymbol{x})$ for all possible states starting from the final time epoch $t = T$. Please see section 2.3.3. Applying the Bellman equation in every single state solves the model to optimality. By applying Bellman equations to any given time epoch $t < T$, and under an optimal policy, the expected total reward i.e. the Bayes-expected number of successes on both arms, for time epoch $t+1$ to T, is

$$\begin{aligned} F_t^C\Big(s_C, f_C, s_D, f_D\Big) =& q_{C,(x,i),1}\cdot\Big(1 + F_{t+1}\Big(s_C+1, f_C, s_D, f_D\Big)\Big) \\ &+ q_{C,(x,i),0}\cdot\Big(0 + F_{t+1}\Big(s_C, f_C+1, s_D, f_D\Big)\Big) \\ F_t^D\Big(s_C, f_C, s_D, f_D\Big) =& q_{D,(x,i),1}\cdot\Big(1 + F_{t+1}\Big(s_C, f_C, s_D+1, f_D\Big)\Big) \\ &+ q_{D,(x,i),0}\cdot\Big(0 + F_{t+1}\Big(s_C, f_C, s_D, f_D+1\Big)\Big) \end{aligned} \tag{3.4}$$

therefore, based on the *principle of optimality*, for any $0 \le t \le T-1$ we have:

$$\begin{aligned} &F_t\Big(s_C, f_C, s_D, f_D\Big) = max\Big\{F_t^C\Big(s_C, f_C, s_D, f_D\Big), F_t^D\Big(s_C, f_C, s_D, f_D\Big)\Big\} \\ &F_t\Big(s_C, f_C, s_D, f_D\Big) = 0, \text{otherwise.} \end{aligned} \tag{3.5}$$

Note that, for full generality, Jacko (2019b) recommends considering the information state i, which is potentially dependent on the current physical state $\boldsymbol{x}$ and any other changing during the trial. This may include real-world evidence and/or modelling assumptions. The latter could be the type of prior distribution assumed in the trial, while the former can be the probability of mistakes in statistical analysis and/or administration processes. Hence, the states can be defined by a pair of $(\boldsymbol{x}, i)$. Finally, it is important to point out that our DP model allows for $50:50$ randomisation when the two value functions are equal. In other words, although the allocation procedure determined by DP is purely deterministic, there are several allocation decision states where the chance of selecting an arm as optimal is 0.5 as the up-to-date arms' performances (value functions) are the same for both.

3.2 Maximum Likelihood Estimator (MLE) and its Bias

This section presents a brief overview of the proposed estimation characteristics. Note that estimation criteria, along with corresponding proofs, are explained in the next chapter in detail. For now, we briefly characterize the frequentist MLE and its bias in the context of the designs with binary responses. We follow and build our assumptions based on the framework offered by Bowden and Trippa (2017) to develop the bias of the MLE. To do so, we assume that unknown success probabilities θ_k are Beta-distributed with parameters α_k and β_k:

$$\theta_k \sim Beta(\alpha_k, \beta_k) \quad \forall k \in \mathcal{K}$$

The maximum likelihood estimator at time epoch $\tau \in \mathcal{T} \cup \{T\}$, i.e., having observed $s_k(\tau)$ successes and $f_k(\tau)$ failures on each arm $k \in \mathcal{K}$, is

$$\widehat{\theta}_k(s_k(\tau), f_k(\tau)) = \begin{cases} \dfrac{s_k(\tau)}{n_k(\tau)} & \text{for } n_k(\tau) > 0 \\ \text{any value in } (0,1) & \text{for } n_k(\tau) = 0 \end{cases} \tag{3.6}$$

where $n_k(\tau) = s_k(\tau) + f_k(\tau)$. The bias of MLE is formulated in the study by Bowden and Trippa (2017), whilst Shin et al. (2019a) also proposed applying a similar expression for trials with an early and/or adaptively stopping time. The bias of the MLE is

$$\mathrm{Bias}\Big[\widehat{\theta}_k(\tau)\Big] := E\Big[\widehat{\theta}_k(\tau)\Big] - \theta_k = \begin{cases} \dfrac{-\mathrm{Cov}\Big[n_k(\tau), \widehat{\theta}_k(\tau)\Big]}{E\Big[n_k(\tau)\Big]} & \text{for} \quad n_k(\tau) > 0 \quad \forall k \in \mathcal{K} \\ 0 & \text{for} \quad n_k(\tau) = 0 \end{cases} \tag{3.7}$$

Note that in designs where the allocation is independent of past responses, e.g. EFR with arbitrary randomization ratio, we will have a zero covariance and thus a zero bias on each arm. In contrast, for the designs in which the subject allocation depends on (or is correlated with) past responses, a particular arm may have a bias away from zero. Note that it can be concluded from equation (3.7) that a significant bias will be obtained if the covariance is away from zero and the sample size is small. The sign of the bias is typically negative for each arm under the response-adaptive allocation procedures, which aim at (exactly/approximately/asymptotically) maximizing the expected number of observed successes since these tend to allocate more subjects to arms with the higher estimated success probability. However, this is not guaranteed by the above characterization in general, as shown in Nie et al. (2018) where sufficient conditions are provided for having the estimation bias positive or negative, depending on the trial circumstances. In particular, negative bias is not guaranteed for those designs that use other estimators rather than the MLE for allocation decisions, which, among others, include upper confidence bound procedures and Bayesian procedures. Furthermore, the sign of the bias can be positive for some or all arms for some procedures (Shin et al. (2019a), Shin et al. (2019b)), for instance, for those that aim at maximizing the statistical power, which, in some circumstances, tend to allocate more subjects to arms with lower estimates (Rosenberger et al., 2001).

It is noteworthy to mention that the underlying principle of this thesis is to estimate the MLE using the optimal (and suboptimal) solutions that the DP procedure provides via solving (3.2). As the two-armed Bayesian Beta-Bernoulli model described above is under the RAR procedures umbrella, the arms' effectiveness is estimated with bias. In other words, the MLE estimation using the results stemming from the model above is often associated with bias. Despite the fact that in this thesis, we meticulously explore novel techniques to correct/mitigate the estimation bias, there is a very old idea addressing this trade-off between the accuracy of estimation and subject benefit via *after-trial* studies which is due to (Berry and Eick, 1995). We take advantage of this idea to prepare readers' minds about what we plan to do next in this thesis since after-trial studies suffer from some computational issues discussed in the section below.

3.3 After-Trial Studies

A traditional way of correcting the estimation bias in the Beta-Bernoulli two-armed bandit model is to consider an after-trial population, S, *outside* of the trial so that the optimal criterion is defined for both *inside*, T, and *outside* populations, see (Berry and Eick, 1995; Cheng and Berry, 2007; Zhang et al., 2019). An example of this criterion, which is more or less equivalent to what we assume here, can be a situation where all outside populations are allocated to the arm with the best performance within the trial. Note that what we assume here is to provide a more accurate efficacy estimation by manipulating the arms' value functions F whilst keeping patients (subjects) well-being as high as possible within the trial.

In this circumstance, the multiplication of the maximum current estimate of the arms' efficacy and the remaining population, $(T + S) - n$ (assuming n number of responses have been observed within the trial) results in the expected subject benefit, i.e. the new value functions F. Note that this type of response-adaptive allocation procedure is known as *robust Bayes* (RB) and is described and compared to other randomisation procedures in (Berry and Eick, 1995). For another example of RB, please see (Berry and Stangl, 1996).

We begin with setting up a new definition for the value functions instead of taking them as zero when the trial terminates. Currently, in the backward induction recursion, we assumed the value functions to be at zero when the trial reaches its end, i.e., when $s_C(t) + f_C(t) + s_D(t) + f_D(t) = T$. In other words, if $t = T$, there is nothing to do since all subjects have already been randomised and their outcomes observed. Now, considering the fact that there is a cohort of subjects outside of the trial where either arm will be allocated to all of them leads to setting the value functions to be the multiplication of their current belief, i.e. the Bayes-expected number of successes and the after-trial population size at the end of the trial. Thus, we have:

- if $t = T$:

$$\begin{aligned} F_t^C\Big(s_C, f_C, s_D, f_D\Big) &= q_{C,(x,i),1}.S \\ F_t^D\Big(s_C, f_C, s_D, f_D\Big) &= q_{D,(x,i),1}.S \end{aligned} \tag{3.8}$$

- if $t < T$:

$$\begin{aligned} F_t^C\Big(s_C, f_C, s_D, f_D\Big) =& q_{C,(x,i),1} \cdot \Big(1 + F_{t+1}\Big(s_C + 1, f_C, s_D, f_D\Big)\Big) \\ &+ q_{C,(x,i),0} \cdot \Big(0 + F_{t+1}\Big(s_C, f_C + 1, s_D, f_D\Big)\Big) \\ F_t^D\Big(s_C, f_C, s_D, f_D\Big) =& q_{D,(x,i),1} \cdot \Big(1 + F_{t+1}\Big(s_C, f_C, s_D + 1, f_D\Big)\Big) \\ &+ q_{D,(x,i),0} \cdot \Big(0 + F_{t+1}\Big(s_C, f_C, s_D, f_D + 1\Big)\Big) \end{aligned} \tag{3.9}$$

3.3.1 Average Estimation Bias

We ran the model with three different multiples of 1000 for the after-trial population, $S = 1000$ and $1,000,000$ (1 Million) and $1,000,000,000$ (1 Billion), and also for four different trial sizes, $T = 60, 120, 180$ and 240. Recalling criteria mentioned in section 2.5, we note that all plots depict the relationship between the bias of the MLE (horizontal axis) and the covariance between the MLE and sample size (Vertical axis) for both arm C (circles) and arm D (stars). It is interesting to mention that the values on the vertical axis, which represents the covariance between the MLE and sample size, are actually those obtained from the numerator of equation (3.7) with an opposite sign.

It is also worth mentioning that in the DP procedure, the value functions in the design with an after-trial population can take values from 0 to $T + S$. In turn, in the circumstances where the trial size and after-trial population are large, considering an appropriate level of accuracy in defining the value function matrices along with the bound used for absolute differences to find the optimal arm can give rise to significantly different estimation results. Hence, we present the MLE estimation results for all trial sizes and all above-mentioned after-trial populations in two different accuracy cases through the DP procedure: (i) elements in the value function matrices are saved by a single-precision floating-point format (float32) whilst the bound of the "*error term*" ϵ in subtracting value functions to determine the optimal arm is set to $\epsilon = 10^{-7}.(S+1)$, (ii) elements in the value function matrices are rounded by a double-precision floating-point format (float64) whilst the bound of the error term is set to $\epsilon = 10^{-16}.(S+1)$. Note that considering $S = 0$ leads to our standard design where the after-trial population is zero. Furthermore, owing to the limitation on the available memory capacity, which is 16GB, it is impossible to run a trial with the size of 240 whilst the value functions matrices precision is considered float 64.

The difference in the simulation results brings us to one noteworthy distinction between less biased estimations for the case (i) and those with more bias for the case (ii). From a general point of view in the DP procedure, in a given time epoch, when the differences in the value functions between two arms become smaller than the error term, ϵ, the DP recognises that allocating either arm has the same reward, i.e. both arms are optimal. Consequently, the allocation procedure will be the same as EFR, i.e. 50 : 50 randomisation profile, instead of allocating an arm in a deterministic manner. Therefore, setting up a bigger level of precision, i.e. the error term ϵ, whilst the trial size and after-trial population are relatively large gives rise to observing more 50 : 50 randomisation and, therefore, having estimation results similar to those obtained from EFR designs. This explains the reason for observing appreciable differences between the left-hand column (float32) and right-hand side (float 64) results illustrated in figure 3.3, as the after-trial population is considered at one billion. However, the differences are not huge in figure 3.2, where the after-trial population is one million. Figure 3.1, on the other hand, shows there is no difference in estimation results between cases (i) and (ii) since the after-trial population is one hundred. Note that the multiplication of one hundred by any of the assumed trial sizes results in at most six-digit numbers for value functions. In turn, either case provides a decent accuracy in distinguishing the optimal arm in the DP procedure.

In figures 3.1 and 3.2, using different trial sizes does not significantly reduce the estimation bias when like-for-like plots in right-hand-side, and left-hand-side columns are compared. However, in figure 3.1 larger trial sizes result in higher covariance values, see the numerator of equation (3.7). This is not the case in figures 3.2 and 3.3 since the after-trial populations are quite large, which in turn, leads to the covariance between sample size and the estimator itself in equation (3.7) being more or less the same for different trial sizes. Moreover, focusing on the right-hand-side columns and comparing like-for-like plots, the estimation bias is reduced by ≈ 0.1 in figure 3.2 in comparison with 3.1, overall. Although the estimation bias remains the same for some scenarios in 3.3 in comparison with 3.1, for the majority of then the overall estimation bias is reduced by ≈ 0.05 in 3.3. Therefore, the overall reduced estimation bias is ≈ 0.15 when after-trial population increases from 1000 to 1 billion, please compare like-for-like plots in 3.1 and 3.3.

3.3.2 Subject Benefit

Tables 3.1, 3.2, and 3.3 summarise the subject benefit results obtained from the after-trial designs discussed above. Each cell is composed of the average aggregated success responses on both interventions plus/minus the corresponding standard deviation at the end of the trial. As we discussed in section 2.2 regarding the tension between the second and the last objectives, i.e. the conflict between maximising the subject benefit and minimising the estimation bias, numerical results presented in the tables 3.1, 3.2, and 3.3 below show that mitigating the estimation bias leads to losing subject benefit significantly. For instance, scenario $(0.2, 0.8)$ in two extreme trial sizes: $T = 60$ and $T = 240$, tends to show subject benefit of 46.25 ± 3.51 and not available (NA), respectively in the design with after-trial population of 1000 in table 3.1 right-hand-side column (float64 $[\epsilon = 10^{-16}]$), whilst it reduces to 44.15 ± 3.9 and NA, respectively in the design with after-trial population of 1 million in table 3.1 right-hand-side column (float64 $[\epsilon = 10^{-16}]$). Note that we encounter an NA error (for $T = 240$, (float64 $[\epsilon = 10^{-16}]$), and all assumed after-trial sizes) because the computing machine becomes “out of memory" due to the memory space needed for recording the arithmetic calculations in comparison value functions. The reduction rates of the subject benefit results presented in table 3.3 are interesting. The scenario $(0.2, 0.8)$ in two extreme trial sizes, $T = 60$ and $T = 240$, shows a subject benefit of 29.99 ± 3.85 and 119.99 ± 7.74, respectively in the left-hand-side column (float32 $[\epsilon = 10^{-7}]$), whilst it increases to 43.34 ± 4.20 and NA, respectively in the right-hand-side column (float64 $[\epsilon = 10^{-16}]$). As discussed in section 3.3.1, the design with an after-trial population of 1 billion returns almost unbiased estimation results when value functions are rounded by float32, whilst it is associated with some small bias for float64 cases. In accordance with these circumstances, subject benefit results in the left-hand-side column (float32) are considerably smaller than the right-hand-side column (float64) counterparts. Thus, the conflict between a RAR procedure’s second and last objectives can be simply recognised in these situations.

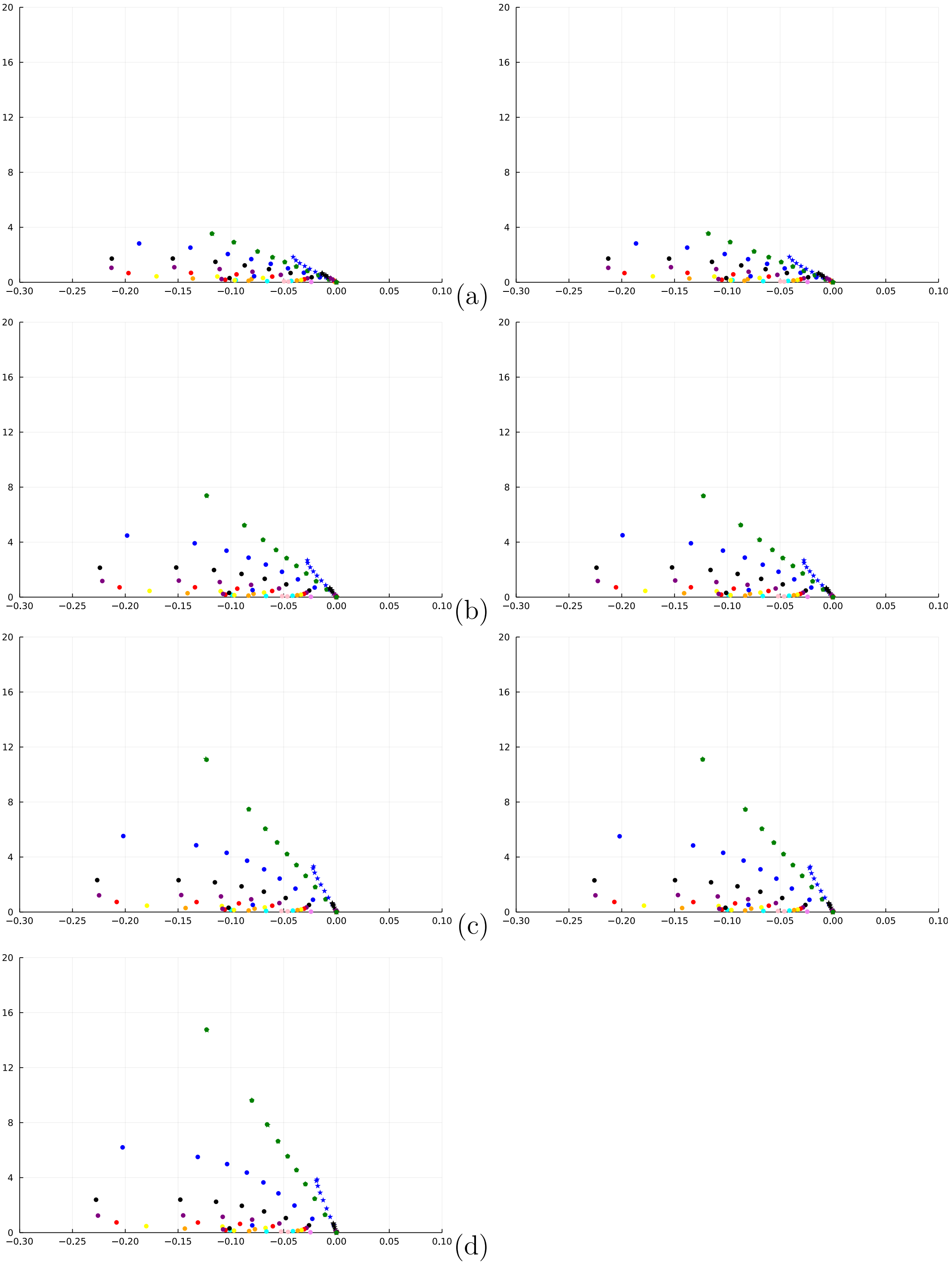


Figure 3.1: After-trial population: 1000. The precision of value functions in the backward induction algorithm is set to be up to about 7 digits in the left-hand-side column whilst it goes up to about 16 digits in the right-hand-side counterparts. (a) $T = 60$ (b) $T = 120$ (c) $T = 180$ (d) $T = 240$. x-axis: Bias of MLE, y-axis: Covariance (MLE, Sample Size).

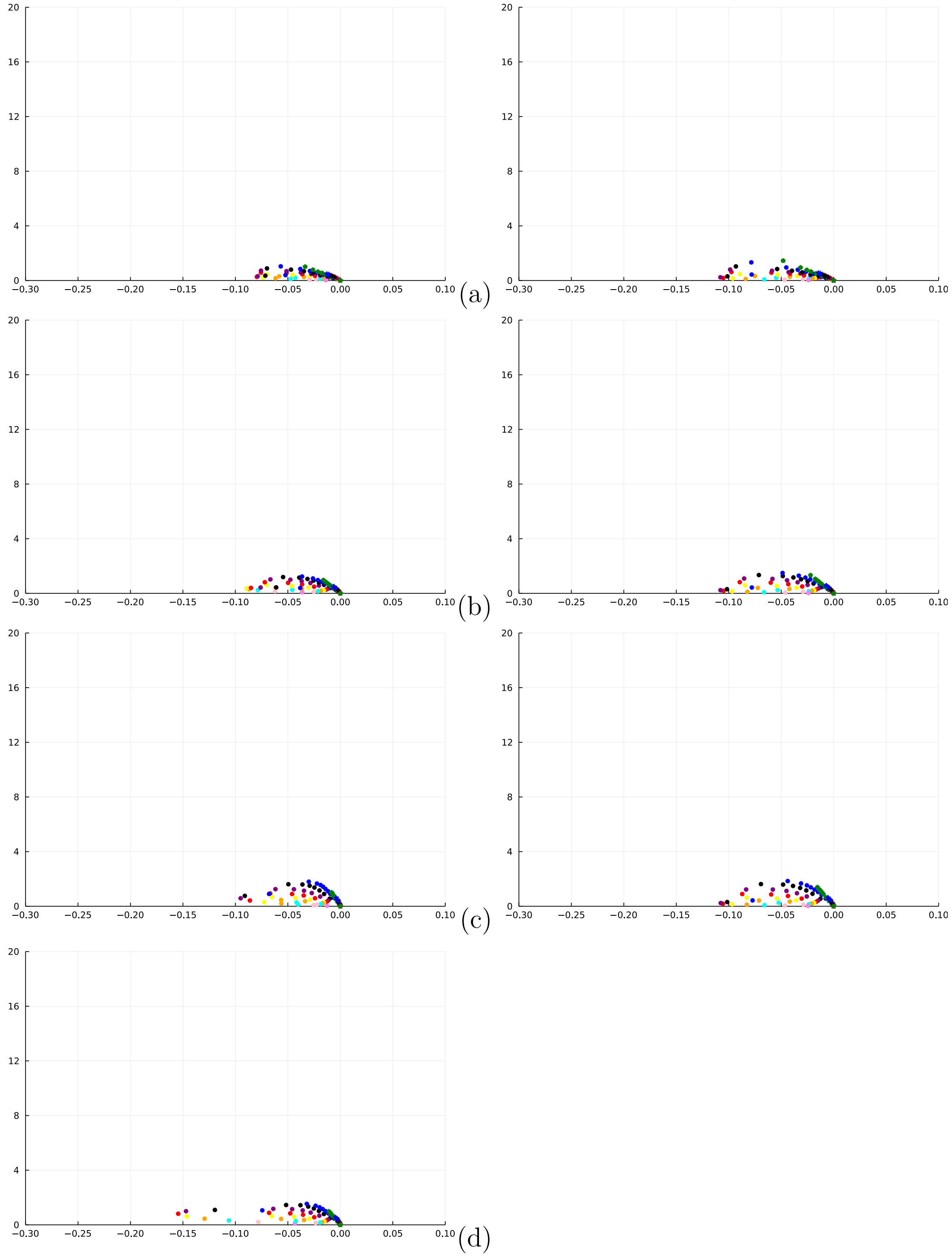


Figure 3.2: After-trial population: 1 million. The precision of value functions in the backward induction algorithm is set to be up to about 7 digits in the left-hand-side column whilst it goes up to about 16 digits in the right-hand-side counterparts. (a) $T = 60$ (b) $T = 120$ (c) $T = 180$ (d) $T = 240$. x-axis: Bias of MLE, y-axis: Covariance (MLE, Sample Size).

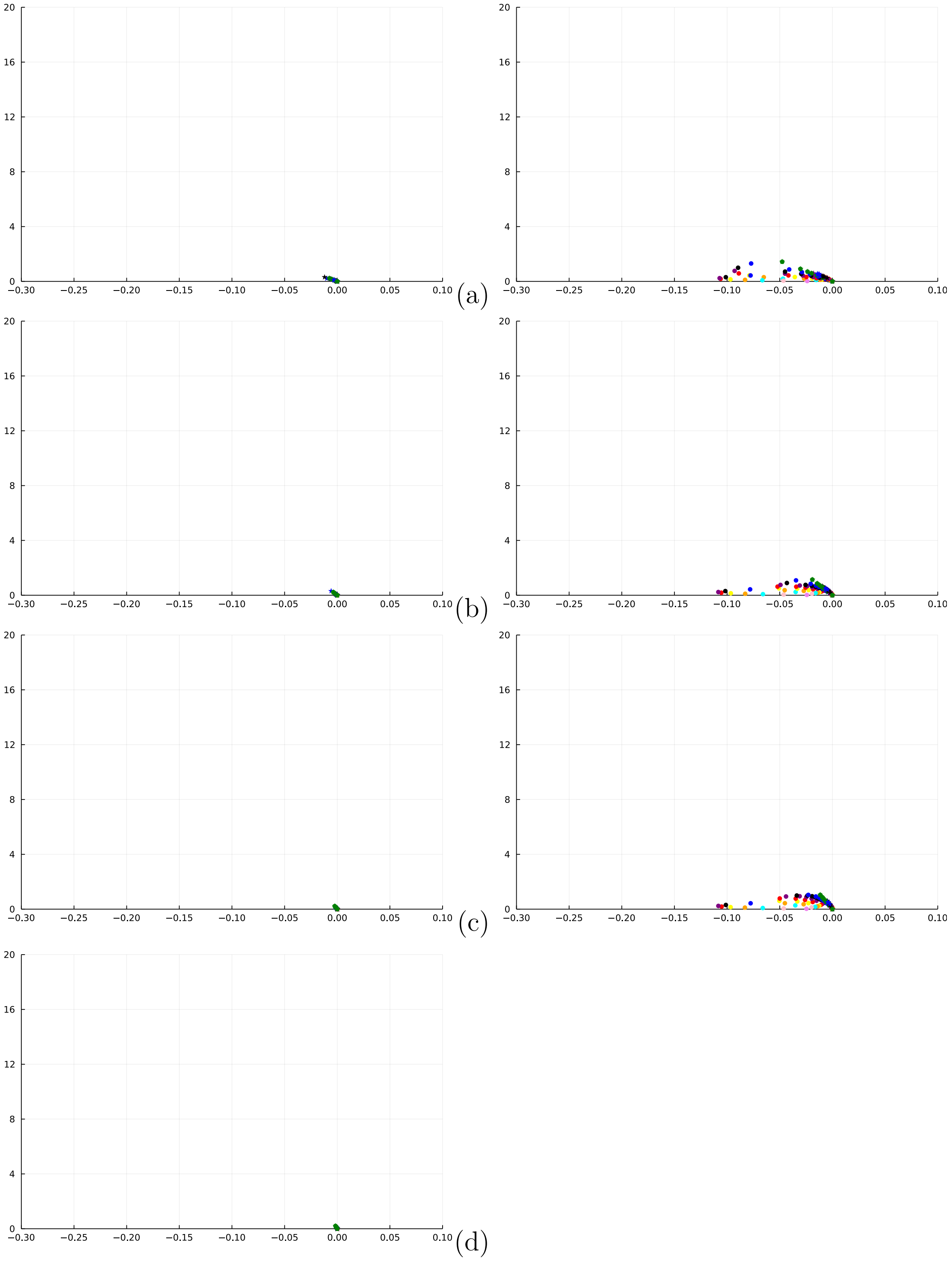


Figure 3.3: After-trial population: 1 billion. The precision of value functions in the backward induction algorithm is set to be up to about 7 digits in the left-hand-side column whilst it goes up to about 16 digits in the right-hand-side counterparts. (a) $T = 60$ (b) $T = 120$ (c) $T = 180$ (d) $T = 240$. x-axis: Bias of MLE, y-axis: Covariance (MLE, Sample Size).

	float32 [$\epsilon = 10^{-7}$]				**float64** [$\epsilon = 10^{-16}$]			
(θ_C, θ_D)	**T=60**	**T=120**	**T=180**	**T=240**	**T=60**	**T=120**	**T=180**	**T=240**
(0 , 0)	0 ± 0	0 ± 0	0 ± 0	0 ± 0	0 ± 0	0 ± 0	0 ± 0	NA
(0 , 0.1)	3.63 ± 2.41	9.05 ± 3.64	14.91 ± 4.33	20.82 ± 4.92	3.64 ± 2.41	9.05 ± 3.65	14.92 ± 4.32	NA
(0 , 0.2)	9.05 ± 3.52	20.8 ± 4.71	32.67 ± 5.63	44.61 ± 6.43	9.04 ± 3.52	20.79 ± 4.7	32.69 ± 5.63	NA
(0 , 0.3)	15.02 ± 3.97	32.83 ± 5.32	50.72 ± 6.4	68.66 ± 7.31	15.02 ± 3.97	32.82 ± 5.32	50.73 ± 6.39	NA
(0 , 0.4)	21.14 ± 4.18	44.98 ± 5.65	68.91 ± 6.82	92.86 ± 7.79	21.15 ± 4.19	44.97 ± 5.65	68.9 ± 6.81	NA
(0 , 0.5)	27.35 ± 4.25	57.22 ± 5.74	87.14 ± 6.93	117.12 ± 7.95	27.35 ± 4.24	57.21 ± 5.75	87.15 ± 6.94	NA
(0 , 0.6)	33.63 ± 4.15	69.53 ± 5.63	105.48 ± 6.78	141.43 ± 7.77	33.63 ± 4.15	69.53 ± 5.63	105.48 ± 6.79	NA
(0 , 0.7)	39.98 ± 3.89	81.9 ± 5.27	123.85 ± 6.34	165.82 ± 7.25	39.98 ± 3.89	81.9 ± 5.27	123.85 ± 6.34	NA
(0 , 0.8)	46.39 ± 3.41	94.33 ± 4.59	142.29 ± 5.51	190.27 ± 6.3	46.39 ± 3.42	94.32 ± 4.59	142.29 ± 5.51	NA
(0 , 0.9)	52.9 ± 2.53	106.91 ± 3.44	160.93 ± 4.15	214.91 ± 4.75	52.89 ± 2.53	106.92 ± 3.44	160.92 ± 4.15	NA
(0 , 1)	59 ± 0	119 ± 0	179 ± 0	239 ± 0	59 ± 0	119 ± 0	179 ± 0	NA
(0.1 , 0.1)	5.99 ± 2.32	12 ± 3.29	18 ± 4.02	24 ± 4.65	6 ± 2.32	12 ± 3.29	18 ± 4.03	NA
(0.1 , 0.2)	9.73 ± 3.14	20.61 ± 4.74	31.99 ± 5.98	43.62 ± 6.99	9.74 ± 3.14	20.6 ± 4.74	31.99 ± 5.99	NA
(0.1 , 0.3)	14.89 ± 3.99	32.27 ± 5.63	50.07 ± 6.74	67.96 ± 7.64	14.9 ± 3.98	32.29 ± 5.64	50.07 ± 6.74	NA
(0.1 , 0.4)	20.87 ± 4.36	44.57 ± 5.87	68.44 ± 7	92.38 ± 7.96	20.87 ± 4.36	44.56 ± 5.86	68.45 ± 6.99	NA
(0.1 , 0.5)	27.12 ± 4.4	56.93 ± 5.88	86.85 ± 7.04	116.79 ± 8.04	27.13 ± 4.4	56.93 ± 5.88	86.84 ± 7.05	NA
(0.1 , 0.6)	33.47 ± 4.25	69.33 ± 5.71	105.26 ± 6.86	141.2 ± 7.84	33.47 ± 4.25	69.31 ± 5.71	105.27 ± 6.86	NA
(0.1 , 0.7)	39.88 ± 3.95	81.77 ± 5.32	123.71 ± 6.39	165.67 ± 7.29	39.87 ± 3.95	81.76 ± 5.32	123.71 ± 6.38	NA
(0.1 , 0.8)	46.33 ± 3.46	94.26 ± 4.63	142.21 ± 5.55	190.19 ± 6.34	46.33 ± 3.46	94.26 ± 4.63	142.21 ± 5.55	NA
(0.1 , 0.9)	52.87 ± 2.55	106.88 ± 3.46	160.88 ± 4.17	214.87 ± 4.78	52.87 ± 2.55	106.89 ± 3.46	160.89 ± 4.17	NA
(0.1 , 1)	59.05 ± 0.22	119.05 ± 0.22	179.05 ± 0.22	239.05 ± 0.22	59.05 ± 0.22	119.05 ± 0.22	179.05 ± 0.22	NA
(0.2 , 0.2)	12 ± 3.1	24 ± 4.38	36.01 ± 5.37	47.99 ± 6.2	12 ± 3.1	24 ± 4.38	36 ± 5.37	NA
(0.2 , 0.3)	15.76 ± 3.64	32.46 ± 5.45	49.62 ± 6.91	67.05 ± 8.17	15.76 ± 3.64	32.46 ± 5.44	49.62 ± 6.9	NA
(0.2 , 0.4)	20.91 ± 4.28	44.08 ± 6.15	67.76 ± 7.43	91.58 ± 8.44	20.91 ± 4.28	44.08 ± 6.15	67.77 ± 7.44	NA
(0.2 , 0.5)	26.94 ± 4.51	56.52 ± 6.11	86.4 ± 7.27	116.3 ± 8.26	26.93 ± 4.51	56.53 ± 6.12	86.39 ± 7.28	NA
(0.2 , 0.6)	33.29 ± 4.37	69.06 ± 5.85	104.97 ± 6.98	140.9 ± 7.94	33.28 ± 4.38	69.06 ± 5.85	104.96 ± 6.98	NA
(0.2 , 0.7)	39.73 ± 4.04	81.59 ± 5.4	123.52 ± 6.46	165.48 ± 7.37	39.73 ± 4.04	81.6 ± 5.4	123.53 ± 6.46	NA
(0.2 , 0.8)	46.25 ± 3.52	94.16 ± 4.68	142.12 ± 5.6	190.08 ± 6.38	46.25 ± 3.51	94.16 ± 4.68	142.12 ± 5.6	NA
(0.2 , 0.9)	52.83 ± 2.58	106.84 ± 3.49	160.84 ± 4.19	214.83 ± 4.8	52.84 ± 2.58	106.84 ± 3.49	160.84 ± 4.19	NA
(0.2 , 1)	59.1 ± 0.3	119.1 ± 0.3	179.1 ± 0.3	239.1 ± 0.3	59.1 ± 0.3	119.1 ± 0.3	179.1 ± 0.3	NA
(0.3 , 0.3)	18 ± 3.55	36 ± 5.02	54 ± 6.15	72 ± 7.1	18 ± 3.55	36.01 ± 5.02	54 ± 6.15	NA
(0.3 , 0.4)	21.8 ± 3.95	44.42 ± 5.86	67.48 ± 7.45	90.79 ± 8.85	21.8 ± 3.95	44.42 ± 5.86	67.46 ± 7.47	NA
(0.3 , 0.5)	27.01 ± 4.41	56.09 ± 6.38	85.7 ± 7.76	115.51 ± 8.86	27.03 ± 4.4	56.09 ± 6.39	85.7 ± 7.77	NA
(0.3 , 0.6)	33.11 ± 4.47	68.68 ± 6.09	104.54 ± 7.25	140.44 ± 8.23	33.11 ± 4.46	68.69 ± 6.09	104.53 ± 7.24	NA
(0.3 , 0.7)	39.56 ± 4.16	81.36 ± 5.54	123.28 ± 6.59	165.22 ± 7.48	39.57 ± 4.16	81.37 ± 5.55	123.29 ± 6.59	NA
(0.3 , 0.8)	46.15 ± 3.6	94.04 ± 4.75	141.99 ± 5.67	189.96 ± 6.46	46.14 ± 3.59	94.04 ± 4.76	141.98 ± 5.67	NA
(0.3 , 0.9)	52.79 ± 2.63	106.8 ± 3.52	160.77 ± 4.23	214.78 ± 4.83	52.79 ± 2.63	106.79 ± 3.52	160.78 ± 4.23	NA
(0.3 , 1)	59.15 ± 0.36	119.15 ± 0.36	179.15 ± 0.36	239.15 ± 0.36	59.15 ± 0.36	119.15 ± 0.36	179.15 ± 0.36	NA
(0.4 , 0.4)	24 ± 3.79	48 ± 5.37	72 ± 6.58	96.02 ± 7.58	24 ± 3.79	47.99 ± 5.37	72 ± 6.58	NA
(0.4 , 0.5)	27.85 ± 4.08	56.45 ± 6.06	85.47 ± 7.71	114.73 ± 9.16	27.84 ± 4.08	56.46 ± 6.07	85.47 ± 7.71	NA
(0.4 , 0.6)	33.17 ± 4.39	68.25 ± 6.36	103.84 ± 7.77	139.65 ± 8.88	33.17 ± 4.39	68.25 ± 6.36	103.85 ± 7.77	NA
(0.4 , 0.7)	39.4 ± 4.27	81.02 ± 5.79	122.87 ± 6.88	164.8 ± 7.79	39.4 ± 4.26	81.02 ± 5.79	122.89 ± 6.88	NA
(0.4 , 0.8)	46.01 ± 3.71	93.86 ± 4.88	141.8 ± 5.78	189.77 ± 6.56	46.01 ± 3.71	93.85 ± 4.88	141.8 ± 5.78	NA
(0.4 , 0.9)	52.72 ± 2.69	106.72 ± 3.58	160.7 ± 4.27	214.7 ± 4.87	52.73 ± 2.69	106.72 ± 3.57	160.71 ± 4.27	NA
(0.4 , 1)	59.2 ± 0.4	119.2 ± 0.4	179.2 ± 0.4	239.2 ± 0.4	59.2 ± 0.4	119.2 ± 0.4	179.2 ± 0.4	NA
(0.5 , 0.5)	30.01 ± 3.87	60 ± 5.48	90 ± 6.72	120 ± 7.74	30 ± 3.87	59.99 ± 5.47	90 ± 6.71	NA
(0.5 , 0.6)	33.91 ± 4.07	68.55 ± 6.04	103.59 ± 7.69	138.85 ± 9.15	33.9 ± 4.07	68.56 ± 6.05	103.59 ± 7.7	NA
(0.5 , 0.7)	39.41 ± 4.22	80.59 ± 6.08	122.25 ± 7.43	164.05 ± 8.49	39.41 ± 4.21	80.59 ± 6.09	122.24 ± 7.43	NA
(0.5 , 0.8)	45.86 ± 3.83	93.58 ± 5.12	141.48 ± 6.07	189.42 ± 6.88	45.86 ± 3.82	93.58 ± 5.13	141.48 ± 6.05	NA
(0.5 , 0.9)	52.64 ± 2.78	106.61 ± 3.67	160.59 ± 4.36	214.58 ± 4.96	52.64 ± 2.79	106.6 ± 3.67	160.59 ± 4.36	NA
(0.5 , 1)	59.25 ± 0.43	119.25 ± 0.43	179.25 ± 0.43	239.25 ± 0.43	59.25 ± 0.43	119.25 ± 0.43	179.25 ± 0.43	NA
(0.6 , 0.6)	36 ± 3.79	72 ± 5.36	108 ± 6.57	144.01 ± 7.58	36 ± 3.8	71.99 ± 5.36	108 ± 6.57	NA
(0.6 , 0.7)	40 ± 3.91	80.74 ± 5.81	121.85 ± 7.42	163.19 ± 8.81	40.01 ± 3.9	80.74 ± 5.81	121.86 ± 7.41	NA
(0.6 , 0.8)	45.78 ± 3.84	93.17 ± 5.47	140.91 ± 6.64	188.75 ± 7.6	45.78 ± 3.84	93.17 ± 5.48	140.91 ± 6.64	NA
(0.6 , 0.9)	52.52 ± 2.92	106.42 ± 3.88	160.39 ± 4.59	214.37 ± 5.2	52.52 ± 2.92	106.42 ± 3.88	160.38 ± 4.6	NA
(0.6 , 1)	59.3 ± 0.46	119.3 ± 0.46	179.3 ± 0.46	239.3 ± 0.46	59.3 ± 0.46	119.3 ± 0.46	179.3 ± 0.46	NA
(0.7 , 0.7)	42 ± 3.55	84 ± 5.02	126 ± 6.15	168 ± 7.1	42 ± 3.56	83.99 ± 5.02	126.01 ± 6.15	NA
(0.7 , 0.8)	46.17 ± 3.55	93.09 ± 5.31	140.34 ± 6.77	187.81 ± 8.05	46.18 ± 3.55	93.09 ± 5.31	140.35 ± 6.77	NA
(0.7 , 0.9)	52.38 ± 3.04	106.09 ± 4.26	159.95 ± 5.16	213.89 ± 5.87	52.38 ± 3.04	106.08 ± 4.26	159.96 ± 5.14	NA
(0.7 , 1)	59.35 ± 0.48	119.35 ± 0.48	179.35 ± 0.48	239.35 ± 0.48	59.35 ± 0.48	119.35 ± 0.48	179.35 ± 0.48	NA
(0.8 , 0.8)	48 ± 3.1	96 ± 4.38	144 ± 5.37	191.99 ± 6.2	48 ± 3.1	96 ± 4.38	144 ± 5.37	NA
(0.8 , 0.9)	52.49 ± 2.88	105.74 ± 4.34	159.26 ± 5.57	212.95 ± 6.63	52.49 ± 2.88	105.74 ± 4.35	159.27 ± 5.56	NA
(0.8 , 1)	59.4 ± 0.49	119.39 ± 0.51	179.39 ± 0.51	239.39 ± 0.51	59.4 ± 0.49	119.39 ± 0.51	179.39 ± 0.51	NA
(0.9 , 0.9)	54 ± 2.32	108 ± 3.28	162 ± 4.02	216 ± 4.65	54 ± 2.32	108 ± 3.28	162 ± 4.03	NA
(0.9 , 1)	59.44 ± 0.54	119.36 ± 0.67	179.34 ± 0.7	239.34 ± 0.72	59.44 ± 0.54	119.36 ± 0.67	179.34 ± 0.7	NA
(1 , 1)	60 ± 0	120 ± 0	180 ± 0	240 ± 0	60 ± 0	120 ± 0	180 ± 0	NA

Table 3.1: The numerical subject benefit results for the design with an after-trial population $S = 1000$, and for different trial sizes $T = 60, 120, 180$, and 240. Each cell is composed of the average number of success responses (first component) added to/subtracted from the corresponding standard deviation (second component) for each scenario (θ_C, θ_D).

	float32 [$\epsilon = 10^{-7}$]				**float64** [$\epsilon = 10^{-16}$]			
(θ_C, θ_D)	**T=60**	**T=120**	**T=180**	**T=240**	**T=60**	**T=120**	**T=180**	**T=240**
(0 , 0)	0.00 ± 0.00	0.00 ± 0.00	0.00 ± 0.00	0.00 ± 0.00	0.00 ± 0.00	0.00 ± 0.00	0.00 ± 0.00	NA
(0 , 0.1)	2.37 ± 1.98	6.13 ± 3.60	10.66 ± 4.74	15.77 ± 5.61	2.42 ± 1.96	6.26 ± 3.69	11.39 ± 4.85	NA
(0 , 0.2)	6.42 ± 3.60	16.63 ± 5.09	27.60 ± 6.00	38.95 ± 7.02	6.46 ± 3.62	17.20 ± 5.32	28.84 ± 6.05	NA
(0 , 0.3)	11.80 ± 4.43	28.22 ± 5.76	45.31 ± 6.93	62.95 ± 7.60	11.95 ± 4.57	29.23 ± 5.74	46.85 ± 6.69	NA
(0 , 0.4)	17.77 ± 4.68	40.28 ± 6.11	63.53 ± 7.17	86.77 ± 8.09	18.16 ± 4.82	41.46 ± 5.98	65.12 ± 7.08	NA
(0 , 0.5)	24.00 ± 4.68	52.66 ± 6.20	81.88 ± 7.39	111.28 ± 8.50	24.59 ± 4.71	53.88 ± 6.05	83.58 ± 7.18	NA
(0 , 0.6)	30.39 ± 4.59	65.30 ± 6.13	100.67 ± 7.27	136.49 ± 8.36	31.07 ± 4.48	66.40 ± 5.88	102.15 ± 7.02	NA
(0 , 0.7)	36.95 ± 4.33	78.18 ± 5.71	119.61 ± 6.74	161.34 ± 7.52	37.64 ± 4.20	79.07 ± 5.53	120.87 ± 6.58	NA
(0 , 0.8)	43.73 ± 3.87	91.12 ± 4.97	138.61 ± 5.94	186.27 ± 6.71	44.39 ± 3.74	91.93 ± 4.86	139.76 ± 5.76	NA
(0 , 0.9)	50.86 ± 3.09	104.33 ± 3.85	157.93 ± 4.55	211.71 ± 5.17	51.49 ± 3.01	105.07 ± 3.75	158.93 ± 4.40	NA
(0 , 1)	58.03 ± 0.17	118.00 ± 0.03	178.01 ± 1.00	237.26 ± 0.67	59.00 ± 0.00	119.00 ± 0.00	179.00 ± 0.00	NA
(0.1 , 0.1)	6.00 ± 2.32	12.00 ± 3.29	17.99 ± 4.02	24.00 ± 4.65	6.00 ± 2.32	12.00 ± 3.28	17.99 ± 4.02	NA
(0.1 , 0.2)	8.86 ± 2.95	18.30 ± 4.40	28.27 ± 5.65	38.60 ± 6.76	8.90 ± 2.96	18.43 ± 4.47	28.65 ± 5.80	NA
(0.1 , 0.3)	12.68 ± 3.90	27.88 ± 5.85	44.20 ± 7.21	61.05 ± 8.32	12.79 ± 3.95	28.41 ± 6.01	45.33 ± 7.30	NA
(0.1 , 0.4)	17.88 ± 4.63	39.58 ± 6.40	62.24 ± 7.65	85.36 ± 8.62	18.15 ± 4.73	40.54 ± 6.41	63.82 ± 7.53	NA
(0.1 , 0.5)	23.87 ± 4.82	51.97 ± 6.46	80.93 ± 7.72	110.08 ± 8.74	24.35 ± 4.87	53.14 ± 6.33	82.60 ± 7.48	NA
(0.1 , 0.6)	30.24 ± 4.71	64.70 ± 6.32	99.98 ± 7.50	135.35 ± 8.55	30.82 ± 4.63	65.85 ± 6.08	101.44 ± 7.22	NA
(0.1 , 0.7)	36.84 ± 4.41	77.66 ± 5.86	119.10 ± 6.91	160.65 ± 7.83	37.43 ± 4.31	78.66 ± 5.66	120.37 ± 6.71	NA
(0.1 , 0.8)	43.67 ± 3.94	90.76 ± 5.14	138.30 ± 6.02	185.94 ± 6.84	44.26 ± 3.83	91.66 ± 4.96	139.44 ± 5.86	NA
(0.1 , 0.9)	50.85 ± 3.13	104.19 ± 3.94	157.78 ± 4.62	211.63 ± 5.17	51.45 ± 3.06	104.95 ± 3.81	158.80 ± 4.46	NA
(0.1 , 1)	58.14 ± 0.41	117.92 ± 0.32	177.95 ± 0.98	237.45 ± 0.74	59.05 ± 0.22	119.05 ± 0.22	179.05 ± 0.22	NA
(0.2 , 0.2)	12.00 ± 3.10	24.00 ± 4.38	36.00 ± 5.37	47.99 ± 6.19	12.00 ± 3.10	24.00 ± 4.38	36.00 ± 5.37	NA
(0.2 , 0.3)	15.10 ± 3.52	30.55 ± 5.09	46.38 ± 6.39	62.51 ± 7.58	15.15 ± 3.54	30.70 ± 5.13	46.75 ± 6.49	NA
(0.2 , 0.4)	19.09 ± 4.15	39.91 ± 6.17	61.84 ± 7.70	84.36 ± 8.92	19.26 ± 4.21	40.48 ± 6.28	62.96 ± 7.78	NA
(0.2 , 0.5)	24.30 ± 4.62	51.60 ± 6.60	80.05 ± 7.98	108.95 ± 9.07	24.64 ± 4.67	52.56 ± 6.58	81.60 ± 7.82	NA
(0.2 , 0.6)	30.32 ± 4.70	64.23 ± 6.46	99.13 ± 7.73	134.34 ± 8.76	30.80 ± 4.66	65.27 ± 6.29	100.63 ± 7.46	NA
(0.2 , 0.7)	36.79 ± 4.46	77.19 ± 5.99	118.45 ± 7.13	159.88 ± 8.04	37.31 ± 4.36	78.24 ± 5.82	119.81 ± 6.87	NA
(0.2 , 0.8)	43.61 ± 4.01	90.41 ± 5.27	137.90 ± 6.14	185.47 ± 6.99	44.15 ± 3.90	91.38 ± 5.07	139.10 ± 5.97	NA
(0.2 , 0.9)	50.84 ± 3.17	104.01 ± 4.02	157.58 ± 4.69	211.42 ± 5.25	51.42 ± 3.12	104.82 ± 3.88	158.63 ± 4.52	NA
(0.2 , 1)	58.24 ± 0.53	117.88 ± 0.46	177.92 ± 0.95	237.62 ± 0.79	59.10 ± 0.30	119.10 ± 0.30	179.10 ± 0.30	NA
(0.3 , 0.3)	18.00 ± 3.55	36.00 ± 5.02	53.99 ± 6.15	72.00 ± 7.10	18.00 ± 3.55	36.00 ± 5.01	53.99 ± 6.15	NA
(0.3 , 0.4)	21.26 ± 3.83	42.77 ± 5.51	64.61 ± 6.86	86.66 ± 8.07	21.32 ± 3.85	42.96 ± 5.57	65.00 ± 6.96	NA
(0.3 , 0.5)	25.46 ± 4.26	52.27 ± 6.28	80.05 ± 7.86	108.43 ± 9.14	25.67 ± 4.30	52.88 ± 6.35	81.17 ± 7.88	NA
(0.3 , 0.6)	30.79 ± 4.51	64.02 ± 6.46	98.43 ± 7.88	133.33 ± 8.99	31.14 ± 4.51	64.93 ± 6.37	99.87 ± 7.70	NA
(0.3 , 0.7)	36.92 ± 4.42	76.79 ± 6.13	117.75 ± 7.36	159.00 ± 8.26	37.36 ± 4.35	77.81 ± 5.97	119.15 ± 7.10	NA
(0.3 , 0.8)	43.61 ± 4.05	90.10 ± 5.38	137.46 ± 6.31	184.91 ± 7.18	44.10 ± 3.95	91.06 ± 5.19	138.68 ± 6.11	NA
(0.3 , 0.9)	50.85 ± 3.20	103.81 ± 4.12	157.36 ± 4.78	211.14 ± 5.35	51.40 ± 3.17	104.67 ± 3.95	158.44 ± 4.59	NA
(0.3 , 1)	58.33 ± 0.61	117.88 ± 0.57	177.92 ± 0.91	237.77 ± 0.83	59.15 ± 0.36	119.15 ± 0.36	179.15 ± 0.36	NA
(0.4 , 0.4)	24.00 ± 3.79	48.00 ± 5.36	72.00 ± 6.57	96.01 ± 7.59	24.00 ± 3.79	47.99 ± 5.36	71.99 ± 6.57	NA
(0.4 , 0.5)	27.39 ± 3.97	55.00 ± 5.69	82.89 ± 7.06	110.99 ± 8.29	27.46 ± 3.99	55.25 ± 5.74	83.35 ± 7.14	NA
(0.4 , 0.6)	31.81 ± 4.22	64.72 ± 6.13	98.54 ± 7.68	132.91 ± 8.99	32.04 ± 4.24	65.37 ± 6.14	99.62 ± 7.64	NA
(0.4 , 0.7)	37.32 ± 4.27	76.66 ± 6.14	117.19 ± 7.49	158.16 ± 8.47	37.70 ± 4.24	77.55 ± 6.03	118.51 ± 7.29	NA
(0.4 , 0.8)	43.73 ± 4.02	89.81 ± 5.48	136.97 ± 6.49	184.26 ± 7.37	44.16 ± 3.95	90.77 ± 5.32	138.20 ± 6.28	NA
(0.4 , 0.9)	50.89 ± 3.23	103.62 ± 4.21	157.12 ± 4.88	210.78 ± 5.48	51.41 ± 3.20	104.50 ± 4.03	158.21 ± 4.68	NA
(0.4 , 1)	58.43 ± 0.67	117.92 ± 0.66	177.92 ± 0.86	237.89 ± 0.87	59.20 ± 0.40	119.20 ± 0.40	179.20 ± 0.40	NA
(0.5 , 0.5)	30.00 ± 3.87	60.01 ± 5.48	90.00 ± 6.70	120.01 ± 7.75	30.00 ± 3.87	60.00 ± 5.47	90.00 ± 6.71	NA
(0.5 , 0.6)	33.50 ± 3.95	67.21 ± 5.62	101.17 ± 6.97	135.33 ± 8.17	33.58 ± 3.96	67.49 ± 5.65	101.66 ± 7.01	NA
(0.5 , 0.7)	38.16 ± 4.03	77.25 ± 5.87	117.27 ± 7.35	157.76 ± 8.52	38.41 ± 4.03	77.91 ± 5.84	118.29 ± 7.25	NA
(0.5 , 0.8)	44.02 ± 3.93	89.70 ± 5.51	136.50 ± 6.65	183.58 ± 7.57	44.39 ± 3.87	90.55 ± 5.39	137.68 ± 6.46	NA
(0.5 , 0.9)	50.97 ± 3.22	103.46 ± 4.29	156.83 ± 5.01	210.38 ± 5.63	51.45 ± 3.20	104.33 ± 4.14	157.92 ± 4.81	NA
(0.5 , 1)	58.53 ± 0.71	118.00 ± 0.74	177.90 ± 0.82	237.99 ± 0.89	59.25 ± 0.43	119.25 ± 0.43	179.25 ± 0.43	NA
(0.6 , 0.6)	36.00 ± 3.80	72.00 ± 5.36	108.00 ± 6.57	144.00 ± 7.59	36.00 ± 3.79	72.00 ± 5.37	108.00 ± 6.58	NA
(0.6 , 0.7)	39.62 ± 3.77	79.44 ± 5.39	119.55 ± 6.67	159.77 ± 7.80	39.70 ± 3.78	79.73 ± 5.41	120.01 ± 6.70	NA
(0.6 , 0.8)	44.63 ± 3.73	90.12 ± 5.31	136.43 ± 6.59	183.15 ± 7.65	44.90 ± 3.72	90.77 ± 5.27	137.42 ± 6.48	NA
(0.6 , 0.9)	51.14 ± 3.18	103.39 ± 4.35	156.51 ± 5.15	209.93 ± 5.79	51.56 ± 3.16	104.19 ± 4.22	157.58 ± 4.95	NA
(0.6 , 1)	58.66 ± 0.75	118.13 ± 0.81	177.90 ± 0.79	238.07 ± 0.90	59.30 ± 0.46	119.30 ± 0.46	179.30 ± 0.46	NA
(0.7 , 0.7)	42.00 ± 3.55	84.00 ± 5.02	126.00 ± 6.15	168.02 ± 7.10	42.00 ± 3.55	84.01 ± 5.02	126.01 ± 6.15	NA
(0.7 , 0.8)	45.80 ± 3.43	91.84 ± 4.88	138.11 ± 6.03	184.57 ± 7.05	45.90 ± 3.44	92.12 ± 4.90	138.60 ± 6.05	NA
(0.7 , 0.9)	51.47 ± 3.05	103.61 ± 4.28	156.36 ± 5.19	209.52 ± 5.93	51.80 ± 3.05	104.26 ± 4.21	157.31 ± 5.06	NA
(0.7 , 1)	58.82 ± 0.76	118.32 ± 0.86	177.97 ± 0.78	238.15 ± 0.89	59.35 ± 0.48	119.35 ± 0.48	179.35 ± 0.48	NA
(0.8 , 0.8)	48.00 ± 3.10	96.00 ± 4.38	144.00 ± 5.36	192.00 ± 6.20	48.00 ± 3.10	96.00 ± 4.38	144.00 ± 5.36	NA
(0.8 , 0.9)	52.18 ± 2.78	104.61 ± 3.95	157.21 ± 4.85	210.00 ± 5.65	52.31 ± 2.80	104.94 ± 3.98	157.80 ± 4.85	NA
(0.8 , 1)	59.01 ± 0.74	118.57 ± 0.89	178.18 ± 0.78	238.33 ± 0.85	59.40 ± 0.49	119.40 ± 0.49	179.40 ± 0.49	NA
(0.9 , 0.9)	54.00 ± 2.33	108.00 ± 3.29	162.00 ± 4.03	215.99 ± 4.65	54.01 ± 2.32	108.00 ± 3.29	162.00 ± 4.03	NA
(0.9 , 1)	59.25 ± 0.66	118.96 ± 0.84	178.58 ± 0.71	238.67 ± 0.79	59.44 ± 0.52	119.45 ± 0.50	179.45 ± 0.50	NA
(1 , 1)	60.00 ± 0.00	120.00 ± 0.00	180.00 ± 0.00	240.00 ± 0.00	60.00 ± 0.00	120.00 ± 0.00	180.00 ± 0.00	NA

Table 3.2: The numerical subject benefit results for the design with an after-trial population $S = 1$ million, and for different trial sizes $T = 60, 120, 180,$ and 240. Each cell is composed of the average number of success responses (first component) added to/subtracted from the corresponding standard deviation (second component) for each scenario (θ_C, θ_D).

	float32 [$\epsilon = 10^{-7}$]				**float64** [$\epsilon = 10^{-16}$]			
(θ_C, θ_D)	**T=60**	**T=120**	**T=180**	**T=240**	**T=60**	**T=120**	**T=180**	**T=240**
(0 , 0)	0.00 ± 0.00	0.00 ± 0.00	0.00 ± 0.00	0.00 ± 0.00	0.00 ± 0.00	0.00 ± 0.00	0.00 ± 0.00	NA
(0 , 0.1)	2.14 ± 1.72	5.27 ± 2.87	8.69 ± 3.33	11.94 ± 3.49	2.00 ± 1.66	4.61 ± 3.25	8.51 ± 4.82	NA
(0 , 0.2)	5.18 ± 2.79	11.89 ± 3.48	17.99 ± 4.04	24.00 ± 4.65	5.13 ± 3.31	13.92 ± 5.82	25.05 ± 6.89	NA
(0 , 0.3)	8.62 ± 3.15	17.99 ± 3.92	27.00 ± 4.79	36.00 ± 5.53	9.92 ± 4.73	25.88 ± 6.50	43.19 ± 7.13	NA
(0 , 0.4)	11.90 ± 3.23	24.00 ± 4.37	36.00 ± 5.36	47.99 ± 6.19	15.96 ± 5.36	38.32 ± 6.43	61.53 ± 7.37	NA
(0 , 0.5)	14.97 ± 3.36	30.00 ± 4.73	45.00 ± 5.80	60.00 ± 6.69	22.54 ± 5.23	50.90 ± 6.34	80.14 ± 7.42	NA
(0 , 0.6)	17.99 ± 3.52	35.99 ± 5.00	54.00 ± 6.13	72.00 ± 7.08	29.21 ± 4.84	63.61 ± 6.14	98.91 ± 7.24	NA
(0 , 0.7)	21.00 ± 3.66	41.99 ± 5.20	63.01 ± 6.38	83.99 ± 7.37	36.02 ± 4.52	76.49 ± 5.76	117.92 ± 6.80	NA
(0 , 0.8)	24.00 ± 3.75	48.00 ± 5.34	71.99 ± 6.55	96.00 ± 7.57	43.18 ± 4.17	89.70 ± 5.10	137.22 ± 5.98	NA
(0 , 0.9)	27.00 ± 3.80	54.00 ± 5.41	81.01 ± 6.64	108.00 ± 7.69	51.06 ± 3.41	103.48 ± 4.09	157.01 ± 4.64	NA
(0 , 1)	30.01 ± 3.81	60.00 ± 5.43	89.99 ± 6.67	120.00 ± 7.70	59.00 ± 0.00	119.00 ± 0.00	179.00 ± 0.00	NA
(0.1 , 0.1)	6.00 ± 2.33	12.00 ± 3.28	18.00 ± 4.03	24.01 ± 4.66	6.00 ± 2.32	12.00 ± 3.29	18.00 ± 4.03	NA
(0.1 , 0.2)	8.61 ± 2.82	17.49 ± 4.00	26.57 ± 4.90	35.68 ± 5.63	8.69 ± 2.93	17.44 ± 4.27	26.74 ± 5.50	NA
(0.1 , 0.3)	11.43 ± 3.31	23.72 ± 4.52	35.93 ± 5.41	47.97 ± 6.21	12.08 ± 3.90	25.78 ± 6.04	41.50 ± 7.67	NA
(0.1 , 0.4)	14.68 ± 3.57	29.96 ± 4.77	45.00 ± 5.80	60.01 ± 6.70	16.95 ± 4.76	37.37 ± 6.87	59.81 ± 8.01	NA
(0.1 , 0.5)	17.91 ± 3.62	36.00 ± 5.01	54.00 ± 6.14	71.99 ± 7.10	22.90 ± 5.04	50.06 ± 6.74	78.73 ± 7.85	NA
(0.1 , 0.6)	20.99 ± 3.69	42.01 ± 5.21	63.00 ± 6.40	84.00 ± 7.38	29.34 ± 4.86	62.89 ± 6.38	97.82 ± 7.53	NA
(0.1 , 0.7)	24.00 ± 3.77	48.01 ± 5.35	72.01 ± 6.56	95.99 ± 7.58	36.05 ± 4.56	75.94 ± 5.95	117.12 ± 7.02	NA
(0.1 , 0.8)	27.00 ± 3.82	54.00 ± 5.42	80.99 ± 6.65	107.99 ± 7.69	43.22 ± 4.22	89.34 ± 5.25	136.70 ± 6.13	NA
(0.1 , 0.9)	29.99 ± 3.83	60.00 ± 5.45	89.99 ± 6.69	119.99 ± 7.72	51.09 ± 3.41	103.36 ± 4.19	156.78 ± 4.73	NA
(0.1 , 1)	33.00 ± 3.80	66.01 ± 5.42	99.00 ± 6.64	132.01 ± 7.68	59.05 ± 0.22	119.05 ± 0.22	179.05 ± 0.22	NA
(0.2 , 0.2)	12.00 ± 3.10	24.01 ± 4.39	36.00 ± 5.37	48.00 ± 6.20	12.00 ± 3.10	24.00 ± 4.38	36.01 ± 5.36	NA
(0.2 , 0.3)	14.83 ± 3.41	29.72 ± 4.80	44.72 ± 5.85	59.77 ± 6.76	15.05 ± 3.53	30.01 ± 5.06	45.32 ± 6.31	NA
(0.2 , 0.4)	17.76 ± 3.67	35.79 ± 5.10	53.91 ± 6.17	71.97 ± 7.11	18.89 ± 4.15	38.54 ± 6.25	59.67 ± 7.91	NA
(0.2 , 0.5)	20.88 ± 3.76	41.97 ± 5.24	63.00 ± 6.41	84.00 ± 7.38	23.92 ± 4.65	49.97 ± 6.76	77.84 ± 8.14	NA
(0.2 , 0.6)	23.97 ± 3.80	48.00 ± 5.36	72.00 ± 6.57	96.00 ± 7.58	29.85 ± 4.73	62.49 ± 6.49	96.88 ± 7.76	NA
(0.2 , 0.7)	27.00 ± 3.83	54.01 ± 5.44	81.00 ± 6.66	108.00 ± 7.69	36.33 ± 4.51	75.49 ± 6.10	116.33 ± 7.22	NA
(0.2 , 0.8)	29.99 ± 3.85	60.01 ± 5.45	89.99 ± 6.69	119.99 ± 7.74	43.34 ± 4.20	89.02 ± 5.39	136.17 ± 6.29	NA
(0.2 , 0.9)	33.00 ± 3.82	65.99 ± 5.43	99.00 ± 6.66	131.99 ± 7.68	51.13 ± 3.39	103.24 ± 4.29	156.53 ± 4.83	NA
(0.2 , 1)	36.00 ± 3.75	72.01 ± 5.34	108.01 ± 6.55	143.99 ± 7.58	59.10 ± 0.30	119.10 ± 0.30	179.10 ± 0.30	NA
(0.3 , 0.3)	18.00 ± 3.55	35.99 ± 5.02	53.99 ± 6.15	72.00 ± 7.10	18.01 ± 3.55	36.00 ± 5.02	54.00 ± 6.15	NA
(0.3 , 0.4)	20.94 ± 3.72	41.87 ± 5.28	62.86 ± 6.44	83.87 ± 7.41	21.26 ± 3.84	42.53 ± 5.53	63.92 ± 6.87	NA
(0.3 , 0.5)	23.92 ± 3.83	47.93 ± 5.40	71.97 ± 6.60	95.98 ± 7.59	25.43 ± 4.28	51.49 ± 6.24	78.55 ± 7.85	NA
(0.3 , 0.6)	26.98 ± 3.86	53.99 ± 5.45	81.00 ± 6.67	108.00 ± 7.70	30.71 ± 4.52	62.79 ± 6.35	96.40 ± 7.82	NA
(0.3 , 0.7)	30.00 ± 3.86	59.98 ± 5.47	90.00 ± 6.70	119.99 ± 7.74	36.77 ± 4.39	75.31 ± 6.14	115.63 ± 7.40	NA
(0.3 , 0.8)	33.00 ± 3.83	65.99 ± 5.43	98.99 ± 6.66	132.00 ± 7.70	43.54 ± 4.14	88.76 ± 5.50	135.64 ± 6.46	NA
(0.3 , 0.9)	36.00 ± 3.77	72.00 ± 5.35	108.00 ± 6.56	144.00 ± 7.58	51.19 ± 3.37	103.14 ± 4.39	156.26 ± 4.94	NA
(0.3 , 1)	39.00 ± 3.66	77.99 ± 5.20	117.00 ± 6.38	156.00 ± 7.36	59.15 ± 0.36	119.15 ± 0.36	179.15 ± 0.36	NA
(0.4 , 0.4)	24.00 ± 3.80	47.99 ± 5.37	72.01 ± 6.57	96.01 ± 7.59	23.99 ± 3.80	48.00 ± 5.36	72.00 ± 6.57	NA
(0.4 , 0.5)	26.99 ± 3.86	53.97 ± 5.46	80.98 ± 6.68	107.97 ± 7.72	27.42 ± 3.99	54.93 ± 5.69	82.59 ± 7.03	NA
(0.4 , 0.6)	30.00 ± 3.89	59.98 ± 5.48	90.00 ± 6.71	120.00 ± 7.75	31.91 ± 4.23	64.27 ± 5.99	97.38 ± 7.45	NA
(0.4 , 0.7)	33.03 ± 3.86	66.00 ± 5.45	99.00 ± 6.67	132.00 ± 7.71	37.40 ± 4.24	75.67 ± 6.02	115.36 ± 7.43	NA
(0.4 , 0.8)	36.03 ± 3.80	72.00 ± 5.36	108.01 ± 6.57	144.00 ± 7.58	43.83 ± 4.05	88.73 ± 5.54	135.17 ± 6.60	NA
(0.4 , 0.9)	39.01 ± 3.69	78.01 ± 5.21	117.00 ± 6.40	156.01 ± 7.38	51.26 ± 3.32	103.11 ± 4.47	156.00 ± 5.07	NA
(0.4 , 1)	42.00 ± 3.53	84.00 ± 5.00	126.00 ± 6.13	168.00 ± 7.09	59.20 ± 0.40	119.20 ± 0.40	179.20 ± 0.40	NA
(0.5 , 0.5)	30.00 ± 3.87	60.00 ± 5.47	90.00 ± 6.70	120.00 ± 7.75	30.00 ± 3.87	60.00 ± 5.47	90.00 ± 6.71	NA
(0.5 , 0.6)	33.02 ± 3.86	66.03 ± 5.47	99.03 ± 6.69	132.03 ± 7.72	33.56 ± 3.96	67.21 ± 5.60	101.01 ± 6.90	NA
(0.5 , 0.7)	36.08 ± 3.83	72.08 ± 5.40	108.03 ± 6.59	144.02 ± 7.59	38.31 ± 4.02	76.97 ± 5.75	116.28 ± 7.11	NA
(0.5 , 0.8)	39.12 ± 3.76	78.03 ± 5.24	117.01 ± 6.40	156.01 ± 7.39	44.23 ± 3.91	89.07 ± 5.48	134.98 ± 6.63	NA
(0.5 , 0.9)	42.10 ± 3.62	84.01 ± 5.01	126.00 ± 6.15	168.00 ± 7.09	51.37 ± 3.27	103.17 ± 4.50	155.77 ± 5.18	NA
(0.5 , 1)	45.03 ± 3.39	90.00 ± 4.73	135.00 ± 5.80	180.01 ± 6.70	59.25 ± 0.43	119.25 ± 0.43	179.25 ± 0.43	NA
(0.6 , 0.6)	36.00 ± 3.79	72.00 ± 5.37	108.00 ± 6.58	144.01 ± 7.59	36.00 ± 3.80	72.00 ± 5.36	107.99 ± 6.57	NA
(0.6 , 0.7)	39.07 ± 3.72	78.13 ± 5.27	117.15 ± 6.44	156.12 ± 7.43	39.69 ± 3.78	79.49 ± 5.38	119.46 ± 6.61	NA
(0.6 , 0.8)	42.24 ± 3.67	84.20 ± 5.09	126.09 ± 6.17	168.03 ± 7.12	44.84 ± 3.74	90.04 ± 5.27	135.66 ± 6.42	NA
(0.6 , 0.9)	45.33 ± 3.57	90.05 ± 4.78	135.00 ± 5.81	180.00 ± 6.70	51.52 ± 3.20	103.40 ± 4.45	155.70 ± 5.27	NA
(0.6 , 1)	48.15 ± 3.32	96.00 ± 4.37	143.99 ± 5.36	191.99 ± 6.19	59.30 ± 0.46	119.30 ± 0.46	179.30 ± 0.46	NA
(0.7 , 0.7)	42.01 ± 3.55	84.00 ± 5.02	126.00 ± 6.14	168.01 ± 7.11	42.00 ± 3.55	84.00 ± 5.02	126.00 ± 6.15	NA
(0.7 , 0.8)	45.17 ± 3.41	90.29 ± 4.80	135.29 ± 5.87	180.25 ± 6.76	45.89 ± 3.44	91.96 ± 4.89	138.10 ± 5.99	NA
(0.7 , 0.9)	48.57 ± 3.32	96.29 ± 4.54	144.09 ± 5.42	192.03 ± 6.21	51.78 ± 3.06	103.86 ± 4.30	156.10 ± 5.19	NA
(0.7 , 1)	51.50 ± 3.27	102.01 ± 3.93	153.01 ± 4.79	204.00 ± 5.53	59.35 ± 0.48	119.35 ± 0.48	179.35 ± 0.48	NA
(0.8 , 0.8)	48.00 ± 3.10	96.00 ± 4.38	144.00 ± 5.37	191.99 ± 6.20	48.00 ± 3.10	95.99 ± 4.38	144.00 ± 5.37	NA
(0.8 , 0.9)	51.40 ± 2.83	102.53 ± 4.02	153.47 ± 4.92	204.35 ± 5.64	52.31 ± 2.81	104.85 ± 3.98	157.45 ± 4.85	NA
(0.8 , 1)	54.98 ± 2.85	108.21 ± 3.59	162.08 ± 4.19	216.01 ± 4.65	59.40 ± 0.49	119.40 ± 0.49	179.40 ± 0.49	NA
(0.9 , 0.9)	54.00 ± 2.32	108.00 ± 3.29	162.00 ± 4.02	216.00 ± 4.65	54.00 ± 2.33	108.00 ± 3.29	161.99 ± 4.02	NA
(0.9 , 1)	57.95 ± 1.70	114.93 ± 2.91	171.79 ± 3.60	228.24 ± 3.67	59.44 ± 0.52	119.45 ± 0.50	179.45 ± 0.50	NA
(1 , 1)	60.00 ± 0.00	120.00 ± 0.00	180.00 ± 0.00	240.00 ± 0.00	60.00 ± 0.00	120.00 ± 0.00	180.00 ± 0.00	NA

Table 3.3: The numerical subject benefit results for the design with an after-trial population $S = 1$ billion, and for different trial sizes $T = 60, 120, 180,$ and 240. Each cell is composed of the average number of success responses (first component) added to/subtracted from the corresponding standard deviation (second component) for each scenario (θ_C, θ_D).

Chapter 4

On the Estimation Bias of the Bayesian Decision-Theoretic Response-Adaptive Randomization Procedure

4.1 Introduction

Response-adaptive randomisation (RAR), in which allocation probabilities change as data is being accrued, has the potential to provide improvements to the design of sequential experiments in certain situations. The primary feature of any RAR procedure is the ability to modify the allocation ratio of subjects to arms to follow the objectives of an experiment. When the objectives are defined based on unknown parameters that need to be learnt during the experiment, this goal is achieved by finding an appropriate balance between *learning* i.e., identifying the unknown parameters correctly, and *earning* i.e., allocating subjects to achieve the objectives during the experiment. That is, there are two major objectives which, perhaps with different weights, are typically present when one designs any sequential experiment: (i) minimisation of "learning" errors, such as estimation bias and/or variability often captured by considering the Type I and/or Type II errors; and (ii) minimisation of "earning" errors, such as regret (subject loss).

A very popular design in today's practice is based on the statistical theory of the design of (non-sequential) experiments proposed more than a century ago, *equal fixed randomisation* (EFR), known as the randomised controlled trial in medicine, as the between-group comparison in social sciences, or as the A/B testing in digital marketing. This procedure allocates subjects to all arms with equal probability, and is fixed during the course of the experiment, i.e., it is not response-adaptive. Unfortunately, although being very simple, this procedure does not optimise any of the above objectives (except in very special cases) (Robertson et al., 2020).

Thompson (1933) and Robbins (1952) are among the pioneers of developing RAR procedures with the aim of improving over the suboptimality of EFR. Thompson (1933) proposed a simple procedure within the Bayesian setting, randomising each subject by matching the allocation probability to the probability of each arm being the best, aiming at improving the expected subject benefit, i.e., objective (ii). Robbins (1952) clarified that in order to obtain an estimator of the difference of efficacies with minimal variability (i.e., objective (i)), one should randomise the subjects using fixed randomisation with unequal weights, in which the weights are given by the standard deviations of each arm's efficacy. Robbins (1952) also considered objective (ii) and showed that the "stay-with-a-winner&switch-on-a-loser" rule, which allocates the current subject based only on the allocation and the observation from the last subject, is superior over EFR and there exist response-adaptive procedures that perform even better (asymptotically, as the number of subjects approaches infinity).

The literature on RAR is rich and abounds with ad hoc procedures whose properties are however not easy to understand theoretically for finite-time experiments, and thus most of the theoretical results are asymptotic, while the standard of evaluation and comparison for finite-time experiments is based on computational (typically simulation) studies. Notable contributions include Wei and Durham (1978), who proposed the randomised play-the-winner rule described as an urn model with replacement where the allocation process is specified by drawing a ball from the urn. Addressing some common criticisms to adaptive procedures, Berry and Eick (1995) compared four RAR procedures with the EFR. They showed that, in the case of less prevalent disease, RAR procedures should be the preferred choice. Studies such as Rosenberger et al. (2001) who considered objective (i), proposed a RAR procedure with which the optimal allocation can be derived asymptotically, and Ivanova and Rosenberger (2001) where, considering *Neyman allocation*, the authors show that the allocations obtained by two RAR procedures based on sequential maximum likelihood estimation (MLE) and an urn model are very close to the optimal.

The trade-off between learning and earning is huge. Typically, procedures optimised for one of these two objectives perform very poorly in the other one. The most illustrative of this is the Bayesian decision-theoretic RAR procedure (see, e.g., Jacko (2019b) for a review) which minimises the Bayes-expected earning errors, and results in subject allocations which are deterministic (as functions of the past observations and the remaining size of the experiment) except when several arms are identified as equivalent, when any deterministic or randomised allocation between them is allowed. When using traditional statistical inference methods (which are not specifically adapted to RAR procedures), the operational characteristics such as statistical power or estimation bias are very far from the levels that could be achieved using EFR. One approach to decrease its learning errors to acceptable levels is by forcing randomisation over these deterministic decisions, see, e.g., Cheng and Berry (2007); Williamson et al. (2017). Another approach is to use inference methods which are better suited for RAR procedures, such as randomisation tests (Rosenberger et al., 2019) and estimation-correcting methods (Bowden and Trippa, 2017). However, these typically do not apply to deterministic RAR procedures.

4.1.1 Estimation Bias under RAR Procedures

In the design of experiments, one of the key “learning” measures is the bias and the variance of efficacy estimators. All traditional estimators are biased when used on data obtained by any RAR procedures. That is because of the fact that the randomisation probabilities can be adjusted to skew allocations to achieve their objectives as the experiment moves along and responses accrue. Bowden and Trippa (2017) showed that the maximum likelihood estimator (MLE) under RAR procedures is calculated as the sample mean (i.e., in the same way as in non-adaptive experiments) when the parameter of interest is the efficacy. The authors developed a formula for the magnitude of the bias of the MLE induced by RAR procedures, along with proposing several methods for obtaining unbiased estimates, which, however, suffer from large variance. Ji et al. (2019) evaluated the estimation bias in biomarker-stratified RAR experiments. They showed that estimation of log odds ratio in experiments in which randomisation ratios depend on both covariates and subject responses, are usually associated with bias and proposed an asymptotically unbiased estimator which utilises the experiment size (total enrolled subjects). Hadad et al. (2021) introduced a particular class of test statistics and, in turn, asymptotically normally distributed estimators to develop frequentist confidence intervals for intervention-efficacy in adaptive experimental settings. The paper shows that for any non-deterministic RAR procedures, though the inverse-probability weighting (IPW) estimator tends to be unbiased, it does not result in a normal asymptotic distribution. To address this, the authors formulate a novel estimator called adaptively weighted augmented IPW. Estimation bias in deterministic RAR procedures is much less understood. Nie et al. (2018) focused on deterministic RAR procedures and proved that when the data collection process satisfies certain conditions (*exploit* and *independence of irrelevant option*), then each arm's MLE (i.e., sample mean) is associated with negative bias. In Shin et al. (2019a) and Shin et al. (2019b), adaptive *sampling*, adaptive *stopping*, adaptive *choosing*, and adaptive *rewinding* were identifed as the sources of the bias of the MLE in the context of deterministic RAR procedures. In both studies, the authors illustrated that the sign of the bias varies according to the assumptions of the procedure. Considering the monotonic behaviors of the data collection strategies and a new notion called "optimism", they delivered a comprehensive insight into not only the bias, but risk and consistency of the MLE. Furthermore, they argued that the reason why Nie et al. (2018) encountered negative bias for MLE is that they considered an optimistic sampling for a given arm at a fixed time, while it is not the case when deceptively stopping and/or choosing rules are also assumed in the experiment. Another factor which may affect the magnitude and direction of the bias is the time trend (Villar et al., 2018; Jiang et al., 2020). Considering a time trend together with the early stopping assumption in an experiment can cause one to estimate treatment effect with even larger bias. To make some notes on the importance of choosing a prior in Bayesian settings, in the field of educational technology Rafferty et al. (2019) showed that although MAB experiments perform very well in comparison to traditional designs in terms of students' benefit, to maintain statistical power at least twice as many students participating in a trial are needed. To alleviate this problem, they suggested using optimistic (*prior above*) and pessimistic (*prior below*) distributions in which prior parameters are not equal necessarily. Finally, a decreasingly informative prior (DIP) model to mitigate variability in adaptive allocation ratio, has been introduced by Sabo (2014). A so-called natural "lead-in" period where the allocation weights are not being changed and are being allocated based on the mode of prior distribution has been incorporated through making prior distributions in a Bayesian design.

4.1.2 Our Contributions

In this chapter we study the estimation bias of the Bayesian decision-theoretic RAR procedure, which focuses on objective (ii) and delivers the minimal Bayes-expected earning errors. For binary (success/failure) responses, it means maximising the Bayes-expected number of successes by the end of the experiment. It can be obtained as a solution to the so-called finite-horizon Bayesian multi-armed bandit problem by dynamic programming (DP) (Williamson et al., 2017; Zhang et al., 2019) or approximately by using the so-called Whittle index (Villar et al., 2015a;

Williamson et al., 2017; Villar, 2018; Panigrahi et al., 2016) and performs as the best or nearly-best also in the sense of frequentist expectation of the number of successes (or, equivalently, cumulative regret) when compared to other procedures in a wide range of scenarios (Jacko, 2019b).

We consider the usual model of the Bayesian Beta-Bernoulli two-armed problem with binary responses in section 3.1. Our contributions are as follows:

- We characterise theoretically the estimation bias of the MLE and prove that it is always associated with negative bias in section 4.2. We also discuss other estimators and investigate their performance, for instance the bias-corrected estimators proposed by Bowden and Trippa (2017), which however do not apply to deterministic RAR procedures and we illustrate that the inverse probability weighted (IPW) estimator has a notably larger bias than the MLE.

- The weaknesses of all the above estimators in bias correction motivate us in section 4.3 to introduce the *augmented estimator*, in which observations are augmented with pseudo-observations, and we derive a formula for its estimation bias, which gives insights into possible approaches by which estimation bias could be substantially mitigated.

- In section 4.4 we characterise the estimation bias of the MLE and other estimators numerically in an extensive simulation study, showing that the bias is unacceptably large in many scenarios. We investigate different combinations of augmentations in order to tune the augmented estimator and numerically characterise its bias. Furthermore, we also illustrate that the estimation bias of the IPW estimator can be reduced using thoughtfully selected augmentations.

Simulation set-up information is presented in section 2.5. We outline the main conclusions of this paper and highlight areas for future research in section 4.5. All proofs and additional computational experiments can be found in the appendix 4.6.

4.2 Estimation Bias: Theoretical Characterization

In this section, we provide a comprehensive analysis of theoretical aspects of the bias of the MLE in the problem described in section 3.1. Then, we introduce a novel estimator called an *augmented estimator*, and subsequently characterise the bias associated with it. In section 4.4 using extensive simulation studies, we illustrate that the augmented estimator can mitigate the bias significantly. Note that both the MLE and Bayesian estimator can be recovered with appropriate choices of priors and augmentations of the augmented estimator. To do so, we assume that unknown success probabilities θ_k are Beta-distributed with parameters α_k and β_k:

$$\theta_k \sim Beta(\alpha_k, \beta_k) \quad \forall k \in \mathcal{K}$$

4.2.1 Maximum Likelihood Estimator and its Bias

To begin with, we characterize the MLE and its bias in the context of the design with binary responses with any fixed or response-adaptive randomization. See, e.g., Bowden and Trippa (2017), which we follow and build on.

Theorem 4.1. *The joint likelihood function at time epoch* $\tau \in \mathcal{T} \cup \{T\}$, *i.e., having observed* $s_k(\tau)$ *successes and* $f_k(\tau)$ *failures on each arm* $k \in \mathcal{K}$, *is proportional to the joint likelihood function for* K *observations, each having independent Beta distribution with parameters* $\alpha_k = s_k(\tau) + 1$ *and* $\beta_k = f_k(\tau) + 1$.

Theorem 4.2. *The maximum likelihood estimator at time epoch* $\tau \in \mathcal{T} \cup \{T\}$, *i.e., having observed* $s_k(\tau)$ *successes and* $f_k(\tau)$ *failures on each arm* $k \in \mathcal{K}$, *is*

$$\widehat{\theta}_k(s_k(\tau), f_k(\tau)) = \begin{cases} \dfrac{s_k(\tau)}{n_k(\tau)} & \text{for } n_k(\tau) > 0 \\ \text{any value in } (0,1) & \text{for } n_k(\tau) = 0 \end{cases} \tag{4.1}$$

where $n_k(\tau) = s_k(\tau) + f_k(\tau)$.

In what follows, we will also refer to the MLE briefly as the *Frequentist estimator*. We will also use the short-hand notation $\widehat{\theta}_k(\tau) = \widehat{\theta}_k(s_k(\tau), f_k(\tau))$. Next we state a result which we will later use to draw analogies between the Frequentist and Bayesian estimators.

Theorem 4.3. *The maximum likelihood estimator at time epoch* $\tau \in \mathcal{T} \cup \{T\}$, *i.e., having observed* $s_k(\tau)$ *successes and* $f_k(\tau)$ *failures on each arm* $k \in \mathcal{K}$, *coincides with the mode of the Beta distribution with parameters* $\alpha_k = s_k(\tau) + 1$ *and* $\beta_k = f_k(\tau) + 1$.

The bias of MLE induced in RAR procedures is initially formulated in the study by (Bowden and Trippa, 2017); Shin et al. (2019a) also proposed applying the similar formula to the trials with early and/or adaptively stopping time.

Theorem 4.4. *The bias of the maximum likelihood estimator given in* (4.1) *is*

$$Bias\Big[\widehat{\theta}_k(\tau)\Big] := E\Big[\widehat{\theta}_k(\tau)\Big] - \theta_k = \begin{cases} \dfrac{-Cov\Big[n_k(\tau), \widehat{\theta}_k(\tau)\Big]}{E\Big[n_k(\tau)\Big]} & for \quad n_k(\tau) > 0 \quad \forall k \in \mathcal{K} \\ 0 & for \quad n_k(\tau) = 0 \end{cases} \tag{4.2}$$

According to Theorem 4.2, and in the case of $n_k(\tau) = 0$ we can choose an arbitrary value like γ and set $\gamma := \widehat{\theta}_k(\tau) \in (0,1)$. Then, $\text{Bias}\Big[\widehat{\theta}_k(\tau)\Big] = n_k(\tau).\Big(\gamma - \theta_k\Big) = 0 \quad \forall k \in \mathcal{K}$.

It is worth mentioning that the proof of this theorem can be obtained upon appropriately setting out the parameters of interest through the proof of Theorem 4.9. Note that any design whose subject allocation is independent of past responses, e.g., fixed randomization with arbitrary randomization ratio, will have a zero covariance and thus a zero bias on every arm. For the designs whose subject allocation depends on or is correlated with past responses, a particular arm may have a bias away from zero if either the covariance is away from zero, or if $E\Big[n_k(\tau)\Big]$ is small, or both. The sign of the bias is typically negative for each arm under the response-adaptive allocation procedures which aim at (exactly/approximately/asymptotically) maximizing the expected number of observed successes, since these tend to allocate more subjects to arms with higher estimates. However, this is not guaranteed by the above characterization in general, as shown in Nie et al. (2018) who provided sufficient conditions for it to be true. In particular, negative bias is not guaranteed for those designs that use other estimators rather than the MLE for allocation decisions, which, among others, include upper confidence bound procedures and Bayesian procedures. The sign of the bias can be positive for some or all arms for some procedures, for instance for those that aim at maximizing the statistical power, which, in some circumstances, tend to allocate more subjects to arms with lower estimates (Rosenberger et al., 2001).

4.2.2 Bayesian Estimators and their Bias

In this section we characterise the Bayesian estimators along with their associated bias. Note that, these estimators are formulated based on the posterior distribution, and the posterior mean is a natural estimator being used in Bayesian statistics and in Bayesian decision theory. The posterior mode is less commonly used but it is also a reasonable choice for a Bayesian estimator.

Theorem 4.5. *Consider time epoch $\tau \in \mathcal{T} \cup \{T\}$ and for arm $k \in \mathcal{K}$, let $s_k(\tau)$ and $f_k(\tau)$ be the numbers of observed successes and failures, respectively, and let $(\widetilde{s}_k(0), \widetilde{f}_k(0))$ be the parameters of the prior Beta distribution. Then, the posterior distribution is a Beta distribution with parameters $\widetilde{s}_k(\tau) := s_k(\tau) + \widetilde{s}_k(0)$, $\widetilde{f}_k(\tau) := f_k(\tau) + \widetilde{f}_k(0)$, and thus the* posterior mean *is*

$$\widetilde{\theta}_k^{Mn}(\tau) := \begin{cases} \dfrac{\widetilde{s}_k(\tau)}{\widetilde{n}_k(\tau)} & \textit{for } \widetilde{n}_k(\tau) > 0 \\ \textit{any value in } (0,1) & \textit{for } \widetilde{n}_k(\tau) = 0 \end{cases} \tag{4.3}$$

and, assuming $\widetilde{s}_k(\tau), \widetilde{f}_k(\tau) \geq 1$, the posterior mode *is*

$$\widetilde{\theta}_k^{Md}(\tau) := \begin{cases} \dfrac{\widetilde{s}_k(\tau) - 1}{\widetilde{n}_k(\tau) - 2} & \textit{for } \widetilde{n}_k(\tau) > 2 \\ \textit{any value in } (0,1) & \textit{for } \widetilde{n}_k(\tau) = 2 \end{cases} \tag{4.4}$$

Following the convention, we assume that the prior parameters do not change with time, which are thus related to time epoch 0 rather than τ. The following two theorems are obtained as a special case of our main result presented in Theorem 4.9.

Theorem 4.6. *The bias of the posterior mean given in* (4.3) *is*

$$Bias\Big[\widetilde{\theta}_k^{Mn}(\tau)\Big] := E\Big[\widetilde{\theta}_k^{Mn}(\tau)\Big] - \theta_k = \frac{-Cov\Big[\widetilde{n}_k(\tau), \widetilde{\theta}_k^{Mn}(\tau)\Big] + \widetilde{n}_k(0)\Big(\widetilde{\theta}_k^{Mn}(0) - \theta_k\Big)}{E\Big[\widetilde{n}_k(\tau)\Big]} \tag{4.5}$$

Note that $\widetilde{\theta}_k^{\mathrm{Mn}}(0) = \widetilde{s}_k(0)/\widetilde{n}_k(0)$ is the prior mean, thus $\widetilde{\theta}_k^{\mathrm{Mn}}(0) - \theta_k$ is the bias of the prior mean. If it has an opposite sign to the sign of the covariance, then the bias of the posterior mean is further away from zero. This is further exacerbated if the prior is very informative, i.e., if $\widetilde{n}_k(0)$ is large.

Theorem 4.7. *The bias of the posterior mode given in* (4.4) *is*

$$\begin{aligned} Bias\Big[\widetilde{\theta}_k^{Md}(\tau)\Big] :=& E\Big[\widetilde{\theta}_k^{Md}(\tau)\Big] - \theta_k \\ =& \frac{-Cov\Big[\widetilde{n}_k(\tau) - 2, \widetilde{\theta}_k^{Md}(\tau)\Big] + \widetilde{n}_k(0)\Big(\widetilde{\theta}_k^{Mn}(0) - \theta_k\Big) - 2\Big(\frac{1}{2} - \theta_k\Big)}{E\Big[\widetilde{n}_k(\tau) - 2\Big]} \end{aligned} \tag{4.6}$$

The above results lead to the following corollaries.

Corollary 4.1. *If the parameters of the prior Beta distribution are* $(\widetilde{s}_k(0), \widetilde{f}_k(0)) =$

$(0, 0)$ *(the so-called* Haldane prior*), then the posterior mean is*

$$\widetilde{\theta}_k^{Mn}(\tau) = \begin{cases} \dfrac{s_k(\tau)}{n_k(\tau)} & \textit{for } n_k(\tau) > 0 \\ \textit{any value in } (0, 1) & \textit{for } n_k(\tau) = 0 \end{cases} \tag{4.7}$$

and its bias simplifies to

$$Bias\Big[\widetilde{\theta}_k^{Mn}(\tau)\Big] = \frac{-Cov\Big[n_k(\tau), \widetilde{\theta}_k^{Mn}(\tau)\Big]}{E\Big[n_k(\tau)\Big]} \tag{4.8}$$

and thus it coincides with the Frequentist maximum likelihood estimator and its bias.

Corollary 4.2. *If the parameters of the prior Beta distribution are* $(\widetilde{s}_k(0), \widetilde{f}_k(0)) = (1, 1)$ *(the so-called* Bayes prior*), then the posterior mode is*

$$\widetilde{\theta}_k^{Md}(\tau) = \begin{cases} \dfrac{s_k(\tau)}{n_k(\tau)} & \textit{for } n_k(\tau) > 0 \\ \textit{any value in } (0, 1) & \textit{for } n_k(\tau) = 0 \end{cases} \tag{4.9}$$

and its bias simplifies to

$$Bias\Big[\widetilde{\theta}_k^{Md}(\tau)\Big] = \frac{-Cov\Big[n_k(\tau), \widetilde{\theta}_k^{Md}(\tau)\Big]}{E\Big[n_k(\tau)\Big]} \tag{4.10}$$

and thus it coincides with the Frequentist maximum likelihood estimator and its bias.

4.2.3 HT and IPW Estimators

In the study by Bowden and Trippa (2017), the authors propose a simple bias-corrected estimator for success probabilities, called Horvitz–Thompson(HT) which utilises randomisation probabilities $\pi_k(t)$ at each time epoch t, as below:

$$\widehat{\theta}_{k,HT} = \frac{1}{T}\sum_{t=1}^{T}\frac{\delta_k(t)y_k(t)}{\pi_k(t)} \tag{4.11}$$

This estimator, nowadays also called Inverse Probability Weighted (IPW) estimator in the literature as in the study by Hadad et al. (2021) where $Y_k(t)$ represents the random variable (and $y_k(t)$ the realization) of the response at time epoch $t \in \mathcal{T}$ corresponding to arm $k \in \mathcal{K}$, and $\delta_k(t)$ stands for a binary variable equalling 1 if time epoch t is allocated to arm k and 0 otherwise. Denote by $k(t) \in \mathcal{K}$ the actual assigned arm at each time epoch t such that $\delta_{k(t)}(t) \equiv 1$, then, assuming a notion of the inverse probability weighted (IPW) by which the values of HT estimator can be fitted within the unit interval, the "normalised" version of the HT estimator is defined as follows:

$$\widehat{\theta}_{k,IPW} = \frac{T\widehat{\theta}_{k,HT}}{\sum_{t=1}^{T}\frac{\delta_k(t)}{\pi_{k(t)}(t)}} = \frac{\sum_{t=1}^{T}\frac{\delta_k(t)y_k(t)}{\pi_k(t)}}{\sum_{t=1}^{T}\frac{\delta_k(t)}{\pi_{k(t)}(t)}} \tag{4.12}$$

Finally, using Rao-Blackwellization result in which all permutations of the order of columns in the realisation matrix have been computed, a significant improvement on the HT estimator, by averaging out all HT estimators corresponding to each permutation, has been achieved. However, we show that neither of above estimators can necessarily either mitigate or eliminate the bias of estimation in

RAR settings. Therefore, this has been taken into account as a motivation for this study which ultimately leads us to introduce a novel estimator in the following section.

To begin with, we recall the primary assumption about the action set considered in this study. We assume three possible cases for randomisation probabilities, based on which the action set $\mathcal{A}_{(x)}$ is built: $\mathcal{A}_{(x)} \subseteq \Big\{a;\ \pi^a_C + \pi^a_D = 1,\ \pi^a_C,\ \pi^a_D \in \{0, 1, 1/2\}\Big\}$. The third case, i.e., $\pi^a_C = \pi^a_D = 1/2$, can happen when there is no difference between arms allocation. Without loss of generality, we can presume that:

$$\pi_k(t) = \begin{cases} \epsilon & \text{if } \delta_k(t) = 0 \quad \forall k \in \mathcal{K} \\ \\ 1 - \epsilon & \text{if } \delta_k(t) = 1 \end{cases} \tag{4.13}$$

Assuming a non-randomised RAR procedure rearranges the equation (4.11) (by applying a limit on the summation of each time epoch divided by the corresponding randomisation probability, when epsilon tends to zero) in the following equality in which the number of success observations is divided by the trial size rather than the sample size on a given arm.

$$\widehat{\theta}_{k,HT} = \frac{s_k(T)}{T} \tag{4.14}$$

Hence, in comparison with MLE, it is quite obvious that the next equality tends to estimate any probability of successes with smaller values in magnitude, consequently leading us to estimate success probabilities with more bias at the end of the trial. Figure 4.5 (left-hand-side column) represents HT estimators for four different trial sizes assumed in this study. It is also noteworthy to mention that the normalised version of HT estimator (4.12) is functioning as an MLE, since the term $\sum_{t=1}^{T} \frac{\delta_k(t)}{\pi_{k(t)}(t)}$ in the denominator represents the sample size on a given arm at the end of the trial. In section 4.4 we introduce a thoughtfully-selected augmentation by which the bias of the IPW estimator is notably reduced. However, in the study by Bowden and Trippa (2017) the authors employ Rao-Blackwellization technique, which suffers from a rather high computational complexity associated with computing permutations to achieve $\widehat{\theta}_{RBHT}$, to improve the HT estimator. Considering the assumption of not having any early stopping or adaptive stopping in the trial one needs to carry out a factorial of trial size i.e., $T!$ computations along with computing the HT estimator in each replica in order to calculate the $\widehat{\theta}_{RBHT}$. To overcome this, the authors suggest using Monte Carlo approximation, since integration used for likelihood function becomes unfeasible in large enough trials. Hence, the assumed trial size by Bowden and Trippa (2017) is only 25 which is computationally achievable to calculate, whilst in the present study trial sizes are a multiple of 60. Since in practical settings the size of trials are assumed quite large, the performance of the $\widehat{\theta}_{RBHT}$ estimator is poor and at some points it is expensive to compute.

4.3 Augmented Estimator and its Bias

From the practical point of view, and due to potential limitations on resources, budgets and also ethical obligations, trials are usually conducted once or for a limited number of repetitions. Hence, by using extensive simulation studies, we intend to develop a framework by which an adequate augmentation can be selected. Augmentations are formulated based on the trial's characteristics and specifications. By applying the proposed framework in the efficacy calculation process, one can select an appropriate level of augmentation to make the MLE estimation relatively accurate at the end of the trial.

In this section we characterise the Bayesian posterior mean, which is a natural estimator being used in Bayesian designs, and we will derive its bias. We will do so as a special case of our newly-proposed *augmented estimator*, defined next.

Definition 4.1. Consider time epoch $\tau \in \mathcal{T} \cup \{T\}$ and for arm $k \in \mathcal{K}$. Let $s_k(\tau)$ and $f_k(\tau)$ be the numbers of observed successes and failures, respectively, let $(\widetilde{s}_k(0), \widetilde{f}_k(0))$ be the parameters of the prior Beta distribution, and let $\dot{s}_k(\tau), \dot{f}_k(\tau)$ be called the *augmentations* of the numbers of successes and failures, respectively. Denote $\dot{n}_k(\tau) := \dot{s}_k(\tau) + \dot{f}_k(\tau)$ and suppose that $n_k(\tau) + \widetilde{n}_k(0) + \dot{n}_k(\tau) \geq 0$. Then, we define the *augmented estimator* as

$$\widetilde{\theta}_k(\tau) := \begin{cases} \dfrac{s_k(\tau) + \widetilde{s}_k(0) + \dot{s}_k(\tau)}{n_k(\tau) + \widetilde{n}_k(0) + \dot{n}_k(\tau)} & \text{for } n_k(\tau) + \widetilde{n}_k(0) + \dot{n}_k(\tau) > 0 \\ \text{any value in } (0,1) & \text{for } n_k(\tau) + \widetilde{n}_k(0) + \dot{n}_k(\tau) = 0 \end{cases} \tag{4.15}$$

Theorem 4.8. *The augmented estimator $\widetilde{\theta}_k(\tau)$ is asymptotically consistent.*

Proof. According to the definition of consistency we have:

$$p \lim_{n_k \to \infty} \widetilde{\theta}_k(\tau) = \theta_k \quad \forall k \in \mathcal{K}$$

That is, if, for all $\epsilon > 0$

$$\lim_{n_k \to \infty} \Pr\left(\left|\widetilde{\theta}_k(\tau) - \theta_k\right| > \epsilon\right) = 0 \quad \forall k \in \mathcal{K}$$

Recalling the fact that both prior beliefs and augmentations are constants:

$$\begin{aligned}\lim_{n_k\to\infty}\left(\widetilde{\theta}_k(\tau)-\theta_k\right) &= \lim_{n_k\to\infty}\left(\frac{s_k(\tau)+\widetilde{s}_k(0)+\dot{s}_k(\tau)}{n_k(\tau)+\widetilde{n}_k(0)+\dot{n}_k(\tau)}-\theta_k\right)\\ &\equiv \lim_{n_k\to\infty}\left(\frac{s_k(\tau)}{n_k(\tau)}-\theta_k\right)=0\end{aligned} \tag{4.16}$$

The last term indicates that augmented estimator converges to MLE, i.e. $\dfrac{s_k(\tau)}{n_k(\tau)}$, and it has been proved that MLE is asymptotically consistent estimator. □

Note that we allow for the augmentations to change with time, which are thus related to time epoch τ. The augmented estimator can be seen in both the Bayesian and the Frequentist setting; for the Frequentist setting we would simply set $\widetilde{s}_k(0)=\widetilde{f}_k(0)=0$ and thus consider the Frequentist version.

$$\widetilde{\theta}_k(\tau)=\begin{cases}\dfrac{s_k(\tau)+\dot{s}_k(\tau)}{n_k(\tau)+\dot{n}_k(\tau)} & \text{for } n_k(\tau)+\dot{n}_k(\tau)>0\\ \text{any value in } (0,1) & \text{for } n_k(\tau)+\dot{n}_k(\tau)=0\end{cases} \tag{4.17}$$

In turn, the Frequentist MLE is recovered when the augmentations are $\dot{s}_k(\tau)=\dot{f}_k(\tau)=0$. In the Bayesian setting, the posterior mean is recovered when the augmentations are $\dot{s}_k(\tau)=\dot{f}_k(\tau)=0$.

The idea of augmentation of the numbers of successes and failures in statistical inference is not new. For instance, Agresti and Caffo (2000) suggested constructing confidence intervals around an augmented estimator. They showed that these adjusted confidence intervals, which can be derived by adding pseudo observations, half of each type, to the MLE i.e., the sample mean, can simply bypass sample size rules in the Wald approach. They also numerically proved that adding four pseudo observation performs the best among other adjusted cases. The second best, which indicates adding two pseudo observations, coincides with Bayesian estimator with prior beliefs $(\widetilde{s}_k(0), \widetilde{f}_k(0)) = (1,1)$, and equivalently with the mode of the Beta distribution with parameters $\alpha_k = s_k(\tau) + 1$ and $\beta_k = f_k(\tau) + 1$. See Theorem (4.3).

Theorem 4.9. *The bias of the augmented estimator given in (4.15) is*

$$\begin{aligned} Bias\Big[\widetilde{\theta}_k(\tau)\Big] &:= E\Big[\widetilde{\theta}_k(\tau)\Big] - \theta_k \\ &= \frac{-Cov\Big[n_k(\tau) + \widetilde{n}_k(0) + \dot{n}(\tau), \widetilde{\theta}_k(\tau)\Big] + \widetilde{n}_k(0)\Big(\widetilde{\theta}_k(0) - \theta_k\Big) + \dot{n}_k(\tau)\Big(\dot{\theta}_k(\tau) - \theta_k\Big)}{E\Big[n_k(\tau) + \widetilde{n}_k(0) + \dot{n}(\tau)\Big]} \end{aligned} \tag{4.18}$$

where $\widetilde{\theta}_k(0)$ *is the prior mean, and* $\dot{\theta}_k(\tau) = \frac{\dot{s}(\tau)}{\dot{n}(\tau)}$.

4.4 Estimation Bias: Numerical Characterization for Decision-Theoretic Designs

Figure 4.2 represents the MLE simulation results corresponding to frequentist estimator i.e., the equation (4.1) in the left-hand-side plots, and the Bayesian one i.e., equation (4.3) with Bayes prior in the right-hand-side. As a general observation, the covariance of the trial size and either the frequentist or the Bayesian estimator increases when a larger time horizon is taken into account. However, trial size variation fails to lessen the bias induced significantly.

The frequentist estimator tends to have a larger (negative) bias on the worse arm, and may reach as much as ≈ -0.23 on the worse arm and ≈ -0.18 on the better arm. The single-arm bias tends to deteriorate with increased arm D effect and with decreased efficacy values. Although the bias on both arms is negative, the single-arm bias may affect also the estimate of the arm D effect. This estimator tends to notably inflate the effect, e.g., if $\theta_D - \theta_C = 0.1$ (blue circles and stars), the bias in the effect estimation may be as large as $+0.14$ for the time horizon $T = 240$ (estimating the effect to be 0.24). One of the exceptional scenarios of this class is $(0.9, 1)$. On the contrary to other scenarios in this class, where the larger values of θ_C and θ_D are assumed, as the bias and covariance increase, the bias of scenario $(0.9, 1)$ falls to less than -0.1. It is also quite noticeable that for those scenarios with bigger arm D effects, the gap between both arms is wider. This phenomenon confirms that due to lack of observations on worse arm (arm C in these cases), the bias tends to take larger (negative) values.

On the other hand, the bias of the Bayesian estimator is much more complicated to capture, because besides the two effects present for the Frequentist estimator it tends to estimate each arm closer to its prior mean due to the fact that $\frac{\widetilde{s}_k}{\widetilde{s}_k+\widetilde{f}_k}$ is between the values of the Frequentist estimator $\frac{s_k}{s_k+f_k}$ and the prior mean $\frac{\widetilde{s}_k(0)}{\widetilde{s}_k(0)+\widetilde{f}_k(0)}$. This effect goes in the same direction (negative bias) as the other two effects if θ_k is greater than the prior mean, but goes in the opposite direction if θ_k is lower than the prior mean, and it is particularly notable on the worse arm as it tends to have fewer allocations. When using the Bayes prior (which is uninformative, with the mean 0.5), the bias on the worse arm can reach as much as -0.2 or $+0.35$ (scenario $(0.9, 1)$ is the only case out of this range). Some general conclusions on the Bayesian estimator one: (i) the covariance values decline in comparison with the frequentist case (scenarios with equal success probabilities are more or less the same), (ii) estimation for worse arm in scenario $(0, 1)$ has been calculated ≈ 0.34, (iii) the bias of the inferior arms for those scenarios with $\theta_D - \theta_C \geq 0.5$

has been estimated positively while superior ones are unbiased, (iv) the density around origin confirms that the bigger trial size, the less biased the superior arms, and (v) in those scenarios with positive bias values, as the difference between success probabilities decreases the inferior arm estimates with smaller (positive) bias values. This estimator thus may either inflate or deflate the effect, e.g., if $\theta_D - \theta_C = 0.1$, the bias in the effect estimation may be as large as $+0.04$ (estimating the effect to be 0.14) or as low as -0.22 (estimating the effect to be 0.21), depending on whether the efficacy values are closer to 0 or closer to 1. Unfortunately, in situations of practical relevance such as short ($T \approx 10^1$—10^2) and moderate ($T \approx 10^3$—10^4) trial duration, the bias on each arm is non-negligible.

4.4.1 Modification criterion and estimation functionality

In order to mitigate these effects, in this thesis we explore a number of heuristical modifications applying to the end-of-trial Frequentist MLE chiefly, and the Bayesian estimator partially. These modifications, apart from their prior specifications $\left(\widetilde{s}_k(0), \widetilde{f}_k(0)\right)$ which clarify the type of initial estimator, are determined by adjusting augmentations to the desired/intended values. Technically, the augmentation $\dot{f}_k(\tau)$ should be valued at negative numbers. That is because (i) reducing the number of failure observations gives rise to notably modifying MLE, (ii) applying an appropriate negative $\dot{f}_k(\tau)$ converts the Bayesian estimator to MLE, accordingly. However, in this study we assume $\dot{f}_k(\tau) = 0 \ \ \forall \tau$.

To identify the worse-performing arm and modify its corresponding MLE, we compare $\widetilde{\theta}_C^{\mathrm{Mn}}(T)$ and $\widetilde{\theta}_D^{\mathrm{Mn}}(T)$, for each simulation iteration. Once the inferior arm has been determined, the pre-fixed augmentations are being added to both numerator and denominator of (4.3) to form the corresponding augmented estimator (4.15). In other words, we increase the estimator of the arm whose (unmodified) estimator is lower, and we do so being encouraged by the fact that the bias of the Frequentist estimator is always negative (Nie et al., 2018). Although this is not a perfect logic for Bayesian estimator, we apply the novel idea for some specific augmentations to Bayesian estimator. Note that those simulation replications in which $\widetilde{\theta}^{\text{Mn}}_{C}(T) = \widetilde{\theta}^{\text{Mn}}_{D}(T)$, the estimator is not modified, and considered as it is estimated. Finally, the ultimate estimated efficacy, which would be closer to the actual corresponding success probability, and therefore less biased, is obtained by an average over the total number of simulation iterations.

To explain the idea of modification and augmented estimators in general and clear terms, we propose a framework in which each augmentation can be composed of three multiplicative parts: a coefficient, a root function of natural logarithm of trial size (square and cube root perform the best), and a power function of the MLE (4.1) (Bayesian estimator (4.3) could also be the case). Hence, the proposed formula can be considered in a general sense as follows:

$$\dot{n}_k(T) = \dot{s}_k(T) + \dot{f}_k(T)\Big|_{\dot{f}_k(T)=0} = A.\sqrt[n]{\ln(T+1)}.\left(\widehat{\theta}_k\right)^p \tag{4.19}$$

By virtue of a trial and error technique along with extensive simulation studies, we noticed that the function above 4.19 performs better when the exponent in the power function is odd whilst the root for natural logarithm is even. Also, the constant being chosen to scale the effect of two others should not be very large. Table 4.1 summarises the augmentations' specification using the equation (4.19). Note that, corresponding simulation outcomes are presented in figures below, and

		$\dot{s}_k(T)$	Prior belief	Statistics
Fig. 4.1	RHS	0	Bayes	Bayesian
	LHS	0	-	Frequentist
Fig. 4.2	RHS	$\widetilde{\theta}_k^{\text{Mn}}(T)$	Bayes	Bayesian
	LHS	$\widehat{\theta}_k(T)$	-	Frequentist
Fig. 4.3	RHS	$4\ln(T+1)(\widehat{\theta}_k(T))^2$	-	Frequentist
	LHS	$\ln(T+1)\widehat{\theta}_k(T)$	-	Frequentist
Fig. 4.4	RHS	$27\sqrt{\ln(T+1)}(\widehat{\theta}_k(T))^3$	-	Frequentist
	LHS	$9\ln(T+1)(\widehat{\theta}_k(T))^3$	-	Frequentist

Table 4.1: Specifications of modifications

those with Jeffreys prior may be found in the Appendix 4.6.4.

4.4.2 Augmented Estimator Performance

To begin with, figure 4.2 represents augmented estimators in which the value of the augmentation is considered the same as MLE, i.e. $\dot{s}_k(T) = \widehat{\theta}_k(T)$. The bias of the augmented Bayesian estimators (right-hand side column) follows the same pattern as figure 4.1 with two main differences. First, covariance values are slightly declined compared to the original Bayesian equivalent. Second, for those scenarios with success probabilities on arm D closer to zero, estimation is too optimistic. In other words, the bias values lie far from the origin on large positive values, while the reductions in the negative side do not seem substantial. Note that we ignore the augmented Bayesian estimator and do not present the simulation results for further augmentations later in this section.

On the other hand, the Frequentist results illustrate that: (i) overall, a slight improvement, which looks as if a horizontal transfer has been occurred, in bias values can be observed, (ii) as in the augmented Bayesian case, lower covariance values can be discerned in comparison with the original Frequentist equivalent, (iii) the inflation patterns in each trial size are also in same proportion to the original equivalent. It is quite wise to try some other augmentations with which the improvements in either bias values or inflation patterns could be more significant and substantial. To do so, we consider a well-chosen multiplier which draws a connection between trial size and assumed augmentation. Further variants on this multiplier and the augmented estimator together with allocating an appropriate constant led us to have ≈ 0.15 improvement in the worse case scenario.

Figure (4.3) compares simulation results obtained from the augmentations $\dot{s}_k(T) = \ln(T+1)\widehat{\theta}_k(T)$ on the left-hand-side and $\dot{s}_k(T) = 4\ln(T+1)(\widehat{\theta}_k(T))^2$ on the right hand side with one another. It is quite obvious that the overall magnitude of the reduction in bias values, in the worst case scenario, reports around 0.05 which emphases the importance of considering an appropriate constant as a multiplier and variant, square function in this case, of the augmented estimator. In addition, the bias in the arm D equal to 0.1, i.e. $\theta_D - \theta_C = 0.1$ noticeably diminishes since the bias of arm C, which has been considered inferior arm, is modified by large and well-chosen augmentations. Note that, like in previous cases, slight reductions in covariance values are made as the augmentations are enlarged.

Finally, the figure 4.4 provides a comparison between augmentations $\dot{s}_k(T) = 9\ln(T+1)(\widehat{\theta}_k(T))^3$ on the left-hand-side and $\dot{s}_k(T) = 27\sqrt{\ln(T+1)}(\widehat{\theta}_k(T))^3$ in the right hand side. First of all, by comparing the right-hand-side plots with their equivalent on left-hand-side in figure 4.1, which illustrates the original Frequentist MLE, it is quite discernible that the magnitude of reduced bias lies in the $[0.1, 0.13]$. Furthermore, a significant reduction in covariance values also has occurred. It is noteworthy that those scenarios with the effect of 0.1 and equal success probabilities i.e., green circles and stars, tend to constantly be among the worst case scenarios. The former is known for showing arm C as an inferior arm while the effect is just 0.1. Even for the simulation result in the figure 4.4 this is going to be the case. By this, we mean that although the augmented estimators considerably diminish bias values on arm C, the reduction rate is not in proportion to other scenarios with an arm D effect bigger than 0.1. Note that, in almost all trials presented in the figure 4.4, and for any given arm C, the lower bound of success probability estimation will be $\theta_C - 0.1$ which is negligible in practice.

4.4.3 Augmented HT in a non-Randomised RAR procedure

As we showed in section 4.2.3, the performance of the HT estimator in a non-randomised RAR trial because of extreme randomised probabilities, as well as division by trial size, is worse than MLE. To compensate for this, and using initial fixed allocations i.e., the first time epoch allocates to control arm C and the second one to research arm D, we define the augmentation for a given arm as follows:

- If the first allocation is a success, then we add some pseudo success observations to the sample size of the other arm to the numerator of (4.14) at the end of the trial i.e., $\dot{s}_k(T) = n_{k'}(T)$.

$$\widetilde{\theta}_{k,HT} = \frac{s_k(T) + n_{k'}(T)}{T} \tag{4.20}$$

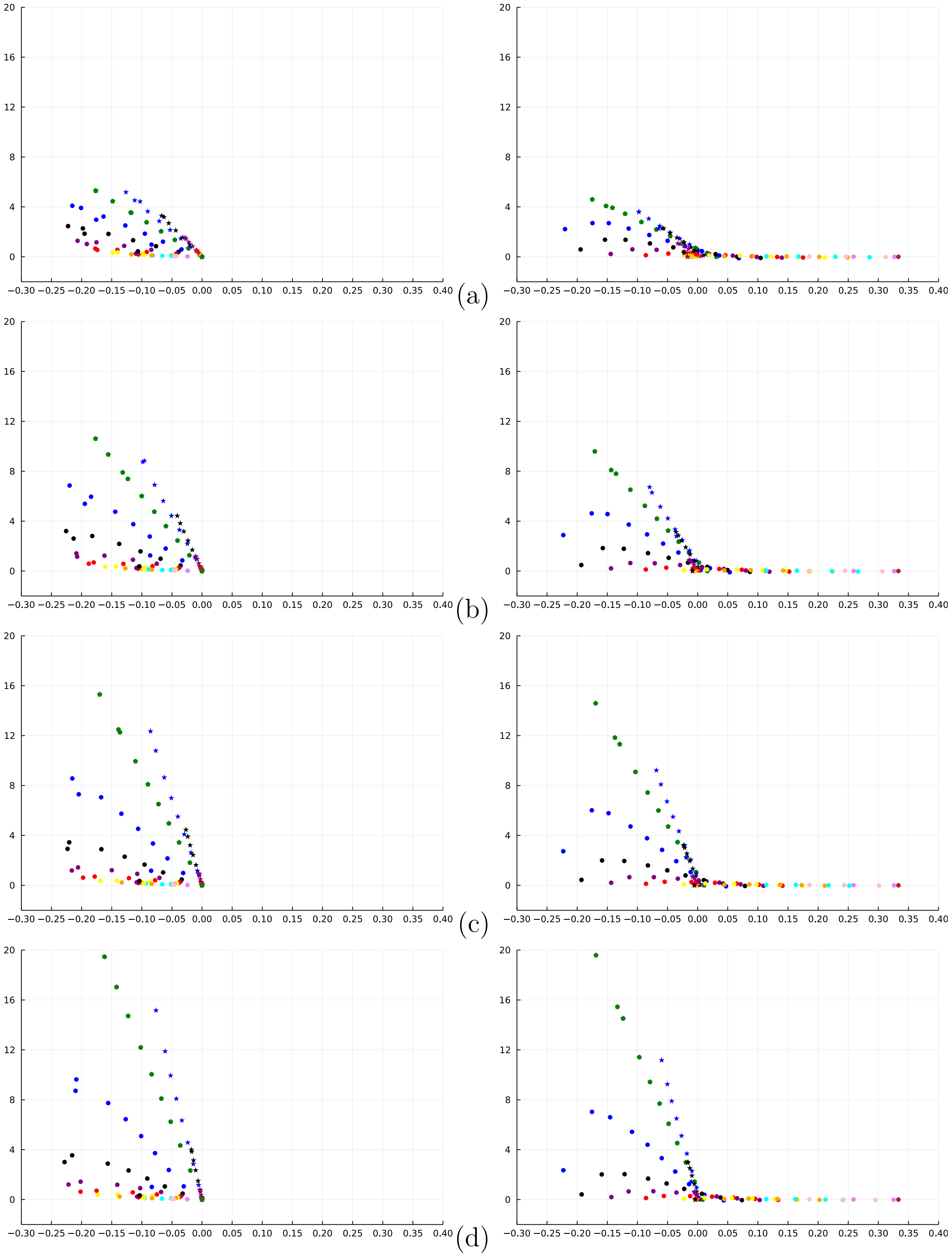


Figure 4.1: $\left(\dot{s}_k(T) = 0, \dot{f}_k(T) = 0, \widetilde{s}_k(0) = 0, \widetilde{f}_k(0) = 0\right)$ vs. $\left(\dot{s}_k(T) = 0, \dot{f}_k(T) = 0, \widetilde{s}_k(0) = 1, \widetilde{f}_k(0) = 1\right)$ i.e., $\Big($Frequentist MLE eq. (4.1)$\Big)$ vs. $\Big($Bayesian estimator eq. (4.3)$\Big)$: (a) T=60 (b) T=120 (c) T=180 (d) T=240. x-axis: Bias of estimator, y-axis: Covariance (Estimator, Sample Size).

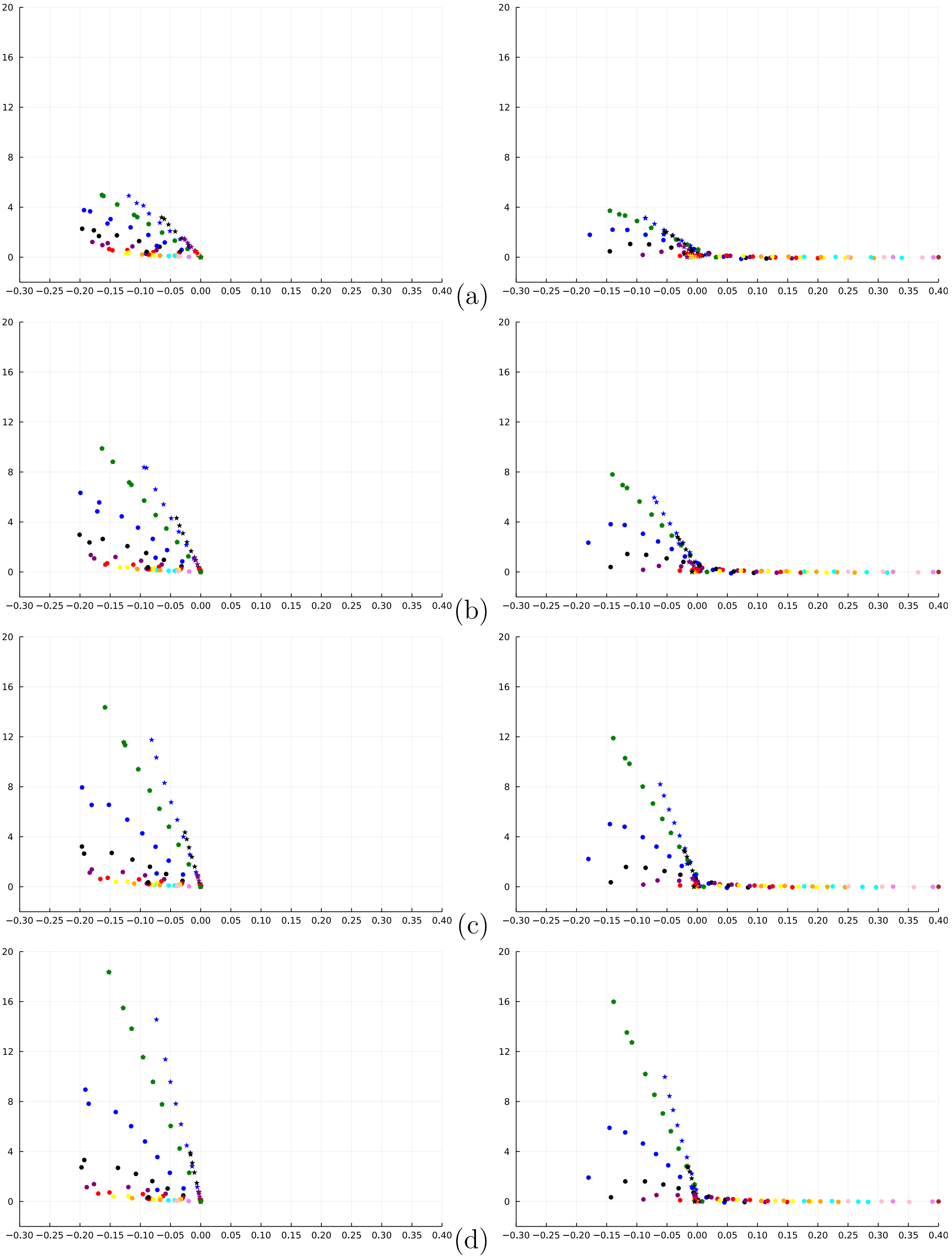


Figure 4.2: $\left(\dot{s}_k(T) = \widehat{\theta}_k(T), \dot{f}_k(T) = 0, \widetilde{s}_k(0) = 0, \widetilde{f}_k(0) = 0\right)$ vs. $\left(\dot{s}_k(T) = \widehat{\theta}_k(T), \dot{f}_k(T) = 0, \widetilde{s}_k(0) = 1, \widetilde{f}_k(0) = 1\right)$: (a) T=60 (b) T=120 (c) T=180 (d) T=240. x-axis: Bias of estimator, y-axis: Covariance (Estimator, Sample Size).

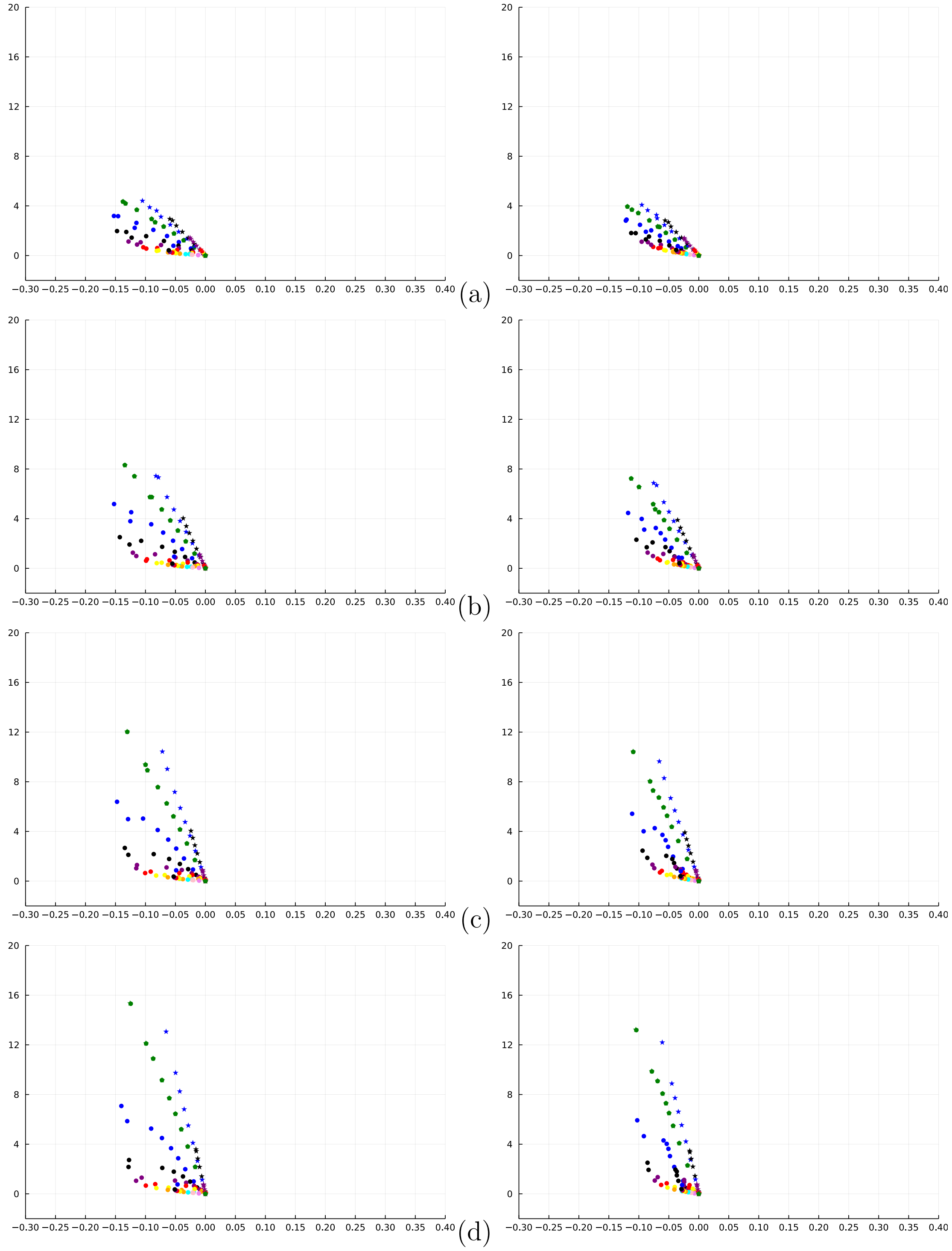


Figure 4.3: $\left(\dot{s}_k(T) = \ln(T+1)\widehat{\theta}_k(T), \dot{f}_k(T) = 0, \widetilde{s}_k(0) = 0, \widetilde{f}_k(0) = 0\right)$ vs. $\left(\dot{s}_k(T) = 4\ln(T+1)(\widehat{\theta}_k(T))^2, \dot{f}_k(T) = 0, \widetilde{s}_k(0) = 0, \widetilde{f}_k(0) = 0\right)$: (a) T=60 (b) T=120 (c) T=180 (d) T=240. x-axis: Bias of estimator, y-axis: Covariance (Estimator, Sample Size).

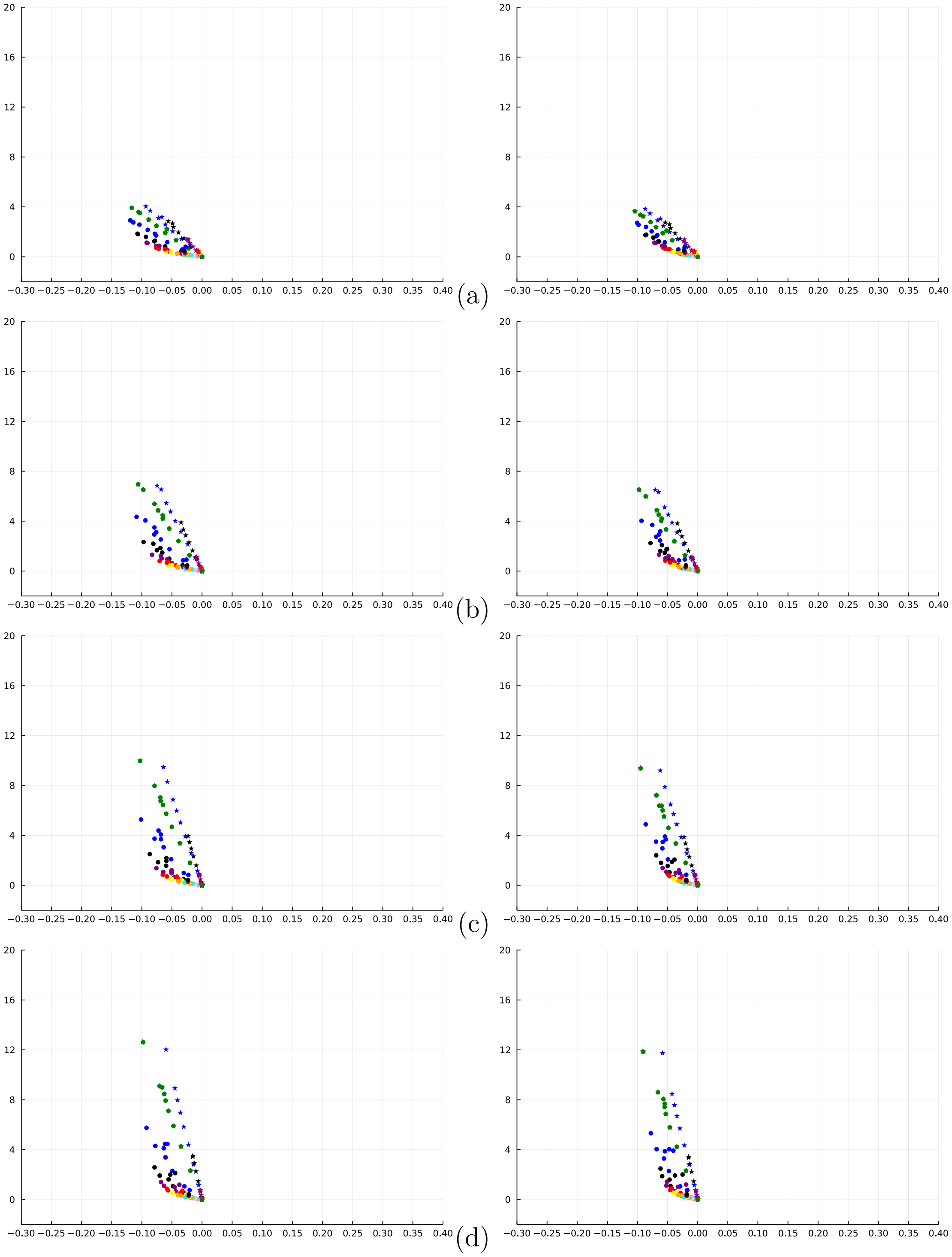


Figure 4.4: $\left(\dot{s}_k(T) = 9\ln(T+1)(\widehat{\theta}_k(T))^3, \dot{f}_k(T) = 0, \widetilde{s}_k(0) = 0, \widetilde{f}_k(0) = 0\right)$ vs. $\left(\dot{s}_k(T) = 27\sqrt{\ln(T+1)}(\widehat{\theta}_k(T))^3, \dot{f}_k(T) = 0, \widetilde{s}_k(0) = 0, \widetilde{f}_k(0) = 0\right)$: (a) T=60 (b) T=120 (c) T=180 (d) T=240. x-axis: Bias of estimator, y-axis: Covariance (Estimator, Sample Size).

- If the first allocation is a failure, then we keep the estimator as in (4.14).

$$\widetilde{\theta}_{k,HT} = \widehat{\theta}_{k,HT} = \frac{s_k(T)}{T} \tag{4.21}$$

In other words, for those simulation replications where the first allocation of a given arm is a success, it is also counted as a success response up to the sample size of the other arm. One might interpret it as a circumstance in which we complete the lack of observations on a given arm by the number of times when the other arm is being pulled. Additionally, it can be easily proven that augmented HT is asymptotically consistent. Recalling the proof procedure used for theorem 4.8, it follows that:

$$\begin{aligned} \lim_{T\to\infty} \left(\widetilde{\theta}_{k,HT} - \widehat{\theta}_{k,HT}\right) &= \lim_{T\to\infty} \left(\frac{s_k(T) + n_{k'}(T)}{T} - \widehat{\theta}_{k,HT}\right) \\ &= \lim_{T\to\infty} \left(\frac{n_{k'}(T)}{T}\right) = 0 \end{aligned} \tag{4.22}$$

The right-hand-side column of figure 4.5 represents augmented HT estimators for the different trial sizes. It is quite obvious that both bias and covariance values are notably reduced. For the trial size 240, estimation bias is reduces by -0.8 whilst it is -0.65 for $T = 60$ approximately, since the augmentation in the former is larger due to the bigger assumed trial size. Moreover, as a general trend that represents the relationship between covariance and bias values, one can see that less biased estimations are associated with higher covariance values as the trial size increases. Finally, the augmented HT estimator is biased no more than -0.07 in any trial size.

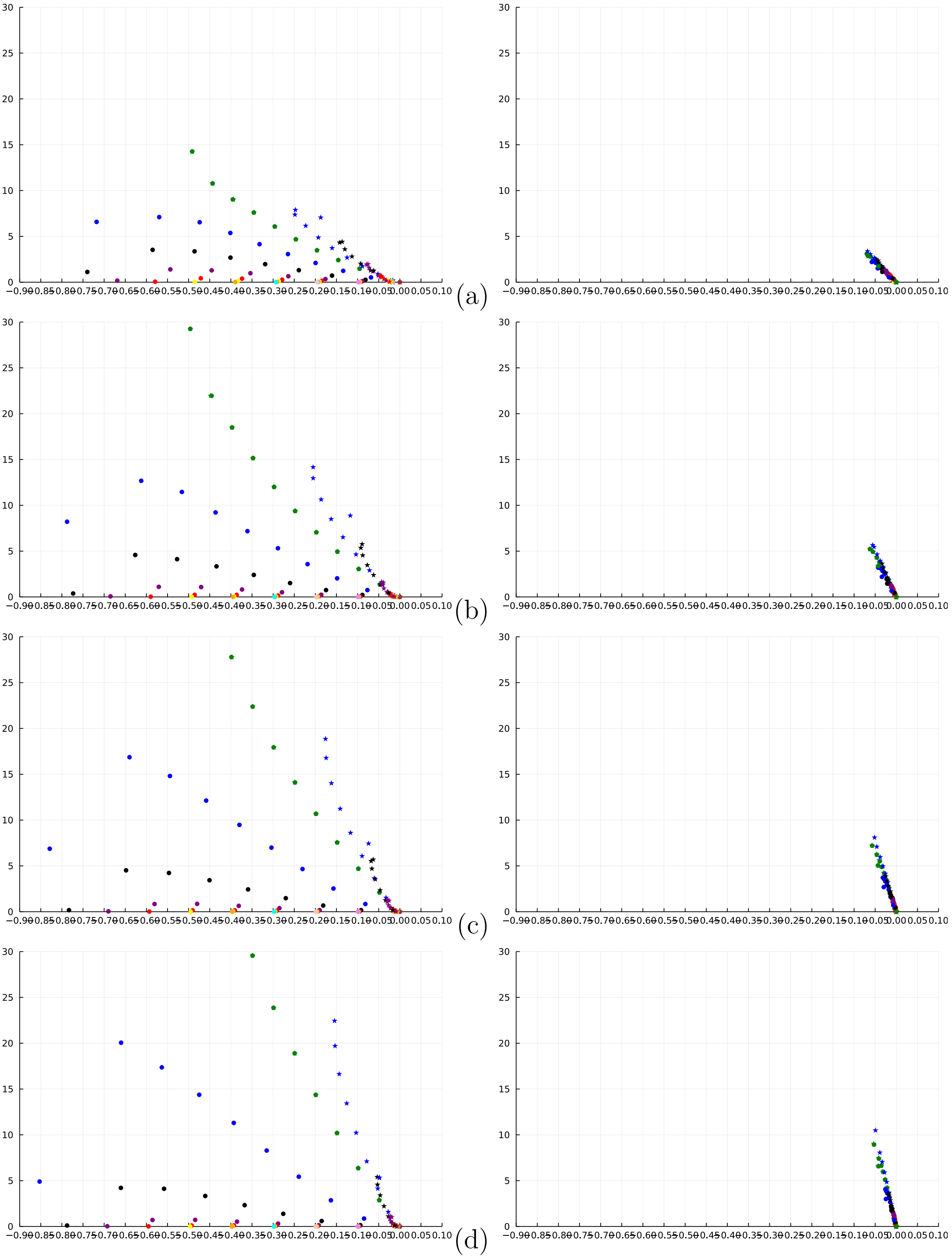


Figure 4.5: HT (IPW) estimator vs augmented equivalent: (a) T=60 (b) T=120 (c) T=180 (d) T=240. x-axis: Bias of estimator, y-axis: Covariance (Estimator, Sample Size).

4.5 Discussion

In this chapter, using DP designs, we evaluate different modifications for both Frequentist and Bayesian MLE estimators. First of all, we formulate the DP using the novel unified terminologies proposed by Jacko (2019b). Then we propose the augmented estimator with which the derived bias of MLE can be mitigated in a considerable manner. Furthermore, to check the performance of this novel estimator, we conduct extensive simulation studies in the largest possible range of scenarios for both Bayes and Jeffreys priors. Those with Jeffreys prior can be found in the Appendix 4.6.4. In addition, we also prove that DP as a selection function fulfils the so-called *Exploit* property proposed by Nie et al. (2018).

Using the augmented estimator, the magnitude of the reduced bias of MLE can lie in the interval $[0.1, 0.13]$ in this study. Although our proposed estimator performs very well in the Frequentist context, this is not the case of the Bayesian setting, since the bias of the Bayesian estimator is more complicated to capture and control. As an area of further work, we recommend applying our augmented estimator to previous adaptive designs and estimation processes to modify the derived bias as much as possible.

In this study, we try to find the optimal augmentation which can eradicate the bias from estimations totally. However, due to the complexity and data dependency this has not been achieved completely, while we noticed that in order to create an augmentation $\dot{s}_k(T)$, one had better consider an odd exponent for $\widehat{\theta}_k(T)$ together with a well-chosen coefficient. Hence, another area of future work can be formulating $\dot{s}_k(T)$ in an optimal manner by which an unbiased estimation can be obtained. To do so, it will be necessary to carefully analyse the behaviour of the

MLE along with the effects of time horizons in choosing the augmentation.

Appendix 4.6.3 consists of simulation results obtained from two not suitable modifications. Figure 4.5 (left-hand side column) shows a situation in which augmentations are identically added to both arms at the end of the trial. It is clear, the estimation ended up with several positive bias values in some scenarios while quite a few negative bias observations are still there. Note that, negative bias values, in this case, are even larger than those in Figure 4.4 (both sides). On the other hand, Figure 4.5 (right) shows an overshooting in the positive quadrant which mainly happens due to assuming a large coefficient for $\dot{s}_k(T)$.

One can also consider expanding the practical framework of covariate-adaptive RAR design offered by Ji et al. (2019) in which randomisation probabilities depend on both covariates and patients response as future work. Note that, in the commentary by Saville and Meurer (2019), some objections associated with the study of Ji et al. (2019) have been raised. The foremost ones which can be served as a motivation for future studies, are the difficulty of extrapolating and poor understanding of statistical properties such as power, bias etc related to the proposed framework, to a multi-armed context. Another practical and useful problem for future study can be obtained by considering time-trend with or without early stopping assumptions in a multi-armed bandit setting. The details of two-armed case have been investigated in the work of Jiang et al. (2020).

Practical implications of this study are countless. Generally, one can take advantage of the present study in any decision making context. For instance, clinical trials where the efficacy of an experimental treatment arm is being estimated within a trial and compared with the control treatment arm is one of the practical implications. As a rule in clinical trials settings, the treatment with the smallest true efficacy shows the largest bias, and this bias grows as the difference between the superior and inferior treatment increases (Bowden and Trippa, 2017). Benefiting from our proposed estimator, the bias can be substantially reduced and the treatment efficacy can be estimated more accurately. Digital marketing and social networks are among the other practical settings in which implications of this paper might be the case. For example, the augmented estimators may give the decision maker more profound insight into the popularity of a post or product on Instagram when people who follow the page like or dislike the post.

In summary, the RAR design is useful in many settings to improve the overall response in the trial. Although the arm effect estimation is usually associated with relatively small and negative bias, we propose a framework with which the bias is mitigated significantly. In order to have minimal bias, more studies need to be carried out to adjust the augmentation appropriately. Note that, in this paper we only focus on DP designs while one can apply the proposed contributions in other randomisation procedures to improve the performance of the estimators.

4.6 Appendix

4.6.1 DP satisfies ***Exploit***

Nie et al. (2018) prove that, in adaptive data collection context, for any selection function which satisfies natural conditions i.e., *Exploit* and *Independence of Irrelevant Option (IIO)* sample means of the data have negative biases. Since we consider a two-armed Bayesian Beta-Bernoulli model described in section (3.1), *(IIO)* conditions need not to be taken into account Nie et al. (2018), while we show that DP, which plays the role of selection function, satisfies *Exploit*. According to Nie et al. (2018) "*Exploit* means that for any given time epoch t if an arm is selected in a scenario in which it has lower sample average, then it would also be selected in a scenario where it has higher sample average."

Since $s_C + f_C + s_D + f_D = t$ represents the history of the trial, for arm C for instance, we can define two histories with same length and different combinations of s_C and f_C. To do so, we suppose $s_C + f_C = s^{'}_C + f^{'}_C$ such that $s_C \leq s^{'}_C$ and $f_C \geq f^{'}_C$ then sample means $\frac{s_C}{s_C+f_C} \leq \frac{s^{'}_C}{s^{'}_C+f^{'}_C}$. Furthermore, due to the fact that prior beliefs are among non-negative pseudo-observations, the monotonicity of current sample means holds in the Bayesian sense as well:

$$\frac{s_C}{s_C + f_C} \leq \frac{s^{'}_C}{s^{'}_C + f^{'}_C} \Leftrightarrow \frac{s_C + \widetilde{s_C}(0)}{s_C + \widetilde{s_C}(0) + f_C + \widetilde{f_C}(0)} \leq \frac{s^{'}_C + \widetilde{s_C}(0)}{s^{'}_C + \widetilde{s_C}(0) + f^{'}_C + \widetilde{f_C}(0)} \Leftrightarrow$$
$$\frac{\widetilde{s_C}}{\widetilde{s_C} + \widetilde{f_C}} \leq \frac{\widetilde{s^{'}_C}}{\widetilde{s^{'}_C} + \widetilde{f^{'}_C}} \Leftrightarrow q_{C,(x,i),1} \leq q^{'}_{C,(x,i),1} \tag{4.23}$$

Let $F_t(s_C, f_C, s_D, f_D) = max\Big\{F^C_t(s_C, f_C, s_D, f_D), F^D_t(s_C, f_C, s_D, f_D)\Big\}$ be the value function determining the maximum Bayes-expected number of successes after the time epoch t towards the end of the trial when the history is $\Big(s_C, f_C, s_D, f_D\Big)$. Without loss of generality if we assume that arm C is being selected to allocate at any time epoch, t i.e., $p^a_C = 1$, then a quite similar proof can be applied for arm D as well. Note that in the case of equality in the equation (4.23) i.e., $q_{C,(x,i),1} = q^{'}_{C,(x,i),1}$, DP always satisfies *Exploit*, and therefore, all proofs below will be trivial. Now, considering the backward induction algorithm together with the value functions $F^C_t\Big(s_C, f_C, s_D, f_D\Big)$ and $F^{'C}_t\Big(s^{'}_C, f^{'}_C, s_D, f_D\Big)$ corresponding to the histories $\Big(s_C, f_C, s_D, f_D\Big)$ and $\Big(s^{'}_C, f^{'}_C, s_D, f_D\Big)$ of arm C respectively:

- If $t = T$, we are at the end of the trial and there is nothing to do. Thus, $F_T\Big(s_C, f_C, s_D, f_D\Big) = 0$.

- If $t = T - 1$, there is only one state left, which needs to be allocated. Hence, the value functions, which are expected one-period rewards, can be formulated as follow:

$$\begin{aligned} F_{T-1}\Big(s_C, f_C, s_D, f_D\Big) &= max\Big\{F^C_{T-1}\Big(s_C, f_C, s_D, f_D\Big), F^D_{T-1}\Big(s_C, f_C, s_D, f_D\Big)\Big\} \\ &= F^C_{T-1}\Big(s_C, f_C, s_D, f_D\Big) \\ &= q_{C,(x,i),1} \end{aligned} \tag{4.24}$$

Similarly, $F^{'C}_{T-1}\Big(s^{'}_C, f^{'}_C, s_D, f_D\Big) = q^{'}_{C,(x,i),1}$.

According to (4.23), since $q_{C,(x,i),1} \leq q^{'}_{C,(x,i),1}$, then $F^C_{T-1}(s_C, f_C, s_D, f_D) \leq F^{'C}_{T-1}(s^{'}_C, f^{'}_C, s_D, f_D)$. Hence, we have:

$$\begin{aligned} F_{T-1}\Big(s_C, f_C, s_D, f_D\Big) &= max\Big\{F^{'C}_{T-1}\Big(s^{'}_C, f^{'}_C, s_D, f_D\Big), F^D_{T-1}\Big(s_C, f_C, s_D, f_D\Big)\Big\} \\ &= F^{'C}_{T-1}\Big(s^{'}_C, f^{'}_C, s_D, f_D\Big) \end{aligned} \tag{4.25}$$

- et cetera.

First of all, at any given time epoch t, the value functions can be expressed in general form as follows:

$$\begin{aligned}F_t^C\Big(s_C, f_C, s_D, f_D\Big) =& q_{C,(x,i),1}\cdot\Big(1 + F_{t+1}\Big(s_C + 1, f_C, s_D, f_D\Big)\Big)\\ &+ q_{C,(x,i),0}\cdot\Big(0 + F_{t+1}\Big(s_C, f_C + 1, s_D, f_D\Big)\Big)\\ F_t^D\Big(s_C, f_C, s_D, f_D\Big) =& q_{D,(x,i),1}\cdot\Big(1 + F_{t+1}\Big(s_C, f_C, s_D + 1, f_D\Big)\Big)\\ &+ q_{D,(x,i),0}\cdot\Big(0 + F_{t+1}\Big(s_C, f_C, s_D, f_D + 1\Big)\Big)\end{aligned}$$

according to our primary assumption $F_t(s_C, f_C, s_D, f_D) = F_t^C(s_C, f_C, s_D, f_D)$. Second, we presume two conditions: $s_C^{'} = s_C + \hat{s}$ and $f_C^{'} = f_C - \hat{s}$ for the history $\Big(s_C^{'}, f_C^{'}, s_D, f_D\Big)$ in which $\hat{s} \geq 1$. Then, the posterior probability, or in other words the current belief of success for arm C can be written as follows:

$$\begin{aligned}q_{C,(x,i),1}^{'} =& \frac{\widetilde{s_C^{'}}}{\widetilde{s_C^{'}} + \widetilde{f_C^{'}}} = \frac{\widetilde{s_C} + \hat{s}}{\widetilde{s_C^{'}} + \widetilde{f_C^{'}}} = \frac{\widetilde{s_C}}{\widetilde{s_C^{'}} + \widetilde{f_C^{'}}} + \frac{\hat{s}}{\widetilde{s_C^{'}} + \widetilde{f_C^{'}}}\\ =& q_{C,(x,i),1} + \frac{\hat{s}}{\widetilde{s_C^{'}} + \widetilde{f_C^{'}}}\end{aligned} \tag{4.26}$$

In turn, the value functions corresponding to the histories $\Big(s_C, f_C, s_D, f_D\Big)$ and $\Big(s_C^{'}, f_C^{'}, s_D, f_D\Big)$ of arm C by applying the equation (4.26) can be written as follows:

$$\begin{aligned}F_t^C\Big(s_C, f_C, s_D, f_D\Big) =& q_{C,(x,i),1}\cdot\Big(1 + F_{t+1}\Big(s_C + 1, f_C, s_D, f_D\Big)\Big)\\ &+ q_{C,(x,i),0}\cdot\Big(0 + F_{t+1}\Big(s_C, f_C + 1, s_D, f_D\Big)\Big)\end{aligned} \tag{4.27}$$

$$\begin{aligned}
&F_t^{'C}\Big(s_C^{'}, f_C^{'}, s_D, f_D\Big)\\
&= q_{C,(x,i),1}^{'} \cdot \Big(1 + F_{t+1}\Big(s_C^{'} + 1, f_C^{'}, s_D, f_D\Big)\Big)\\
&+ q_{C,(x,i),0}^{'} \cdot \Big(0 + F_{t+1}\Big(s_C^{'}, f_C^{'} + 1, s_D, f_D\Big)\Big)\\
&= \Big(q_{C,(x,i),1} + \frac{\hat{s}}{\widetilde{s_C^{'}} + \widetilde{f_C^{'}}}\Big) . \Big(1 + F_{t+1}\Big(s_C + \hat{s} + 1, f_C - \hat{s}, s_D, f_D\Big)\Big)\\
&+ \Big(q_{C,(x,i),0} - \frac{\hat{s}}{\widetilde{s_C^{'}} + \widetilde{f_C^{'}}}\Big) . \Big(0 + F_{t+1}\Big(s_C + \hat{s}, f_C - \hat{s} + 1, s_D, f_D\Big)\Big)
\end{aligned} \tag{4.28}$$

If we show that $F_t^C(s_C, f_C, s_D, f_D) < F_t^{'C}(s_C^{'}, f_C^{'}, s_D, f_D)$, then one can simply conclude that:

$$\begin{aligned}
F_t\Big(s_C, f_C, s_D, f_D\Big) =& max\Big\{F_t^{'C}\Big(s_C^{'}, f_C^{'}, s_D, f_D\Big), F_t^D\Big(s_C, f_C, s_D, f_D\Big)\Big\}\\
=& F_t^{'C}\Big(s_C^{'}, f_C^{'}, s_D, f_D\Big)
\end{aligned}$$

which satisfies the *Exploit* for this case. To do so, we rearrange equations above in order to have a convex combination of F_{t+1}s in the right hand sides:

$$\begin{aligned}
F_t^C\Big(s_C, f_C, s_D, f_D\Big) - q_{C,(x,i),1} = q_{C,(x,i),1} . F_{t+1}\Big(s_C + 1, f_C, s_D, f_D\Big)\\
+ q_{C,(x,i),0} . F_{t+1}\Big(s_C, f_C + 1, s_D, f_D\Big)
\end{aligned} \tag{4.29}$$

$$
\begin{aligned}
F_t^{'C}\Big(s_C^{'}, f_C^{'}, s_D, f_D\Big) - q_{C,(x,i),1}& \\
&- \frac{\hat{s}}{\widetilde{s_C^{'}} + \widetilde{f_C^{'}}}.\Big[1 + F_{t+1}\Big(s_C + \hat{s} + 1, f_C - \hat{s}, s_D, f_D\Big) \\
&- F_{t+1}\Big(s_C + \hat{s}, f_C - \hat{s} + 1, s_D, f_D\Big)\Big] \\
&= q_{C,(x,i),1}.F_{t+1}\Big(s_C + \hat{s} + 1, f_C - \hat{s}, s_D, f_D\Big) \\
&+ q_{C,(x,i),0}.F_{t+1}\Big(s_C + \hat{s}, f_C - \hat{s} + 1, s_D, f_D\Big)
\end{aligned}
\tag{4.30}
$$

Lemma 4.1. *For any given $\hat{s}$ in the recursion (4.30), it can be concluded that:*

$$
F_{t+1}\Big(s_C + \hat{s}, f_C - \hat{s} + 1, s_D, f_D\Big) < F_{t+1}\Big(s_C + \hat{s} + 1, f_C - \hat{s}, s_D, f_D\Big) \quad \forall \hat{s} \tag{4.31}
$$

Proof. Using Mathematical induction with an assumption for time epoch $t + 1$ such that:

$$
F_{t+1}\Big(s_C + \hat{s}, f_C - \hat{s} + 1, s_D, f_D\Big) < F_{t+1}\Big(s_C + \hat{s} + 1, f_C - \hat{s}, s_D, f_D\Big) \quad \forall \hat{s}
$$

and considering backward induction, we need to prove:

$$
F_t\Big(s_C + \hat{s}, f_C - \hat{s} + 1, s_D, f_D\Big) < F_t\Big(s_C + \hat{s} + 1, f_C - \hat{s}, s_D, f_D\Big) \quad \forall \hat{s}
$$

According to the value function definition, we have:

$$
\begin{aligned}
&F_t\Big(s_C + \hat{s}, f_C - \hat{s} + 1, s_D, f_D\Big) \\
&\quad = max\Big\{F_t^C\Big(s_C + \hat{s}, f_C - \hat{s} + 1, s_D, f_D\Big), F_t^D\Big(s_C + \hat{s}, f_C - \hat{s} + 1, s_D, f_D\Big)\Big\}
\end{aligned}
\tag{4.32}
$$

$$
\begin{aligned}
&F_t\Big(s_C+\hat{s}+1, f_C-\hat{s}, s_D, f_D\Big)\\
&= max\Big\{F_t^C\Big(s_C+\hat{s}+1, f_C-\hat{s}, s_D, f_D\Big), F_t^D\Big(s_C+\hat{s}+1, f_C-\hat{s}, s_D, f_D\Big)\Big\}
\end{aligned}
\tag{4.33}
$$

- Arm C comparison:

  Using definition, value functions corresponding to arm C in the equations (4.32) and (4.33) can be written as follows:

$$
\begin{aligned}
&F_t^C\Big(s_C+\hat{s}, f_C-\hat{s}+1, s_D, f_D\Big)\\
&= \frac{\widetilde{s_C}+\hat{s}}{\widetilde{s_C}+\widetilde{f_C}+1}.\\
&\Big(1+F_{t+1}\Big(s_C+\hat{s}+1, f_C-\hat{s}+1, s_D, f_D\Big)\Big)\\
&+\frac{\widetilde{f_C}-\hat{s}}{\widetilde{s_C}+\widetilde{f_C}+1}.\\
&\Big(0+F_{t+1}\Big(s_C+\hat{s}, f_C-\hat{s}+2, s_D, f_D\Big)\Big)\\
&F_t^C\Big(s_C+\hat{s}+1, f_C-\hat{s}, s_D, f_D\Big)\\
&= \frac{\widetilde{s_C}+\hat{s}+1}{\widetilde{s_C}+\widetilde{f_C}+1}.\\
&\Big(1+F_{t+1}\Big(s_C+\hat{s}+2, f_C-\hat{s}, s_D, f_D\Big)\Big)\\
&+\frac{\widetilde{f_C}-\hat{s}}{\widetilde{s_C}+\widetilde{f_C}+1}.\\
&\Big(0+F_{t+1}\Big(s_C+\hat{s}+1, f_C-\hat{s}+1, s_D, f_D\Big)\Big)
\end{aligned}
\tag{4.34}
$$

  Applying some arithmetic simplifications and properties of inequalities, namely: if $a<b$ and $c<d$ then $a+c<b+d$, together with the induction assumption for time epoch $t+1$, which mainly confirms that:

$$
F_{t+1}\Big(s_C+\hat{s}+1, f_C-\hat{s}+1, s_D, f_D\Big) < F_{t+1}\Big(s_C+\hat{s}+2, f_C-\hat{s}, s_D, f_D\Big)
$$

$$F_{t+1}\Big(s_C + \hat{s}, f_C - \hat{s} + 2, s_D, f_D\Big) < F_{t+1}\Big(s_C + \hat{s} + 1, f_C - \hat{s} + 1, s_D, f_D\Big)$$

it can be concluded that:

$$F_t^C\Big(s_C + \hat{s}, f_C - \hat{s} + 1, s_D, f_D\Big) < F_t^C\Big(s_C + \hat{s} + 1, f_C - \hat{s}, s_D, f_D\Big) \quad (4.35)$$

– Arm D comparison:

Similarly, we can also set up the value functions of arm D in the equations (4.32) and (4.33) as below:

$$\begin{aligned}
&F_t^D\Big(s_C + \hat{s}, f_C - \hat{s} + 1, s_D, f_D\Big) \\
&= q_{D,(x,i),1}.\Big(1 + F_{t+1}\Big(s_C + \hat{s}, f_C - \hat{s} + 1, s_D + 1, f_D\Big)\Big) \\
&+ q_{D,(x,i),0}.\Big(0 + F_{t+1}\Big(s_C + \hat{s}, f_C - \hat{s} + 1, s_D, f_D + 1\Big)\Big) \\
&F_t^D\Big(s_C + \hat{s} + 1, f_C - \hat{s}, s_D, f_D\Big) \\
&= q_{D,(x,i),1}.\Big(1 + F_{t+1}\Big(s_C + \hat{s} + 1, f_C - \hat{s}, s_D + 1, f_D\Big)\Big) \\
&+ q_{D,(x,i),0}.\Big(0 + F_{t+1}\Big(s_C + \hat{s} + 1, f_C - \hat{s}, s_D, f_D + 1\Big)\Big)
\end{aligned} \quad (4.36)$$

Due to the fact that the histories of arm D in both equations above are the same, so are the posterior probabilities i.e., $q_{D,(x,i),1}$ and $q_{D,(x,i),0}$ in the right hand sides. Recalling the induction assumption, we have:

$$F_{t+1}\Big(s_C + \hat{s}, f_C - \hat{s} + 1, s_D + 1, f_D\Big) < F_{t+1}\Big(s_C + \hat{s} + 1, f_C - \hat{s}, s_D + 1, f_D\Big)$$

$$F_{t+1}\Big(s_C + \hat{s}, f_C - \hat{s} + 1, s_D, f_D + 1\Big) < F_{t+1}\Big(s_C + \hat{s} + 1, f_C - \hat{s}, s_D, f_D + 1\Big)$$

which leads to us concluding that:

$$F_t^D\Big(s_C + \hat{s}, f_C - \hat{s} + 1, s_D, f_D\Big) < F_t^D\Big(s_C + \hat{s} + 1, f_C - \hat{s}, s_D, f_D\Big) \quad (4.37)$$

Finally, considering (4.35) and (4.37) results that:

$$max\Big\{F_t^C\Big(s_C+\hat{s}, f_C-\hat{s}+1, s_D, f_D\Big), F_t^D\Big(s_C+\hat{s}, f_C-\hat{s}+1, s_D, f_D\Big)\Big\}$$

$$<$$

$$max\Big\{F_t^C\Big(s_C+\hat{s}+1, f_C-\hat{s}, s_D, f_D\Big), F_t^D\Big(s_C+\hat{s}+1, f_C-\hat{s}, s_D, f_D\Big)\Big\}$$

and therefore, $F_t\Big(s_C+\hat{s}, f_C-\hat{s}+1, s_D, f_D\Big) < F_t\Big(s_C+\hat{s}+1, f_C-\hat{s}, s_D, f_D\Big) \quad \forall\hat{s}$, which confirms that the induction assumption is true. □

Now, recalling convex combination properties together with lemma 1, and considering the fact that $q_{C,(x,i),1}$ and $q_{C,(x,i),0}$ are the posterior probabilities (weights in weighted mean) in the equation (4.29) and (4.30), we have:

$$\begin{aligned} F_t^C\Big(s_C, f_C, s_D, f_D\Big) - q_{C,(x,i),1} &= q_{C,(x,i),1}.F_{t+1}\Big(s_C+1, f_C, s_D, f_D\Big) \\ &+ q_{C,(x,i),0}.F_{t+1}\Big(s_C, f_C+1, s_D, f_D\Big) \\ &< F_{t+1}\Big(s_C+1, f_C, s_D, f_D\Big) \end{aligned} \tag{4.38}$$

$$\begin{aligned} F_t^{'C}\Big(s_C^{'}, f_C^{'}, s_D, f_D\Big) &- q_{C,(x,i),1} \\ &- \frac{\hat{s}}{\widetilde{s_C^{'}}+\widetilde{f_C^{'}}}.\Big[1+F_{t+1}\Big(s_C+\hat{s}+1, f_C-\hat{s}, s_D, f_D\Big) \\ &- F_{t+1}\Big(s_C+\hat{s}, f_C-\hat{s}+1, s_D, f_D\Big)\Big] \\ &= q_{C,(x,i),1}.F_{t+1}\Big(s_C+\hat{s}+1, f_C-\hat{s}, s_D, f_D\Big) \\ &+ q_{C,(x,i),0}.F_{t+1}\Big(s_C+\hat{s}, f_C-\hat{s}+1, s_D, f_D\Big) \\ &> F_{t+1}\Big(s_C+\hat{s}, f_C-\hat{s}+1, s_D, f_D\Big) \end{aligned} \tag{4.39}$$

It is noteworthy to mention that for $\hat{s} = 1$ right hand sides of (4.38) and (4.39) will be the same, and thereafter for any given $\hat{s} > 1$, according to the lemma 1, we have:

$$\begin{aligned}
F_{t+1}\Big(s_C+1, f_C, s_D, f_D\Big) &< F_{t+1}\Big(s_C+\hat{s}, f_C-\hat{s}+1, s_D, f_D\Big)\\
F_t^C\Big(s_C, f_C, s_D, f_D\Big) - q_{C,(x,i),1} &< F_t^{'C}\Big(s_C^{'}, f_C^{'}, s_D, f_D\Big) - q_{C,(x,i),1}\\
&- \frac{\hat{s}}{\widetilde{s_C^{'}} + \widetilde{f_C^{'}}}.\\
&\Big[1 + F_{t+1}\Big(s_C+\hat{s}+1, f_C-\hat{s}, s_D, f_D\Big)\\
&- F_{t+1}\Big(s_C+\hat{s}, f_C-\hat{s}+1, s_D, f_D\Big)\Big]\\
F_t^C\Big(s_C, f_C, s_D, f_D\Big) &< F_t^{'C}\Big(s_C^{'}, f_C^{'}, s_D, f_D\Big) - \frac{\hat{s}}{\widetilde{s_C^{'}} + \widetilde{f_C^{'}}}.\Big[1+\delta\Big]
\end{aligned} \tag{4.40}$$

where δ is positive value quantifying the difference of value functions for two neighbour time epochs. Finally, the last line of (4.40) implies that:

$$F_t^C\Big(s_C, f_C, s_D, f_D\Big) < F_t^{'C}\Big(s_C^{'}, f_C^{'}, s_D, f_D\Big).$$

4.6.2 Proofs of the main results

Proof. of Theorem (4.1): Let $Y_k(t)$ represent the random variable (and $y_k(t)$ the realization) of the response at time epoch $t \in \mathcal{T}$ corresponding to arm $k \in \mathcal{K}$. Denote by $k(t) \in \mathcal{K}$ the actual assigned arm at each time epoch t. The $2 \times t$ matrix of the total information in the trial before time epoch t is as follows:

$$\mathbf{D}(t-1) = \begin{pmatrix} k(0) & ... & k(t-1) \\ y_{k(0)}(0) & ... & y_{k(t-1)}(t-1) \end{pmatrix} \tag{4.41}$$

This information matrix keeps the record of allocations and responses from all the previous time epochs, and thus plays a significant role in the allocation at time epoch t. Consider that the responses follow a distribution with parameter θ_k. After allocating and observing the response of time epoch τ, the likelihood function for the parameter vector $\mathbf{\Theta} = (\theta_k)_{k\in\mathcal{K}}$ can be factorised as follows:

$$L(\mathbf{\Theta}|\mathbf{D}(\tau)) = \prod_{t=0}^{\tau} f_{\mathbf{\Theta}}\Big(y_{k(t)}(t)|k(t)\Big) f\Big(k(t)|\mathbf{D}(t-1)\Big) \tag{4.42}$$

Term $f\Big(k(t)|\mathbf{D}(t-1)\Big)$ is the probability of allocating the subject to arm $k(t)$ at time epoch t conditional on previous information $\mathbf{D}(t-1)$. According to the standard likelihood theory (cf. Boos and Stefanski, 2013), and the fact that this term does not depend on $\mathbf{\Theta}$ since it is a known function of data, we can write

$$L(\mathbf{\Theta}|\mathbf{D}(\tau)) \propto \prod_{t=0}^{\tau} f_{\mathbf{\Theta}}\Big(y_{k(t)}(t)|k(t)\Big). \tag{4.43}$$

Assuming that responses are independently Bernoulli distributed, the likelihood function can be written as

$$L(\mathbf{\Theta}|\mathbf{D}(\tau)) \propto \prod_{k\in\mathcal{K}} \theta_k^{s_k(\tau)}(1-\theta_k)^{f_k(\tau)}. \tag{4.44}$$

It is well-known that if $(X_k)_{k\in\mathcal{K}}$ are independent (not identically distributed) random variables each having a Beta distribution with parameters α_k, β_k, then the joint likelihood function for these observations is

$$L(\boldsymbol{\alpha},\boldsymbol{\beta}|\boldsymbol{X}) \propto \prod_{k\in\mathcal{K}} X_k^{\alpha_k-1}(1-X_k)^{\beta_k-1}. \tag{4.45}$$

The right-hand sides of the two proportionality-equations are equal when $\alpha_k = s_k(\tau)+1$ and $\beta_k = f_k(\tau)+1$, therefore the left-hand sides are proportional to each other. $\square$

Proof. of Theorem (4.2): Applying the natural logarithm to obtain the log-likelihood function,

$$l\Big(\mathbf{\Theta}|\mathbf{D}(\tau)\Big) := \ln\Big(L\Big(\mathbf{\Theta}|\mathbf{D}(\tau)\Big)\Big) \propto \sum_{k\in\mathcal{K}}\Big[s_k(\tau)\ln(\theta_k)+f_k(\tau)\ln(1-\theta_k)\Big] \tag{4.46}$$

the critical points for $s_k(\tau)+f_k(\tau)>0$ can be found using partial differentiation with respect to θ_k :

$$\begin{aligned}
\nabla\Big(l\Big(\mathbf{\Theta}|\mathbf{D}(\tau)\Big)\Big) = 0 \Rightarrow & \sum_{k\in\mathcal{K}} \frac{\partial\Big[s_k(\tau)\ln(\theta_k)+f_k(\tau)\ln(1-\theta_k)\Big]}{\partial\theta_k} = 0 \\
\Rightarrow & \sum_{k\in\mathcal{K}}\Bigg[\frac{\partial\Big(s_k(\tau)\ln(\theta_k)\Big)}{\partial\theta_k}+\frac{\partial\Big(f_k(\tau)\ln(1-\theta_k)\Big)}{\partial\theta_k}\Bigg] = 0 \\
\Rightarrow & \frac{s_k(\tau)}{\theta_k}-\frac{f_k(\tau)}{(1-\theta_k)} = 0 \quad \forall k\in\mathcal{K} \\
\Rightarrow & \theta_k = \frac{s_k(\tau)}{s_k(\tau)+f_k(\tau)} \quad \forall k\in\mathcal{K}
\end{aligned} \tag{4.47}$$

which for $s_k(\tau)=f_k(\tau)=0$ is satisfied by any value in $(0,1)$, and for $s_k(\tau)+f_k(\tau)>0$ can be rearranged as $\frac{s_k(\tau)}{s_k(\tau)+f_k(\tau)}$. Due to the fact that the corresponding Hessian matrix $\mathbf{H}$, which is a diagonal matrix of size $\mathcal{K}$, is negative-definite at θ_k, it can be shown θ_k is maxima $\forall k \in \mathcal{K}$:

$$\mathbf{H}\Big(l(\mathbf{\Theta}|\mathbf{D}(\tau))\Big) = \begin{pmatrix} -\Big(\frac{s_1(\tau)}{{\theta_1}^2} + \frac{f_1(\tau)}{(1-\theta_1)^2}\Big) & 0 & \cdots & 0 \\ 0 & -\Big(\frac{s_2(\tau)}{{\theta_2}^2} + \frac{f_2(\tau)}{(1-\theta_2)^2}\Big) & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & -\Big(\frac{s_{\mathcal{K}}(\tau)}{{\theta_{\mathcal{K}}}^2} + \frac{f_{\mathcal{K}}(\tau)}{(1-\theta_{\mathcal{K}})^2}\Big) \end{pmatrix}$$

□

Proof. of Theorem (4.3): Note that the mode of the Beta distribution is

$$\begin{cases} \dfrac{\alpha-1}{\alpha+\beta-2} & \text{for } \alpha,\beta > 1 \\ \text{any value in } (0,1) & \text{for } \alpha=\beta=1 \\ \{0,1\} & \text{for } \alpha,\beta < 1 \\ 0 & \text{for } \alpha \leq 1, \beta > 1 \\ 1 & \text{for } \alpha > 1, \beta \leq 1 \end{cases} \tag{4.48}$$

and thus the equivalence holds under our assumption of $s_k(\tau), f_k(\tau) \geq 0$, which corresponds to $\alpha, \beta \geq 1$. □

Proof. of Theorem (4.9): We start by using the definition of the covariance of two random variables, $\mathrm{Cov}[X,Y] = \mathrm{E}[XY] - \mathrm{E}[X]\mathrm{E}[Y]$, to obtain

$$\mathrm{Cov}\Big[n_k(\tau) + \tilde{n}_k(0) + \dot{n}(\tau), \widetilde{\theta}_k(\tau)\Big] \tag{4.49}$$

$$= \mathrm{E}\Big[\Big(n_k(\tau) + \tilde{n}_k(0) + \dot{n}(\tau)\Big)\widetilde{\theta}_k(\tau)\Big] - \mathrm{E}\Big[n_k(\tau) + \tilde{n}_k(0) + \dot{n}(\tau)\Big]\mathrm{E}\Big[\widetilde{\theta}_k(\tau)\Big] \tag{4.50}$$

which after extracting $\mathrm{E}\Big[\widetilde{\theta}_k(\tau)\Big]$ gives

$$\mathrm{E}[\widetilde{\theta}_k(\tau)] = \frac{-\mathrm{Cov}[n_k(\tau)+\widetilde{n}_k(0)+\dot{n}(\tau),\widetilde{\theta}_k(\tau)] + \mathrm{E}\Big[(n_k(\tau)+\widetilde{n}_k(0)+\dot{n}(\tau))\widetilde{\theta}_k(\tau)\Big]}{\mathrm{E}[n_k(\tau)+\widetilde{n}_k(0)+\dot{n}(\tau)]} \tag{4.51}$$

Subtracting θ_k from both sides and using the definition of the bias gives

$$\begin{aligned}&\mathrm{Bias}\Big[\widetilde{\theta}_k(\tau)\Big]\\ &= \frac{-\mathrm{Cov}\Big[n_k(\tau)+\widetilde{n}_k(0)+\dot{n}(\tau),\widetilde{\theta}_k(\tau)\Big] + \mathrm{E}\Big[\Big(n_k(\tau)+\widetilde{n}_k(0)+\dot{n}(\tau)\Big)(\widetilde{\theta}_k(\tau)-\theta_k)\Big]}{\mathrm{E}\Big[n_k(\tau)+\widetilde{n}_k(0)+\dot{n}(\tau)\Big]}\end{aligned} \tag{4.52}$$

We will now continue with simplification of the second term of the numerator,

$$\begin{aligned}&\mathrm{E}\Big[\Big(n_k(\tau)+\widetilde{n}_k(0)+\dot{n}(\tau)\Big)(\widetilde{\theta}_k(\tau)-\theta_k)\Big]\\ &= \mathrm{E}\Big[\Big(n_k(\tau)+\widetilde{n}_k(0)+\dot{n}(\tau)\Big)\widetilde{\theta}_k(\tau)\Big] - \mathrm{E}\Big[\Big(n_k(\tau)+\widetilde{n}_k(0)+\dot{n}(\tau)\Big)\theta_k\Big]\\ &= \mathrm{E}\Big[\Big(n_k(\tau)+\widetilde{n}_k(0)+\dot{n}(\tau)\Big)\frac{s_k(\tau)+\widetilde{s}_k(0)+\dot{s}(\tau)}{n_k(\tau)+\widetilde{n}_k(0)+\dot{n}(\tau)}\Big]\\ &- \mathrm{E}\Big[\Big(n_k(\tau)+\widetilde{n}_k(0)+\dot{n}(\tau)\Big)\theta_k\Big]\\ &= \mathrm{E}\Big[s_k(\tau)-n_k(\tau)\theta_k\Big] + \widetilde{s}_k(0)-\widetilde{n}_k(0)\theta_k+\dot{s}(\tau)-\dot{n}(\tau)\theta_k\\ &= \widetilde{n}_k(0)\Big(\widetilde{\theta}_k(0)-\theta_k\Big)+\dot{n}_k(\tau)\Big(\dot{\theta}_k(\tau)-\theta_k\Big)\end{aligned} \tag{4.53}$$

which finally gives (4.18). □

4.6.3 Both-arms modification and overshooting

This section consists of simulation results obtained from two inappropriate modifications. The left-hand side represents a situation in which augmentations are applied to both arms apart from being inferior or superior arm. The right-hand side denotes an overshooting in positive axis which mainly happens due to assuming 243 in $\dot{s}_k(T)$.

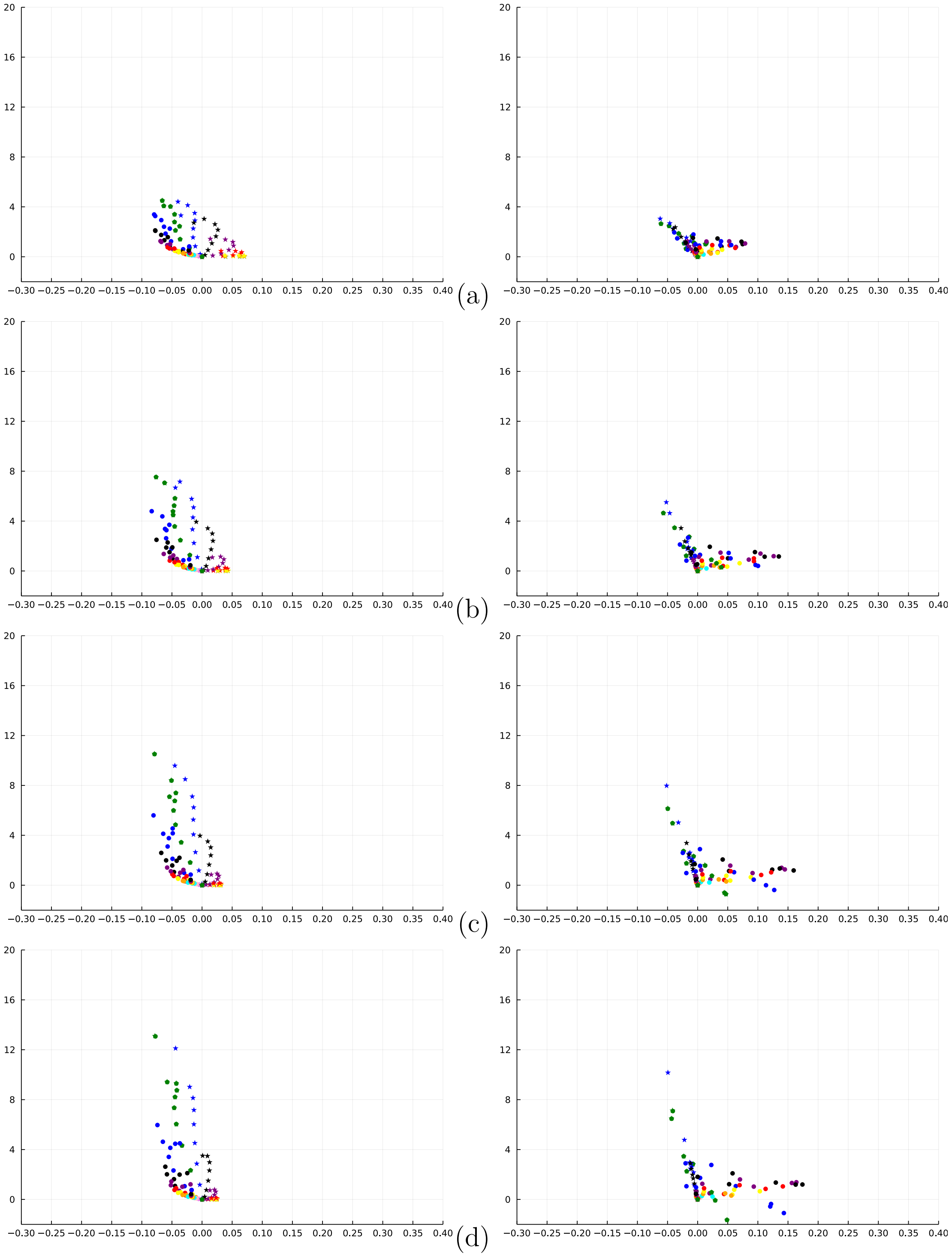


Figure 4.6: Left-hand side: Applying augmentations on both arms $\left(\dot{s}_k(T) = 27\sqrt{\ln(T+1)}(\widehat{\theta}_k(T))^3, \dot{f}_k(T) = 0, \widetilde{s}_k(0) = 0, \widetilde{f}_k(0) = 0\right)$, and right-hand side: Overshooting by too large constant $\left(\dot{s}_k(T) = 243\sqrt{\ln(T+1)}(\widehat{\theta}_k(T))^3, \dot{f}_k(T) = 0, \widetilde{s}_k(0) = 0, \widetilde{f}_k(0) = 0\right)$: (a) T=60 (b) T=120 (c) T=180 (d) T=240. x-axis: Bias of estimator, y-axis: Covariance (Estimator, Sample Size).

4.6.4 Results with Jeffreys prior

In this section, simulation results corresponding to Jeffreys prior are presented. Note that, Jeffreys prior is used for both DP and efficiency estimation process described in section (4.2). In this case, due to the fact that Jeffreys prior performs better than the Bayes counterpart, it is quite obvious that results are usually associated with smaller negative bias values. Furthermore, table (4.2) shows the assumed augmentation information in simulation set-up.

		$\dot{s}_k(T)$	Prior belief	Statistics
Fig. 4.7	RHS	0	Jeffreys	Bayesian
	LHS	0	-	Frequentist
Fig. 4.8	RHS	$\widetilde{\theta}_k^{\mathrm{Mn}}(T)$	Jeffreys	Bayesian
	LHS	$\widehat{\theta}_k(T)$	-	Frequentist
Fig. 4.9	RHS	$4\ln(T+1)(\widehat{\theta}_k(T))^2$	-	Frequentist
	LHS	$\ln(T+1)\widehat{\theta}_k(T)$	-	Frequentist
Fig. 4.10	RHS	$27\sqrt{\ln(T+1)}(\widehat{\theta}_k(T))^3$	-	Frequentist
	LHS	$9\ln(T+1)(\widehat{\theta}_k(T))^3$	-	Frequentist

Table 4.2: Specifications of modifications

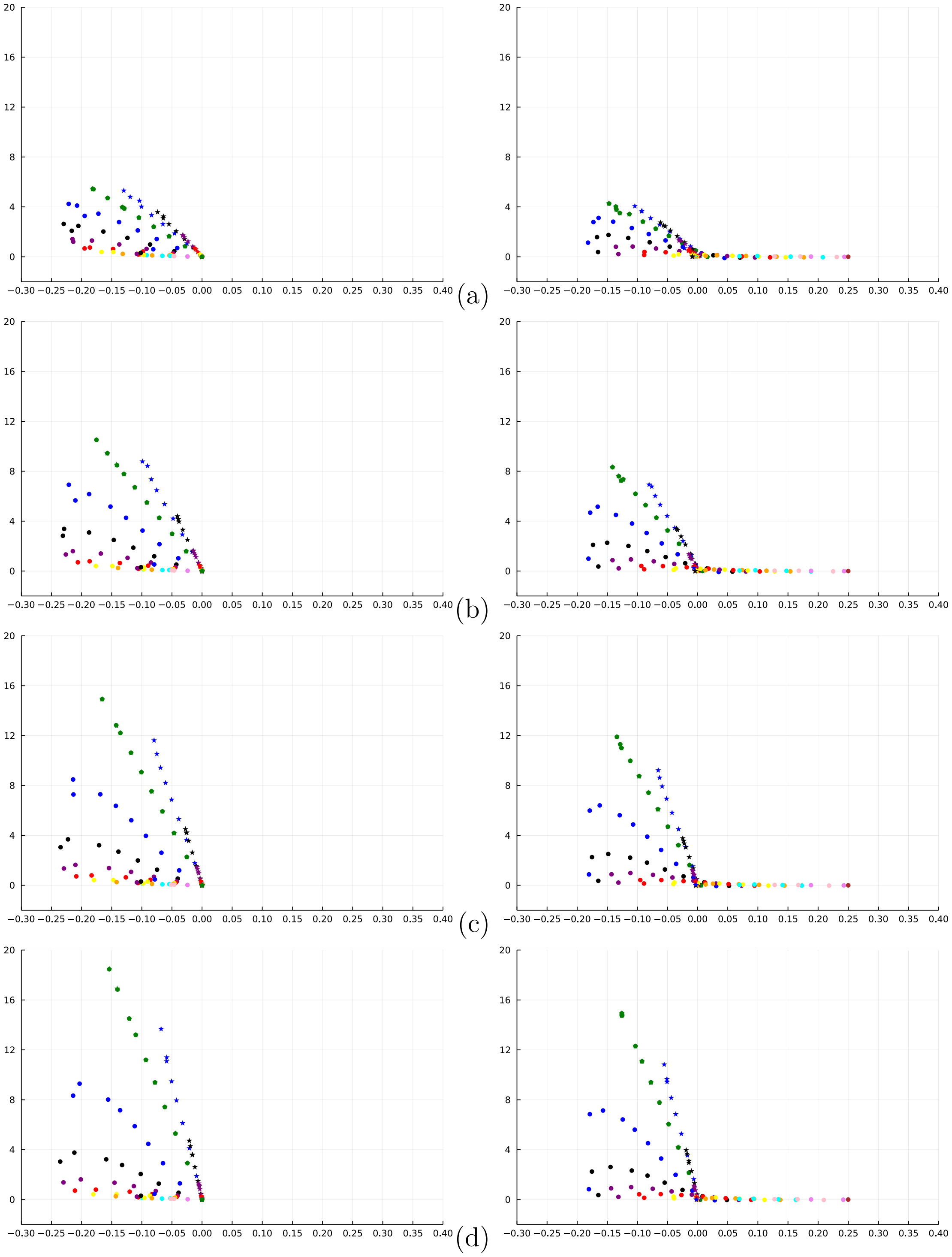


Figure 4.7: $\left(\dot{s}_k(T)=0, \dot{f}_k(T)=0, \widetilde{s}_k(0)=0, \widetilde{f}_k(0)=0\right)$ vs. $\left(\dot{s}_k(T)=0, \dot{f}_k(T)=0, \widetilde{s}_k(0)=1, \widetilde{f}_k(0)=1\right)$ i.e., $\Big($Frequentist MLE eq. (4.1)$\Big)$ vs. $\Big($Bayesian estimator eq. (4.3)$\Big)$: (a) T=60 (b) T=120 (c) T=180 (d) T=240. x-axis: Bias of estimator, y-axis: Covariance (Estimator, Sample Size).

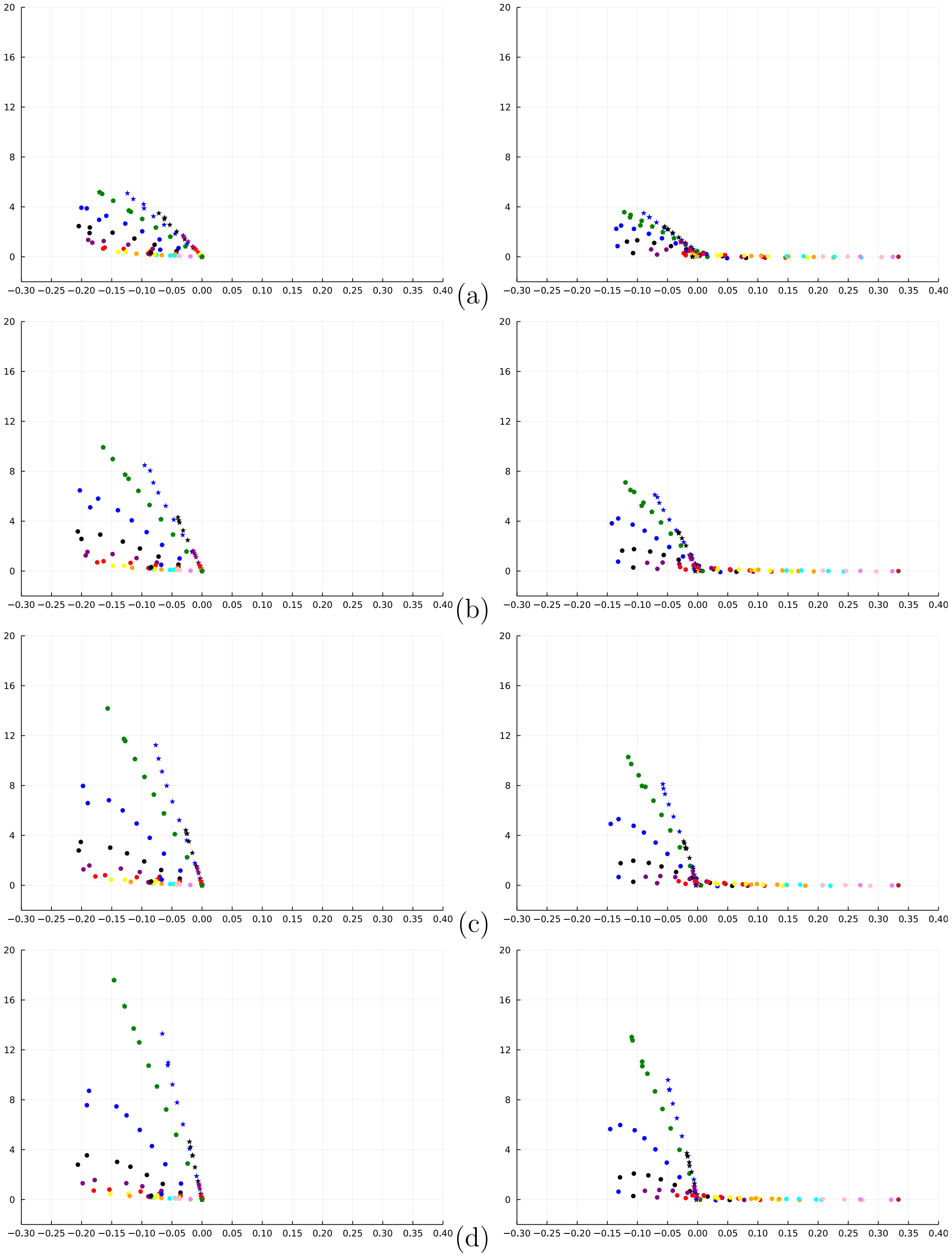


Figure 4.8: $\Big(\dot{s}_k(T) = \widehat{\theta}_k(T), \dot{f}_k(T) = 0, \widetilde{s}_k(0) = 0, \widetilde{f}_k(0) = 0\Big)$ vs. $\Big(\dot{s}_k(T) = \widehat{\theta}_k(T), \dot{f}_k(T) = 0, \widetilde{s}_k(0) = 1/2, \widetilde{f}_k(0) = 1/2\Big)$: (a) T=60 (b) T=120 (c) T=180 (d) T=240. x-axis: Bias of estimator, y-axis: Covariance (Estimator, Sample Size).

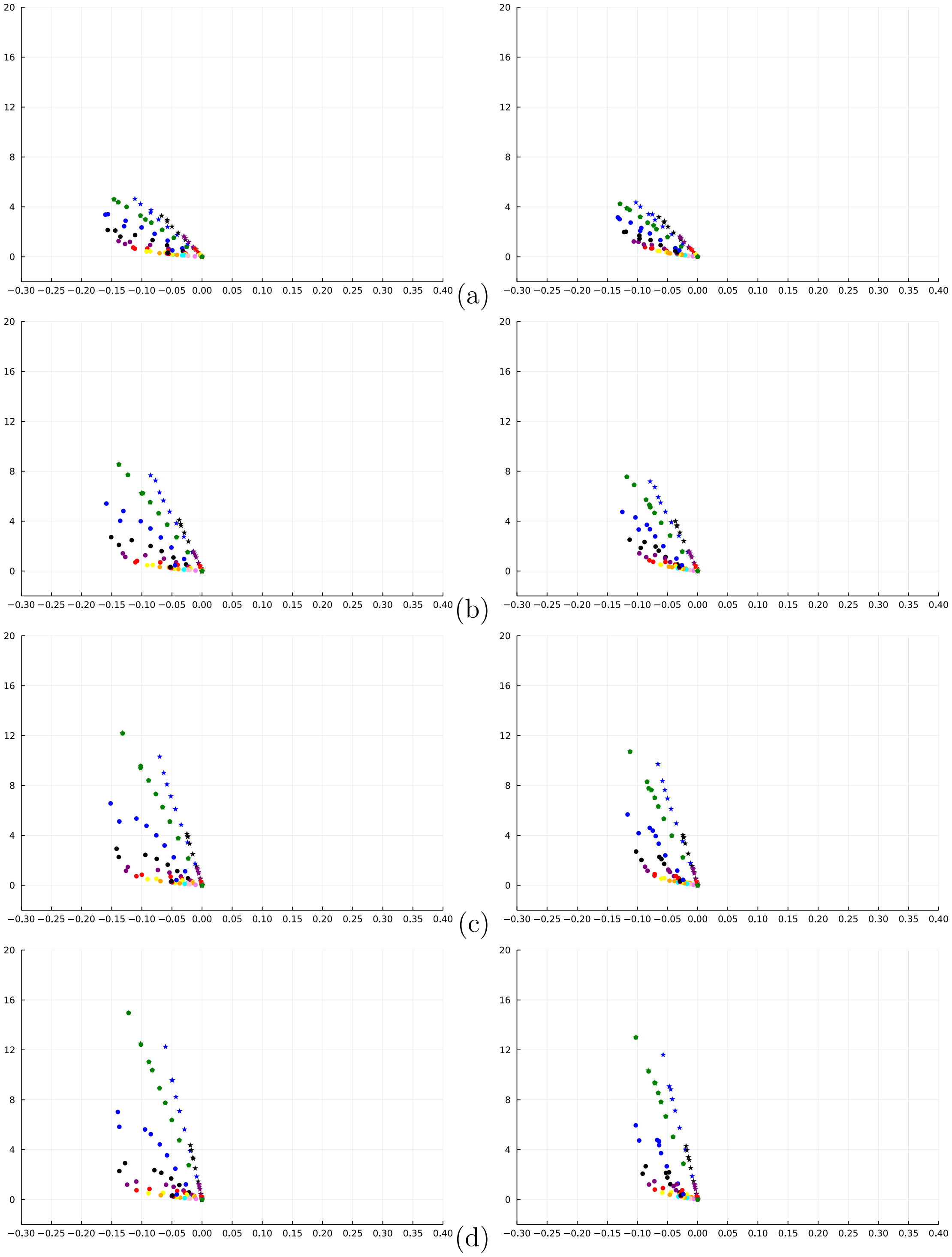


Figure 4.9: $\left(\dot{s}_k(T) = \ln(T+1)\widehat{\theta}_k(T), \dot{f}_k(T) = 0, \widetilde{s}_k(0) = 0, \widetilde{f}_k(0) = 0\right)$ vs. $\left(\dot{s}_k(T) = 4\ln(T+1)(\widehat{\theta}_k(T))^2, \dot{f}_k(T) = 0, \widetilde{s}_k(0) = 0, \widetilde{f}_k(0) = 0\right)$: (a) T=60 (b) T=120 (c) T=180 (d) T=240. x-axis: Bias of estimator, y-axis: Covariance (Estimator, Sample Size).

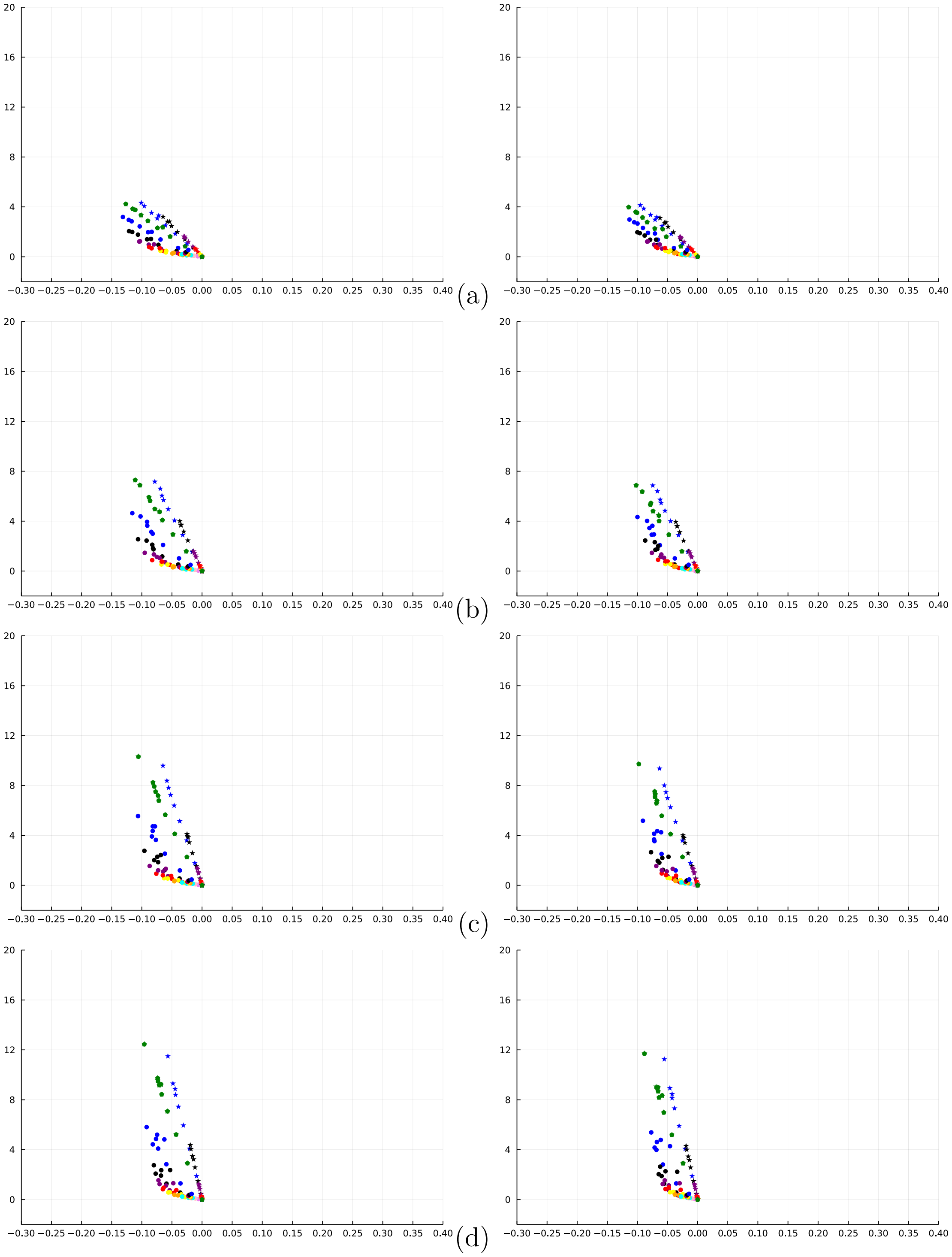


Figure 4.10: $\left(\dot{s}_k(T) = 9\ln(T+1)(\widehat{\theta}_k(T))^3, \dot{f}_k(T) = 0, \widetilde{s}_k(0) = 0, \widetilde{f}_k(0) = 0\right)$ vs. $\left(\dot{s}_k(T) = 27\sqrt{\ln(T+1)}(\widehat{\theta}_k(T))^3, \dot{f}_k(T) = 0, \widetilde{s}_k(0) = 0, \widetilde{f}_k(0) = 0\right)$: (a) T=60 (b) T=120 (c) T=180 (d) T=240. x-axis: Bias of estimator, y-axis: Covariance (Estimator, Sample Size).

Chapter 5

Addressing the trade-off between optimal cumulative reward and unbiased estimation in sequential experiments

5.1 Introduction

As opposed to the standard equal fixed randomisation (EFR) procedures, where allocation probabilities are considered to be fixed during the trial, response-adaptive randomisation (RAR) procedures update allocation probabilities as data is being accrued. This feature of RAR procedures balances: (i) *learning*, i.e. identifying the unknown interventions' efficacies correctly, and (ii) *earning*, i.e. allocating subjects to achieve the objectives during the experiment. For instance, in the clinical trials context, RAR procedures maximise the patients' welfare by reducing exposures to inferior treatment arms based on observed responses.

Conventionally, the design of (non-sequential) experiments, particularly in practice, is expressed based on a (EFR) procedure. EFR, also known as the randomised controlled trial (RCT) in medicine, as the between-group comparison in social sciences, or as the A/B testing in digital marketing, stands for procedures in which allocation probabilities are fixed and equal during the trial. In turn, using data obtained from EFR leads to an unbiased *Maximum Likelihood Estimator* (MLE) for each intervention. However, this procedure is not optimal in terms of resource allocation since half of the available resources have to be directed to the inferior intervention. For example, when it comes to rare disease contexts, allocating a less promising treatment arm to half of the patients is not logically appealing.

This paper considers the MLE to estimate the success probabilities from data collected from a finite-horizon Bayesian Beta-Bernoulli two-armed bandit problem with binary (success/failure) responses. This problem can be considered a fundamental model that usually appears per se or as a sub-problem in bandit-based settings (Jacko, 2019b). Moreover, the standard allocation procedure in the design of sequential experiments obtained by optimising the Bayesian multi-armed bandit problem using dynamic programming (DP) achieves the Bayes-optimal cumulative reward but leads to severely biased MLE.

As mentioned above, both EFR and RAR procedures offer some exclusive advantages, which in a sense might be mutually exclusive. In this chapter, we try to achieve a balance between these advantages in a framework of the RAR procedure as Williamson et al. (2017), Williamson et al. (2022). In chapter 4, we introduced the novel augmented estimator by which the bias induced by DP adaptiveness can be notably mitigated at the end of the trial, whilst this chapter aims to do so by developing novel dynamic modifications and procedures implemented within the trial.

The literature on RAR has been diverse and well-developed ever since Thompson (1933) and Robbins (1952) proposed it for the first time. Whilst Thompson (1933) focused on subject benefit improvement by assuming a Bayesian setting, so Robbins (1952) did for minimising variability in intervention efficacies. Some RAR procedures which emerged after Robbins (1952) showed the "stay-with-a-winner&switch-on-a-loser" rule performs better than EFR. For a comprehensive list, see (Chow and Chang, 2008). Wei and Durham (1978) proposed the "randomised play-the-winner" rule for the first time, and later Williamson et al. (2017) compared the performance of this rule with some other adaptive designs against EFR in a two-armed bandit setting with a binary endpoint. Preliminary results tend to outperform the EFR in terms of subject benefit.

Multi-armed bandit problems (MABP), which nicely balance exploitation vs exploration trade-off, play a pivotal role in the class of problems involving adaptive designs. Bellman is among the pioneers who proposed a backward induction algorithm to study the sequential design of experiments. Due to backwards induction's computational complexity, Gittins (1979) proposed an indexing algorithm called Gittins Index, where allocations take place based on the highest up-to-date indexes. For some RAR designs based on forward-looking Gittins Indices, see (Villar et al., 2015b; Williamson and Villar, 2020), and (Ahuja and Birge, 2016). Whilst Williamson and Villar (2020) propose a forward-looking Gittins index framework for tackling multi-arm bandit models in which the responses are assumed continuous and normally distributed, Ahuja and Birge (2016) using a forward-looking algorithm in the Jointly Adaptive design formulated as the Bayes-adaptive Markov decision process, provide circumstances where one can learn from multiple patients whilst randomised to multiple interventions simultaneously.

A huge stream of literature from clinical trials to social science and A/B testing has focused on the two-armed bandit problem since it serves as the foundation of multi-armed generalisations. For a comprehensive myths review and detailed methodology as well as proposed unified terminology across disciplines, see (Jacko, 2019b). The two-armed bandit problem investigated with the Bayesian learning procedure has been studied by Berry (1978) for the first time. Later, Berry and Eick (1995) by extending the two-armed model, compared some RAR procedures associated with the conflicting goals of patient well-being and high statistical power with EFR. To address the operational characteristics such as statistical power or estimation bias whilst using traditional statistical inference, Cheng and Berry (2007) proposed a constrained adaptive design. In fact, Cheng and Berry (2007) by forcing randomisation over these deterministic decisions, ensure that subjects are allocated to each intervention with a minimum probability of being chosen. Similarly, Williamson et al. (2017) and Williamson et al. (2022) by introducing a family of randomisation procedures referred to as Constrained Randomised Dynamic Programming (CRDP), try to achieve a balance between operational characteristics and subject benefit. The authors in Williamson et al. (2022) also provide an alternative interpretation of CRDP as a bi-level randomisation procedure that can be considered a non-myopic generalisation of the epsilon-greedy algorithm.

In the adaptive sequential experiments, the bias and the variance associated with the estimation of intervention efficacies, i.e. “learning", are amongst critical measures. Skewing allocations to the superior intervention by updating the randomisation probabilities as data accrue gives rise to not having enough observations on the other intervention, and therefore the corresponding estimator will be heavily biased. Hence, all classical estimators utilising data obtained by any RAR procedures are biased. Bowden and Trippa (2017) showed that the maximum likelihood estimator (MLE) under RAR procedures is biased for the sample mean on which the parameter of interest is each arm's efficacy. Nie et al. (2018) proved that the MLE is associated with negative bias if the data collection procedure by which the MLE is estimated satisfies certain conditions. Later, Shin et al. (2019a) and Shin et al. (2019b) by categorising the sources of the estimation bias in any adaptive procedures under the umbrella of adaptive *sampling*, adaptive *stopping*, adaptive *choosing*, and adaptive *rewinding*, showed that, depending on the data collection assumptions, MLE might be associated with positive bias as well as negative. A novel estimator called adaptively weighted augmented inverse-probability weighting (IPW) is introduced by (Hadad et al., 2021). The authors also propose a particular class of test statistics leading to asymptotically normally distributed unbiased estimators to develop frequentist confidence intervals for interventions' efficacy in adaptive experimental settings with normally distributed responses.

5.1.1 Our contribution

This chapter considers the MLE to estimate the success probabilities from data collected from a finite-horizon Bayesian Beta-Bernoulli two-armed bandit problem with binary (success/failure) responses, explained in section 3.1. As opposed to the approach introduced in chapter 4, in which estimations can be corrected by selecting appropriate augmentations at the end of the trial, we propose two novel allocation procedures that can correct the bias induced by the DP procedure during the trial. Taking into account that randomisation can potentially eliminate estimation bias Rosenberger and Lachin (2015), we try to move away from deterministic allocation baseline in DP procedure towards randomising actions containing deterministic allocations (Williamson et al., 2017), (Williamson et al., 2022). The contribution of our proposed procedures beyond existing literature can be summarised as follows:

- A novel modification framework applying to the DP procedure using an *augmented* estimator in order to overcome the passive-aggressive behaviour of classical DP is introduced in section 5.2. As opposed to chapter 4, where we modified MLE using appropriate augmented estimators, allocation decisions are modified during the trial and at every step.

- We develop an RDP procedure for a two-armed bandit problem with binary responses, where Bayes-optimal allocation decisions are perturbed at each decision step. This procedure can be equivalently interpreted as so-called bi-level randomisation. Moreover, the generalisation of the RDP procedure in terms of constrained RDP (CRDP) and bi-level randomisation are proposed by Williamson et al. (2017) and later Williamson et al. (2022) for the first time in the literature.

- The inverse Probability Weighting (IPW) estimator and its normalised version (nIPW) is evaluated in section 5.4. We show that the IPW and nIPW estimators can be estimated unbiasedly using our proposed RDP procedure.

- In section 5.5, we provide a comparison of MLEs' performance in some purposefully-selected designs in terms of root mean squared error (RMSE).

Results confirm that the estimator variability grows as the degree of randomisation increases in the RDP procedures. By visualising the simulated data on MLE using box plot, we compare the performance of this estimator in our proposed designs with EFR and classical DP.

5.2 Response-Adaptive non-Randomised (RAnR) procedure

The two-armed Bayesian Beta-Bernoulli model formulated in section 3.1 could be exactly solved by DP. Since assuming deterministic actions in this model, it can be known as a Response-Adaptive non-Randomised (RAnR) procedure. Moreover, enumeration, by which all possible allocation sequences over a time horizon can be determined, is the most striking feature of DP methods. Although DP suffers from the curse of dimensionality and therefore classifies as a computationally intensive method Bellman (1966), it is considered as a myopic allocation procedure in RAR designs and any online learning strategies. On the other hand, DP is likened to the passive-aggressive family of algorithms in Machine Learning, particularly for some circumstances where the gap between interventions' efficacy is not negligible. Being passive-aggressive means that when DP adheres to a seemingly superior intervention after a few success responses, it will most likely continue allocating subjects to that arm, which in turn gives rise to a relative lack of enough observations on the other arm. Subsequently, this severe shortage of response contributes to MLE estimation being notably biased at the end of the trial. Moreover, the essence of the action set defined in the above bandit model implies the passive-aggressive performance of DP. Based on the action set assumption, we can solely have three different actions as follows:

(i) Action 1 ($a = 1$): the next subject allocates to arm C with $p_C^a = 1$ and to arm D with $p_D^a = 0$

(ii) Action 2 ($a = 2$): the next subject allocates to arm C with $p_C^a = 0$ and to arm D with $p_D^a = 1$

(iii) Action 3 ($a = 3$): the next subject allocates to arm C with $p_C^a = 1/2$ and to arm D with $p_D^a = 1/2$

Except for the third action denoting a simple 50:50 randomisation when there is no difference between interventions' rewards, others are utterly deterministic. Although in the next section we raise a point about *randomised dynamic programming* (RDP), which can overcome the passive-aggressive property of DP to a considerable extent, we now introduce a novel contribution called *Optimistic on Inferior DP* (OIDP).

5.2.1 Optimistic on Inferior Dynamic Programming (OIDP)

The idea of the OIDP originates from the question: Is that possible to force DP to pull the allegedly inferior intervention more frequently so that the number of observations on that intervention increases, which, in turn, will imply that the bias of the MLE will be mitigated most likely? In other words, we try to soften the impact of the passive-aggressive behaviour on DP by making the seemingly inferior intervention more attractive to assign. The question is finely answered by changing the Bayes-expected number of successes on the inferior intervention, which increases the expected total rewards, whilst the principle of optimality still holds. Hence, at each time epoch in the backward induction algorithm, a particular pseudo-success response is applied to the intervention with a lower quantity of $q_{k,\boldsymbol{x},1}$ in the equation (3.1). Bear in mind, at time epochs where there is no difference in the interventions Bayesian posterior expectations and subsequently in the up-to-date value functions, we do not apply modifications on $\widetilde{q}_{k,\boldsymbol{x},1}$. Thus, the expected total rewards for both interventions remain the same. Let k denote the inferior intervention at each time epoch. The modified $q_{k,\boldsymbol{x},1}$ can be formulated as follows:

$$\widetilde{q}_{k,\boldsymbol{x},1} = \frac{\widetilde{s_k} + \dot{s}}{\widetilde{s_k} + \widetilde{f_k} + \dot{s}} \,, \tag{5.1}$$

where $\dot{s}$ is the pseudo-success responses. Although the optimal quantity of the pseudo-success observation in order to have MLE unbiasedly estimated in the assumed RAnR procedure is still unknown, we simulate some trials where the pseudo-increments have been set to $\dot{s} = 1, 2,$ and $\ln(t)$. Note that, $\dot{s} = \ln(t)$ is known as dynamic increments, because of the fact that as the trial moves forward, the larger pseudo-increments are applied to $q_{k,\boldsymbol{x},1}$ corresponding to the inferior intervention. Table 5.1 shows how one can develop an algorithm for the OIDP procedure.

5.2.2 Subject Benefit

The idea of utilising OIDP in a trial design procedure can notably mitigate the estimation bias, whereas it might adversely affect the subject benefit whilst the impact is not appreciable. Tables 5.2 and 5.3 summarise the numerical results of the subject benefit, defined as the average of aggregated success responses on both interventions together with corresponding standard deviation at the end of

Algorithm 1: OIDP procedure

```
Parameter: set up a positive $\dot{s}$.
    for $t = T - 1$ to 0 do
    enumerate all possible combinations of $s_C(t) + f_C(t) + s_D(t) + f_D(t) = t$.
    calculate $q_{C,\boldsymbol{x},1}$ and $q_{D,\boldsymbol{x},1}$ for each combination obtained above.
        if $q_{C,\boldsymbol{x},1} > q_{D,\boldsymbol{x},1}$
            replace $q_{D,\boldsymbol{x},1}$ with $\widetilde{q}_{D,\boldsymbol{x},1}$ in (5.1).
        elseif $q_{C,\boldsymbol{x},1} < q_{D,\boldsymbol{x},1}$
            replace $q_{C,\boldsymbol{x},1}$ with $\widetilde{q}_{C,\boldsymbol{x},1}$ in (5.1).
        else keep $q_{C,\boldsymbol{x},1}$ and $q_{D,\boldsymbol{x},1}$
        determine the optimal arm using (3.5).
        end if
    end for
```

Table 5.1: OIDP procedure pseu-code

the trial. Hence, given a scenario, one can compare the subject benefit obtained from different OIDP designs with one another. For instance, scenario $(0.2, 0.8)$ in two extreme trial sizes: $T = 60$ and $T = 240$, tends to show subject benefit of 46.89 ± 3.35 and 190.69 ± 6.376, respectively in the design with classical DP, whilst it reduces to 45.83 ± 3.48 and 188.84 ± 6.72, respectively in the design with OIDP in which $\dot{s} = 1$. Furthermore, the reduction rate is slightly higher if one considers the design with OIDP in which $\dot{s} = 2$, 45.30 ± 3.71 and 187.87 ± 7, respectively. Note that, owing to the performance of the natural logarithm function, the difference between designs with OIDP where $\dot{s} = 2$ and $\dot{s} = \ln(t)$ is not noticeable in terms of subject benefit. Despite infinitesimal increments happening in some scenarios, which can be attributed to the error and uncertainty in simulation, for some other cases, particularly those with identical success probabilities, the subject benefit remains the same regardless of the modifications applied.

Although it can be concluded that the larger $\dot{s}$ applying to OIDP designs leads to higher reductions in subject benefit and the estimation bias consequently, there is no unique trend describing the relationship between the reduction of subject benefit rate, the quantity of modification in OIDP, and the arm D effect. Note that the trade-off between sacrificing the subject benefit in favour of mitigating the estimation bias depends mostly on researchers and trial experts. Considering that the overall reduction is negligible, the mentioned trade-off may vary in different trial platforms.

5.2.3 Average Estimation Bias

As opposed to the loss of subject benefit upon applying modifications to OIDP, the average estimation bias is notably reduced. Figure 5.2 illustrates the comparison of estimation bias reduction between classical DP (left-hand side group of plots) and OIDP where $\dot{s} = 1$ (right-hand side group of plots), as does figure 5.3 between OIDP where $\dot{s} = 2$ (left-hand side group of plots) and $\dot{s} = \ln(t)$ (right-hand side group of plots). Because of the passive-aggressive performance of the classical DP, both bias and covariance values tend towards large magnitudes. Although the bias values remain more or less the same whilst the trial size increases, covariance values increase noticeably. It is worth mentioning that the latter can be generalised to all other RAR procedures proposed in this paper, as the covariance of sample size (trial size) and the intended estimator is measured.

On the other hand, for those trial designs using OIDP, not only bias and covariance has notably plummeted, but the scatteredness throughout different scenarios and the arms has been significantly reduced. Figure 5.2 shows that the estimation bias is declined by approximately 0.1 for all trial sizes when one moves away from the classical DP baseline to OIDP with $\dot{s} = 1$. It is discernible that the gaps between the estimates for the arms' efficacy (the gaps between stars and circles) are notably reduced in each scenario.

Except for those scenarios where the actual efficacies are extreme, i.e. 0 and/or 1, a generic observation in both classical DP and OIDPs designs is that as the difference in arm D effect grows, the gap in arms efficacies also widens. For example, there is no gap in green circles and stars since they stand for scenarios with equal success probabilities. In contrast, the gaps in blue circles and stars are smaller than black ones since the arm D effect is 0.1 and 0.2 in blue and black scenarios, respectively. Unlike the stars, which denote the research arm and thus are less biased, circles exhibit larger bias quantities. This is mainly because of the fact that in each scenario defined in this paper, θ_D is always considered equal to or greater than θ_C. In turn, by taking into account the passive-aggressive performance of DP based designs, circles are placed farther away than the stars from origin. Note that attention must be paid to the fact that there is no direct relationship between the variation in the location of marker shapes, i.e. stars and circles, and the growing trend in success probabilities in scenarios, necessarily. Hence, the numbers reported in the text may not be identifiable from the figures but written based on numerical excel files.

To be more specific, we draw attention to those scenarios with an arm D effect of 0.1 i.e. blue-colour stars and circles since there is a direct relationship between the rate of growth of bias and covariance values and the magnitudes of success probabilities. Hence, scenario $(0.7, 0.8)$ exhibits the worst performance amongst others (not even $(0.8, 0.9)$ since either arm places relatively close to 1 and therefore might be considered as an extreme efficacy) in all trial designs presented in figure 5.2. Applying equation (3.7) to the designs using classical DP (left-hand side column of figures 5.2) the scenario $(0.7, 0.8)$ estimated for $(0.7 - 0.22, 0.8 - 0.13) = (0.48, 0.67)$ in $T = 60$, and $(0.7 - 0.21, 0.8 - 0.06) = (0.49, 0.74)$ in $T = 240$. However, for the designs using OIDP with $\dot{s} = 1$ (right-hand side column of figures 5.2), it is estimated for $(0.7 - 0.11, 0.8 - 0.05) = (0.59, 0.75)$ and $(0.7 - 0.09, 0.8 - 0.02) = (0.61, 0.78)$ in $T = 60$ and $T = 240$, respectively. The amount of improvement on the MLE estimation process is more discernible in the designs using OIDP with $\dot{s} = 2$ presented in figure 5.3, left-hand side column. As it can be seen the scenario $(07, 08)$ in this class of designs estimated for $(0.7 - 0.09, 0.8 - 0.04) = (0.61, 0.76)$ and $(0.7 - 0.06, 0.8 - 0.01) = (0.64, 0.79)$ in $T = 60$ and $T = 240$, respectively. On the other hand, the variability of the designs using OIDP with dynamic modifications, $\dot{s} = \ln(t)$, is relatively high, because of the fact that $\dot{s}$ values applied to equation (5.1) are not fixed, but change as time epochs and/or the trial size shift. Moreover, utilising the backward induction algorithm in the randomisation procedure incorporated into either DP or OIDP based designs results in exerting more considerable pseudo-increments i.e. modifications at the beginning of the trial rather than the middle or towards the end. Hence, it can be seen from the figure 5.3 (right-hand side column) that the performance of the designs using OIDP with dynamic modifications in some scenarios is slightly better than the classical DP and other OIDP-based designs for all assumed trial sizes. However, it is not comparably good for some other scenarios, particularly those made up of smaller success probabilities. Figure 5.3 (right-hand side column) shows that the scenario $(0.7, 0.8)$ is estimated for $(0.7 - 0.12, 0.8 - 0.03) = (0.61, 0.76)$ and $(0.7 - 0.07, 0.8 - 0.01) = (0.64, 0.79)$ for $T = 60$ and $T = 240$, respectively.

5.3 Randomised Dynamic Programming (RDP)

In this section, we propose a randomisation framework to the DP procedure whereby the passive-aggressive property can be mitigated significantly. On the one hand, this procedure allows for tuning the involved parameters to meet the overall and practical goals of the trial. On the other hand, it minimises the bias of the MLE and subsequently achieves a decent level of accuracy and reliability. As Chow and Liu (2008) discuss randomisation is a vital element in the design of clinical trials. By forcing actions to be randomised throughout the DP procedure at each time epoch, a fully randomised framework named *randomised dynamic programming* (RDP) can be achieved. Williamson et al. (2017) is amongst the pioneers of optimally designing trials using RDP. The first step towards formulating RDP is taking the optimal action instead of the optimal arm at each time epoch. Considering the two-armed Bayesian Betta-Bernoulli model described in section 6.2, we define the following actions so that, at time epoch $t+1$ a given arm can be pulled with a probability of $p_k(t)$, whilst the other has an equivalent complementary probability:

(i) Action 1 ($a = 1$): the next subject allocates to arm C with $p_C(t)$ and to arm D with $1 - p_C(t)$

(ii) Action 2 ($a = 2$): the next subject allocates to arm D with $p_D(t)$ and to arm C with $1 - p_D(t)$

(iii) Action 3 ($a = 3$): equally randomised between these two actions if there is no difference between arms' rewards

In turn, by updating the definition of actions and subsequently, corresponding value functions, the expected total reward, i.e. the Bayes-expected number of success, for time epoch $t+1$ to T can be as below when $a=1$

$$F_t^1\Big(s_C, f_C, s_D, f_D\Big) = p_C(t).F_t^C\Big(s_C, f_C, s_D, f_D\Big) + \Big(1-p_C(t)\Big).F_t^D\Big(s_C, f_C, s_D, f_D\Big)$$

and similarly when $a=2$

$$F_t^2\Big(s_C, f_C, s_D, f_D\Big) = \Big(1-p_D(t)\Big).F_t^C\Big(s_C, f_C, s_D, f_D\Big) + p_D(t).F_t^D\Big(s_C, f_C, s_D, f_D\Big)$$

Therefore, an equivalent expression as (3.5) satisfies

$$F_t\Big(s_C, f_C, s_D, f_D\Big) = max\Big\{F_t^1\Big(s_C, f_C, s_D, f_D\Big), F_t^2\Big(s_C, f_C, s_D, f_D\Big)\Big\} \quad \text{for } 0 \leq t \leq T-1,$$
$$F_t\Big(s_C, f_C, s_D, f_D\Big) = 0, \quad \text{otherwise.}$$

It would be worth mentioning $p_k(t)$ referring to *degree of randomisation* is defined by a set of parameters $0 \leq p_k(t) \leq 1$ for each arm k at each time epoch $0 \leq t \leq T$ (Williamson et al., 2017) and (Williamson et al., 2022). Although $p_k(t)$ may vary arm to arm and differ between time epochs, it is usually set as a fixed parameter not only for both arms but also for the whole course of the trial. In practice, trialists consider $p_k(t)$ no less than 0.5 and usually set $0.5 \leq p_k(t) \leq 1$, although this is not a theoretical obligation. Note that, assuming $p_k(t) = 0.5$ and $p_k(t) = 1$ for both arms and throughout the trial, recovers equal randomisation, EFR, and RAnR procedure, respectively. This study mainly focuses on RDP designs in which $p_k(t) = p = 0.9$ for all trial sizes. Additionally, we investigate those RDP designs where the different "degree of randomisation", $p_k(t) = p = 0.6, 0.7, 0.8$, is used for the trial size $T = 60$, merely. This study mainly focuses on RDP designs in which $p_k(t) = p = 0.9$ for all trial sizes. Additionally, we investigate those RDP designs where the different “degree of randomisation", $p_k(t) = p = 0.6, 0.7, 0.8$, is used for the trial size $T = 60$, merely. Note that we are interested in setting p closer to 1 in order to not being exceedingly deviated from the “exploitation" baseline. As a result, this leads to estimating MLE in a less biased manner at the end of the trial.

5.3.1 Bi-level Randomisation- Alternative Interpretation of RDP

There is an alternative interpretation of the RDP randomisation procedure proposed by (Williamson et al., 2022). The RDP randomisation procedure is equivalent to a situation when there are two parallel branches of the trial: a *fixed branch* and an *adaptive branch*. In the first level of randomisation, and after the first couple of time epochs where both arms are deterministically pulled one after the other (in this study, we set up all the simulation algorithms commencing in this way to prevent trials from ending with no observation on each arm) each time epoch t is randomised between either fixed or adaptive branches. Thus, it might be directed towards the fixed branch with probability $2 - p_C(t) - p_D(t)$, or the adaptive one with the complementary probability $p_C(t) + p_D(t) - 1$. In the second level, if time epoch t has been diverted to the fixed branch, then the randomisation probabilities to arm C and arm D will be $1 - p_C(t)$ and $1 - p_D(t)$, respectively. Otherwise, the time epoch t has been directed to the adaptive branch where the allocation procedure is deterministic, i.e. the allocation probabilities are either 0 or 1. In fact, the allocation procedure follows the RDP matching actions depending on joint data $\boldsymbol{Z}$

available up to time epoch t. Thus, one of the following possibilities will happen to the current time epoch t:

(i) under action 1: it allocates to arm C with probability 1 and to arm D with probability 0

(ii) under action 2: it allocates to arm D with probability 1 and to arm C with probability 0

(iii) under action 3: it equally randomises between two arms if there is no difference between arms' rewards

Note that, in this study, $p_k(t)$ values are considered to be fixed, equal for both arms and determined before the trial, whilst they are time dependent in general and can be different from arm to arm. Randomisation occurs uniformly between arms throughout the fixed branch, resulting in the EFR procedure. Thus, randomisation probabilities are considered 50% in a two-armed trial. Figure 5.1 below depicts the whole bi-level randomisation procedure described above.

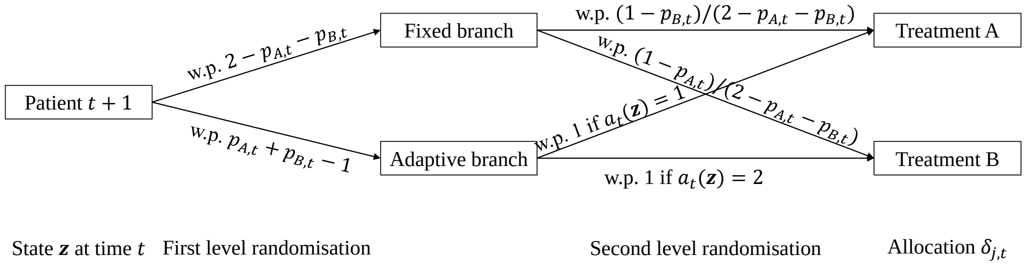

Patient $t + 1$
w.p. $2 - p_{A,t} - p_{B,t}$
w.p. $p_{A,t} + p_{B,t} - 1$
Fixed branch
Adaptive branch
w.p. $(1 - p_{B,t})/(2 - p_{A,t} - p_{B,t})$
w.p. $(1 - p_{A,t})/(2 - p_{A,t} - p_{B,t})$
w.p. 1 if $a_t(\boldsymbol{z}) = 1$
w.p. 1 if $a_t(\boldsymbol{z}) = 2$
Treatment A
Treatment B
State $\boldsymbol{z}$ at time t
First level randomisation
Second level randomisation
Allocation $\delta_{j,t}$

Figure 5.1: Alternative interpretation of the RDP randomisation procedure

To provide a comprehensive insight into the merits and demerits offered by RDP and the bi-level randomisation procedure, we classify after-trial results into three categories: Observations obtained through (i) initial allocations, (ii) fixed branch, and (iii) adaptive branch. Initial allocations refer to the prevention of not having any observation at the end of the trial (default assumption of all simulation setups in this study). We group this category with the fixed branch since allocations are deterministic with an equal ratio of 1:1. Furthermore, the synthesis of all branches, which we call *pooled data*, is compared with the adaptive branch data to demonstrate the effect of randomisation and bi-level procedure in reducing the estimation bias. Figure 5.4 compares the adaptive branch data with corresponding pooled data for all trial sizes. Figure 5.5, on the other hand, depicts the impact of using the different degrees of randomisation on the estimation bias trend for $T = 60$.

5.3.2 Subject Benefit

The left-hand side column of table 5.4 summarises the subject benefit information from the RDP procedure with a degree of randomisation of 0.9 for all trial sizes. Focusing on scenario $(0.2, 0.8)$, we compare subject benefits stemming from RDP design with classical DP counterparts presented in table 5.2. For two extreme trial sizes, $T = 60$ and $T = 240$, classical DP design returns a subject benefit of 46.89 ± 3.35 and 190.69 ± 6.376, respectively, whereas it plummets to 43.55 ± 3.52 and 176.70 ± 6.86, in the RDP procedure, respectively. Also, in trial sizes $T = 120$ and $T = 180$ it plunges from 94.82 ± 4.58 and 142.76 ± 5.56 to 87.93 ± 4.589 and 132.32 ± 5.95. Although the reductions in subject benefits are significant, the average estimation bias declined substantially. In other words, a higher level of randomness gives rise to fewer success responses but more accurate estimation, i.e. less bias at the end of the trial. Thus, we emphasise that it depends on researchers and trialists to identify the extent to which they aim to sacrifice subject benefit in favour of mitigating bias.

The subject benefit information on RDP procedure with different randomisation degrees is summarised on the right-hand side of table 5.4. The aim of presenting this column is to see the effect of altering randomisation degrees on the number of success responses at the end of the trial. The underlying trend in the column labelled by "RDP with different p values for $T = 60$" is opposite of the right-hand side one in table 5.4. In fact, as the assumed degrees of randomisation become bigger and closer to 1 than 0.5, the trial design moves away from the EFR baseline and becomes more adaptive. In turn, the number of success responses and estimation bias noticeably increase. For instance, scenario $(0.2, 0.8)$ results in 30 ± 3.85 and 40.18 ± 3.66 subject benefits when $p = 0.5$ and $p = 0.8$, respectively, whilst it may be seen in figure 5.5 that the corresponding estimated efficacies are associated with significant bias.

5.3.3 Average Estimation Bias

Figure 5.4 illustrates the average bias associated with the arms' efficacies estimation utilising adaptive branches results (left-hand side column of plots) and pooled data (right-hand side column of plots) for designs with RDP where $p = 0.9$. Aside from the fact that bigger trial sizes result in terminating the trial with more available observations and therefore less bias, the impact of randomness entailed in the RDP procedure on average bias reduction and data variability is substantial. For example, under scenario $(0.7, 0.8)$ for $T = 60$, which has the highest bias amongst other scenarios as we discussed in section 5.2.3, the bias is estimated as $(0.7 - 0.17, 0.8 - 0.08) = (0.53, 0.72)$ using data obtained from adaptive branch whilst it is estimated as $(0.7 - 0.05, 0.8 - 0.03) = (0.65, 0.78)$ by pooled data. For the case $T = 240$ it is also estimated as $(0.7 - 0.14, 0.8 - 0.02) = (0.56, 0.78)$ and $(0.7 - 0.02, 0.8 - 0.01) = (0.78, 0.79)$ using adaptive branch and pooled data, respectively. As a result, it can be concluded that the differences for larger trial sizes under the RDP procedure are negligible.

Figure 5.5 also compares the average bias derived from the arms' efficacies estimation process using adaptive branch observation (left-hand side column of plots) with pooled data (right-hand side column of plots). In this case, we presume the trial size to be fixed at $T = 60$, and designs vary according to the assumed degrees of randomisation. It is noteworthy that setting $p = 0.5$ results in an EFR design as the randomisation procedure is ongoing with odds of 0.5 for both arms at each time epoch. Hence, all estimation results presented in figure 5.5 part (a) are unbiased in both adaptive branch and pooled data. As a matter of fact, as the assumed degrees of randomisation become bigger and closer to 1 than 0.5, the trial design moves away from the EFR baseline and becomes more adaptive. Moreover, the intensity of becoming adaptive and acting passive-aggressive as the degree of randomisation grows is notably higher in the adaptive branch than in pooled data cases. For instance, scenario $(0.7, 0.8)$ using pooled data in the design with $p = 0.6$ is estimated as $(0.7 - 0.01, 0.8 - 0.00) = (0.69, 0.8)$, whilst in the design with $p = 0.8$ it comes to $(0.7 - 0.02, 0.8 - 0.02) = (0.68, 0.78)$. Although the differences in arms' efficacies estimation using pooled data set might be negligible, it is not the case when the estimation process is carried out in the adaptive branch. For example, scenario $(0.7, 0.8)$ is estimated as $(0.7 - 0.08, 0.8 - 0.03) = (0.62, 0.77)$ and $(0.7 - 0.15, 0.8 - 0.06) = (0.55, 0.74)$ in the design with $p = 0.6$ and $p = 0.8$, respectively.

5.4 An Unbiased Estimator

In this section, we note the performance of the Inverse Probability Weighting (IPW) estimator, which is usually considered a bias-corrected estimate for MLE Bowden and Trippa (2017) and Hadad et al. (2021), and its normalised version (nIPw) in designs with the RDP procedure. Since IPW utilises randomisation probabilities in the estimation process, it can finely compensate for the lack of observations in a randomised setting such as the RDP designs. Considering $\pi_k(t)$ as randomisation probabilities at each time epoch t, then:

$$\widehat{\theta}_{k,IPW} = \frac{1}{T}\sum_{t=1}^{T} \frac{\delta_k(t) y_k(t)}{\pi_k(t)} \, , \tag{5.2}$$

where $Y_k(t)$ represents the random variable (and $y_k(t)$ the realization) of the response at time epoch $t \in \mathcal{T}$ corresponding to arm $k \in \mathcal{K}$, and $\delta_k(t)$ stands for a binary variable equalling 1 if time epoch t is allocated to arm k and 0 otherwise. Note that the IPW estimator's main drawback is its variability, which may go beyond the acceptable range, i.e. larger than 1. Applying a constraint to overcome this issue, the new estimator will not be unbiased anymore. Hence, we take a similar approach as in Bowden and Trippa (2017) and normalise the IPW estimator to shrink it to be within the unit interval. Denote by $k(t) \in \mathcal{K}$ the actual assigned arm at each time epoch t such that $\delta_{k(t)}(t) \equiv 1$, then, the “normalised" IPW estimator (nIPW) is as follows:

$$\widehat{\theta}_{k,nIPW} = \frac{T\widehat{\theta}_{k,IPW}}{\sum_{t=1}^{T} \dfrac{\delta_k(t)}{\pi_{k(t)}(t)}} \tag{5.3}$$

This study presents IPW and nIPW estimations using the RDP designs with $p = 0.9$ for all trial sizes (see figure 5.6). It is worth mentioning that although randomisation probabilities $\pi_{k(t)}(t)$ can be set in a time-varying manner, in both (5.2) and (5.3) they are considered to be fixed for both arms during the course of the trial. On the other hand, they follow one of the possibilities below at each time epoch:

$$\forall t \in \mathcal{T} \cup \{T\}, \quad \pi_{k(t)}(t) = \pi = \begin{cases} 0.5 & \text{if } t \leq 2 \text{ or no difference between arms' rewards} \\ 0.1 & \text{if } t \text{ randomised to up-to-date inferior arm} \\ 0.9 & \text{otherwise} \end{cases} \tag{5.4}$$

Figure 5.6 (left-hand side column of plots) shows the bias of the IPW for the RDP design with $p = 0.9$ for all trial sizes. Apart from the slight alterations in the covariance axis (owing to utilising different trial sizes), all IPW estimators tend to be unbiased. Similarly, figure 5.6 (right-hand side column of plots) represents the resultant nIPW estimators associated with very little bias. However, as we mentioned above, these estimators are associated with reasonably small variability that is now constrained to $(0, 1)$. Furthermore, it is evident that assuming a large enough trial size can compensate for the bias of nIPW estimators, as we can see throughout designs with $T = 180$ and $T = 240$. As a result, the performances of both IPW and nIPW estimators in a randomised context can be considered almost unbiased. In contrast, variability might be crucial in trialists' and experts' final decisions in favour or against IPW and nIPW estimators.

5.5 Root Mean Square Error (RMSE) and Data Visualisation

In this section, we compare the performance of MLE in terms of quality and variability in some proposed designs. The main focus will be on root mean square error (RMSE) and box plots. Note that we evaluate the performances of proposed design procedures by simulating a million replications of each trial and taking the average values.

5.5.1 Root Mean Square Error (RMSE)

The RMSE can measure the quality and variability of an estimator since it has the same units as the estimators themselves:

$$RMSE(\widehat{\theta}) = \sqrt{\mathbb{E}\left[\left(\widehat{\theta} - \theta\right)^2\right]}\,. \tag{5.5}$$

It can also be expressed as follows:

$$RMSE(\widehat{\theta}) = \sqrt{Bias(\widehat{\theta})^2 + Var(\widehat{\theta})} \tag{5.6}$$

Figure 5.7 compares the RMSE of the MLE, i.e. arm efficacy, as defined by (5.5), in OIDP and RDP designs assuming $T = 60$. The left-hand column of the plots illustrates the RMSE for OIDP designs with all proposed modifications, and as does the right-hand side column for RDP with different degrees of randomisations. Note that in figure 5.7 the horizontal axis represents scenarios (θ_C, θ_D) and the vertical axis shows the corresponding RMSE values. In general, all RDP designs together with classical DP (part a, left-hand side plot) exhibit relatively large RMSE values, which can be attributed to the large estimation bias values we discussed in section 5.2 (see Figures 5.2 and 5.3). A general observation pertinent to the large estimation bias is that the gap between stars and circles increases as the arm D effect increases. For example, the gaps in blue stars and circles, i.e. scenarios in which the arm D effect 0.1, are smaller than those in black and purple with arm D effect of 0.2 and 0.3, respectively. Although differences in RMSE and variabilities amongst scenarios are relatively high in designs with DP and OIDP procedures, compared to designs with RDP counterparts, the general expanding trend related to the arm D effect can also be traced in the designs with RDP procedures. Moving away from the EFR baseline by increasing the degrees of randomisation, the variabilities in RMSE are also notably heightening. However, moving from the classical DP baseline by applying appropriate $\dot{s}$ to OIDP procedures, the variabilities slightly reduce. According to equation (5.6), the variability in RMSE is directly related to estimation bias induced in the MLE, which can be attributed to the passive-aggressive performance of the DP algorithm (see section 5.2 figures 5.2, 5.3, and 5.5).

5.5.2 Data visualisation- Box plot

This section compares the diffusion of MLE estimations for a trial size $T = 240$ across the design using OIDP with $\dot{s} = 2$, RDP with $p = 0.9$, and EFR with the design utilising classical DP. We draw the box plots for a range of scenarios in which θ_C is fixed at 0.5, and θ_D varies from 0 to 1, i.e. $(0.5, 0), (0.5, 0.1), (0.5, 0.2), ..., (0.5, 1)$. Hence, the left-hand side of plots in figure 5.8 depicts arm C variations, as does the right-hand side for arm D. Note that the horizontal axis on both sides shows the range of θ_D, whilst the vertical axis is set to be in the range of $[0, 1]$. Since arm C is considered as the superior arm, for those scenarios where θ_D is less than 0.5, the Inter Quartile Range (IQR) and second quartile, i.e. median, are almost the same for arm C throughout the designs with classical DP, OIDP, and RDP. In fact, the efficacy estimation process in the range of scenarios $(0.5, 0), (0.5, 0.1), ..., (0.5, 0.4)$ for arm C has the same dispersion distribution because of having a higher success probability, whilst the estimation distribution for arm D depends on the level of adaptiveness in the given design. For instance, the range of estimation distribution in the design with classical DP is wider than OIDP and RDP designs, whilst the design with the RDP procedure tends to show a less dispersed estimation distribution. Furthermore, for this range of scenarios, the second quartile, i.e. median, is estimated lower than the actual efficacy, which can be interpreted as a biased estimator. However, as OIDP and RDP designs move away from the baseline of adaptiveness, median estimation becomes closer to the actual efficacies. Note that the same trend in dispersion distribution of estimation process repeats in the range of scenarios $(0.5, 0.6), (0.5, 0.7), ..., (0.5, 1)$ if one replaces one arm with another. Focusing on arm C estimation dispersion in scenario $(0.5, 0.8)$, the passive-aggressive performance of classical DP causes the third quartile to comply with actual efficacy. In fact, the passive-aggressive property in the DP procedure estimates MLE in up to $\%75$ cases less than actual efficacy. However, the third quartile in OIDP and RDP procedures moves above the actual efficacy, resulting in less baised estimation. Hence, it can be concluded that applying modifications, i.e. OIDP and amending the degree of randomisation, i.e. RDP, leads to less baised estimation and less dispersed estimation distribution.

Scenario $(0.5, 0.5)$ returns the same dispersion distribution in all designs presented in figure 6 for both arms. However, the estimation distribution range reduces as the design moves away from the adaptiveness baseline (see figure 5.6 part a to c). Note that estimating efficacy using the EFR procedure results in the same dispersion distribution for both arms and the whole range of scenarios. This mainly occurs because, in the EFR procedure, allocation ratios are pre-determined, and therefore, the estimation process is unbiased (second quartiles comply with actual efficacies).

5.6 Discussion

The EFR randomisation procedure gives rise to an unbiased estimate of the intervention efficacy because of the fixed randomisation ratio and the pre-determined sample size. However, it is not an efficient procedure design when there is a limitation on in-trial subjects such as clinical trials designed for rare diseases. As an alternative, the response-adaptive randomisation procedure allows for learning about the intervention efficacy from up-to-date responses and therefore adjusting allocation ratios in favour of superior interventions accordingly.

Response-adaptive randomisation designs are amongst sequential decision-making procedures. These designs can be modelled by a classical two-armed bandit problem that exemplifies the tradeoff between *exploration* or *learning* vs. *exploitation* or *earning*. In this chapter, using a Bayesian-Bernoulli two-armed problem with binary responses, we propose a response-adaptive design to estimate interventions' efficacy by the Maximum Likelihood Estimator (MLE). However, due to the passive-aggressive feature of dynamic programming in solving the model, MLE, which is calculated based on dynamic programming solutions, tends to be heavily biased. To mitigate this bias in the intervention estimation process, we propose two novel allocation procedures by modifying the allocation decision at every time step.

In section 5.2, we evaluated a modified dynamic programming procedure, where an augmented estimator is obtained by adding a number of pseudo-successes to the inferior intervention is used, performed for different variations in pseudo observations. Results showed that larger values for pseudo observations give rise to smaller values for estimation bias. However, a larger selection may result in missing subject benefit at the end of the trial. Hence, identifying the trade-off between the loss of subject benefit and bias mitigation, and therefore setting a suitable level of pseudo-observation to satisfy this trade-off, depends on the trialist and researcher's criteria.

In section 5.3, we applied a randomisation framework to the DP algorithm to develop RDP, which perturbs the Bayes-optimal allocation decision with a given probability called the degree of randomisation. Then, we showed that RDP could be interpreted as a bi-level randomisation where two parallel branches of the trial, a *fixed branch* and an *adaptive branch* are running at the same time, and each subject can be directed to either branch with pre-determined probabilities. Although the idea of RDP and its being equivalent to bi-level randomisation procedure has been proposed by Williamson et al. (2022) for trials considering delay and normally distributed response, we developed RDP for trials where responses are observable immediately and follow Bernoulli's distribution. We provided the simulation results for different randomisation degrees ($p = 0.6, 0.7, 0.8, 0.9$) and different trial sizes ($T = 60, 120, 180, 240$), all of which demonstrate a considerable reduction in bias, whilst the loss of subject benefit is negligible compared to OIDPs counterparts.

We used Julia programming language to code up all the models and proposed designs on a laptop with 16 GB RAM. Julia is an efficient programming language where syntaxes are implemented to allow for solving large trials via the DP approach. For details, see (Jacko, 2019a). We used the same codes for all computations from chapter 4. Furthermore, four trial sizes ($T = 60, 120, 180, 240$) together with 1 million simulation replications are assumed throughout all computations in this paper.

Inverse Probability Weighting (IPW) estimator and its normalised version are investigated in section 5.4, where the RDP procedure with $p = 0.9$ were used to estimate the IPW estimator for all trial sizes. Simulation results returned an unbiased estimator for all designs because, in the IPW estimation process, success responses are weighted by allocation probabilities.

Limitations. RAR designs have been amongst hot topics and the centre of data scientists' attention over the last two decades, whilst they are not fully applicable in all trial contexts and real-world trials. For instance, the assumption that subject response will be instantly observable is not entirely realistic. Moreover, in a Bayesian setting, evaluating frequentist operating characteristics like confidence intervals and type I error can not be translated accordingly. In reality, trialists, pharmaceutical companies and FDA are more interested in having insight into frequentist characteristics than Bayesian counterparts.

Future work. For extensions of this work, finding the optimal $\dot{s}$ in the OIDP procedure can be considered as a potential area to improve the trade-off between bias reduction and the loss of subject benefit. To do so, we need to investigate the DP algorithm mathematically and therefore modify the Bellman equations accordingly. Another extension can be applied to the bi-level procedure by filtering observations based on the branch they obtained. We mean defining an 8-dimensional vector of observations containing the success and failure responses on both arms for EFR and deterministic DP branches, i.e. $\boldsymbol{x} := \big(s_{C,EFR}, f_{C,EFR}, s_{D,EFR}, f_{D,EFR}, s_{C,DP}, f_{C,DP}, s_{D,DP}, f_{D,DP}\big)$. Then, based on posterior distribution calculated by the equation (3.1) using this vector, allocation policies can be formed in such a way that the MLE estimates have minimum bias and variability. Finally, all proposed procedures can be extended to the multi-arm models, particularly those in which covariates can be added as extra arms. However, computational complexity is a critical factor that often leads to implementation of trials with smaller sizes.

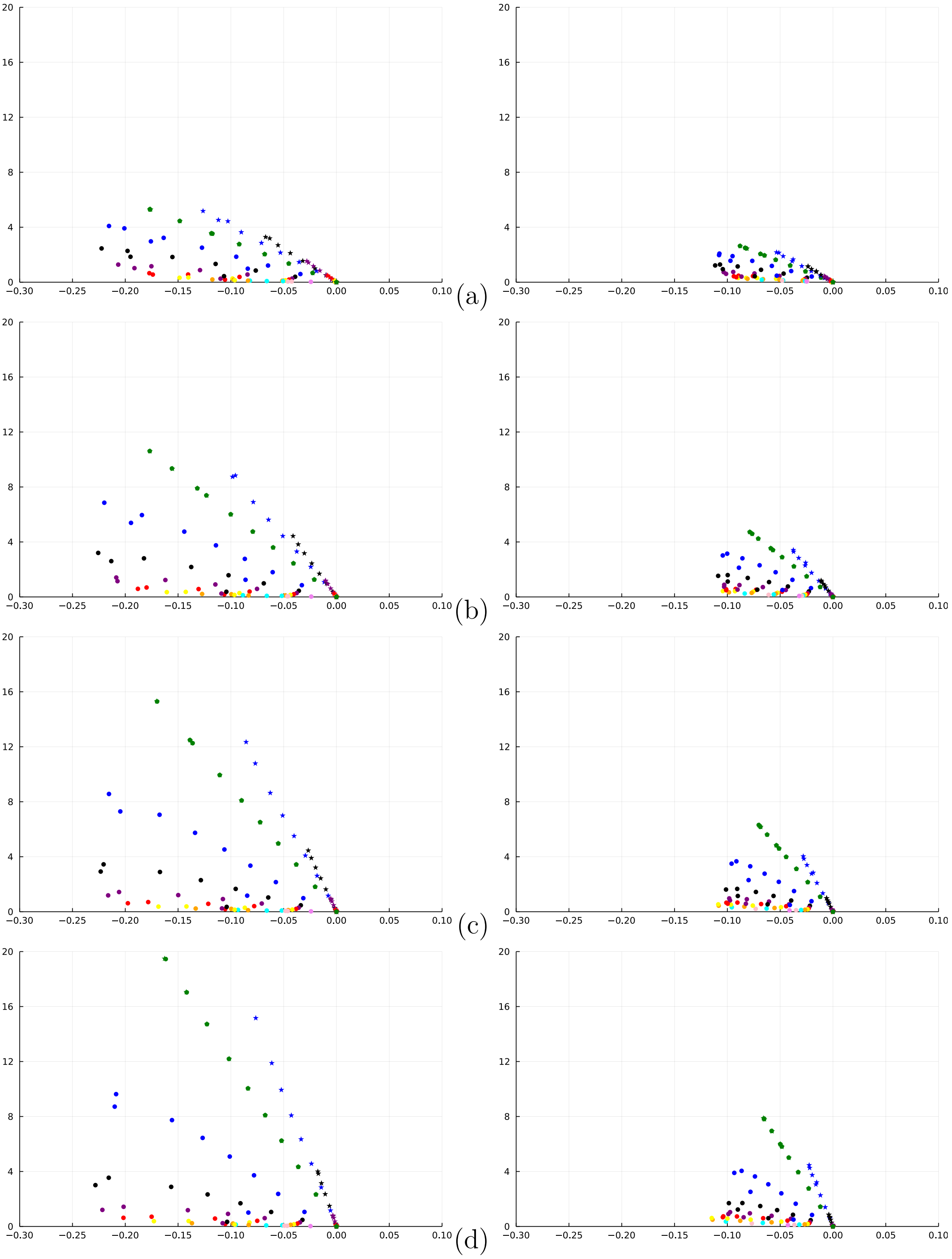


Figure 5.2: The estimation bias and covariance reduction comparison in the design with classical DP (left-hand side column) vs OIDP in which ($\dot{s} = 1$) (right-hand side column) for varying trial sizes: (a) $T = 60$ (b) $T = 120$ (c) $T = 180$ (d) $T = 240$. x-axis: Bias of MLE, y-axis: Covariance (MLE, Sample Size).

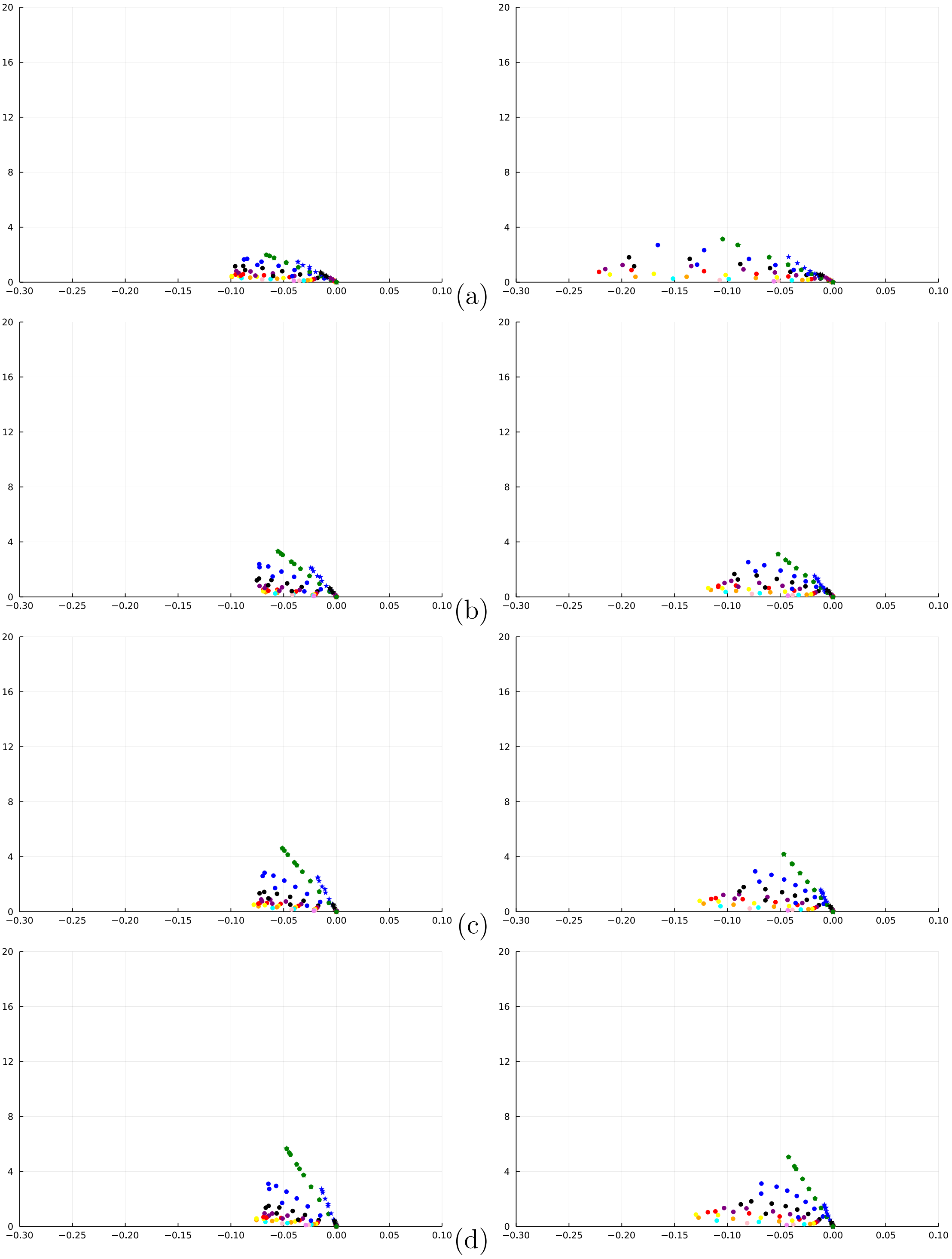


Figure 5.3: The estimation bias and covariance reduction comparison in the design with OIDP in which ($\dot{s} = 2$) (left-hand side column) vs ($\dot{s} = \ln(t)$) (right-hand side column) for varying trial sizes: (a) $T = 60$ (b) $T = 120$ (c) $T = 180$ (d) $T = 240$. x-axis: Bias of MLE, y-axis: Covariance (MLE, Sample Size).

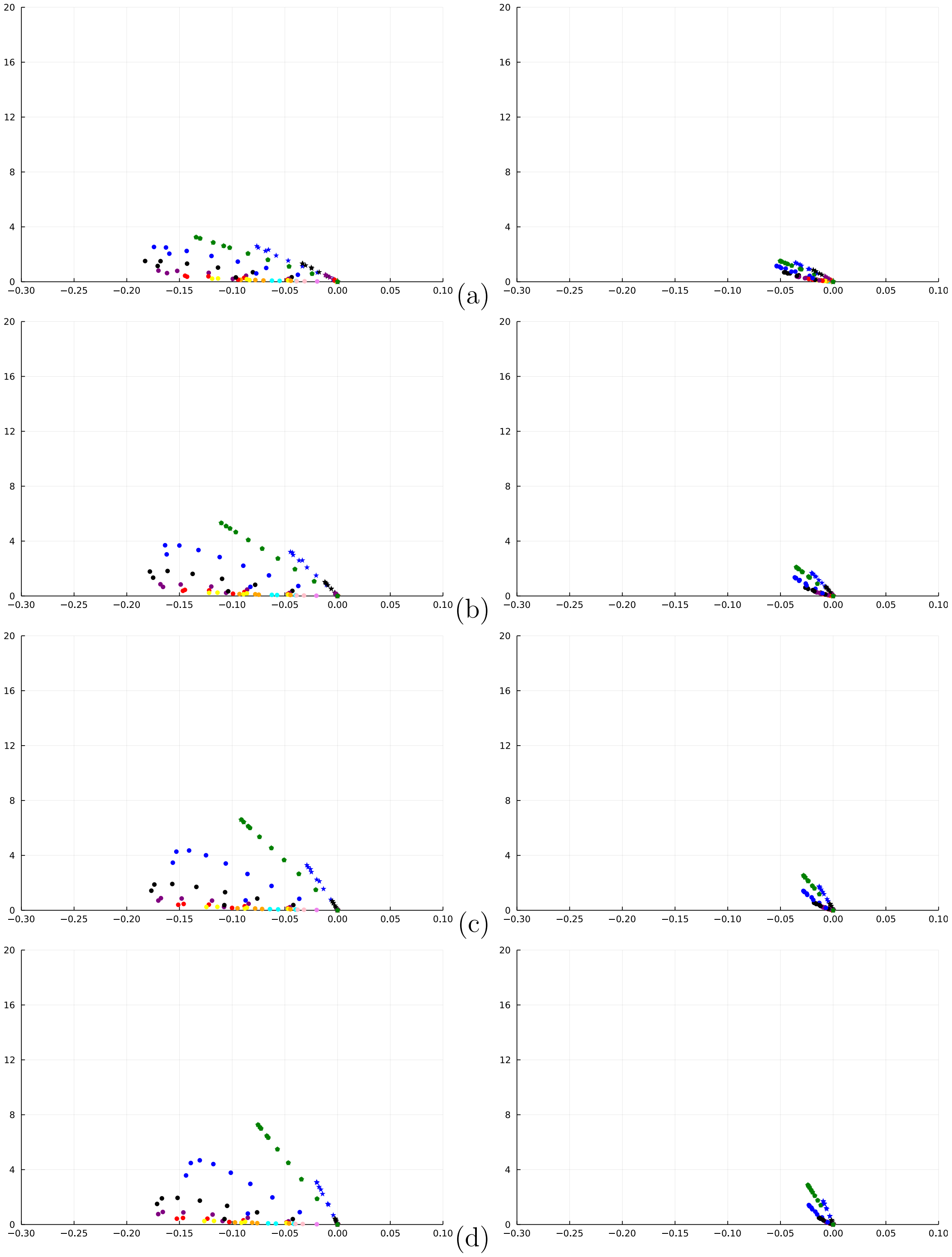


Figure 5.4: RDP procedure (bi-level randomisation): adaptive branch vs. pooled data: (a) $T = 60$ (b) $T = 120$ (c) $T = 180$ (d) $T = 240$. x-axis: Bias of MLE, y-axis: Covariance (MLE, Sample Size).

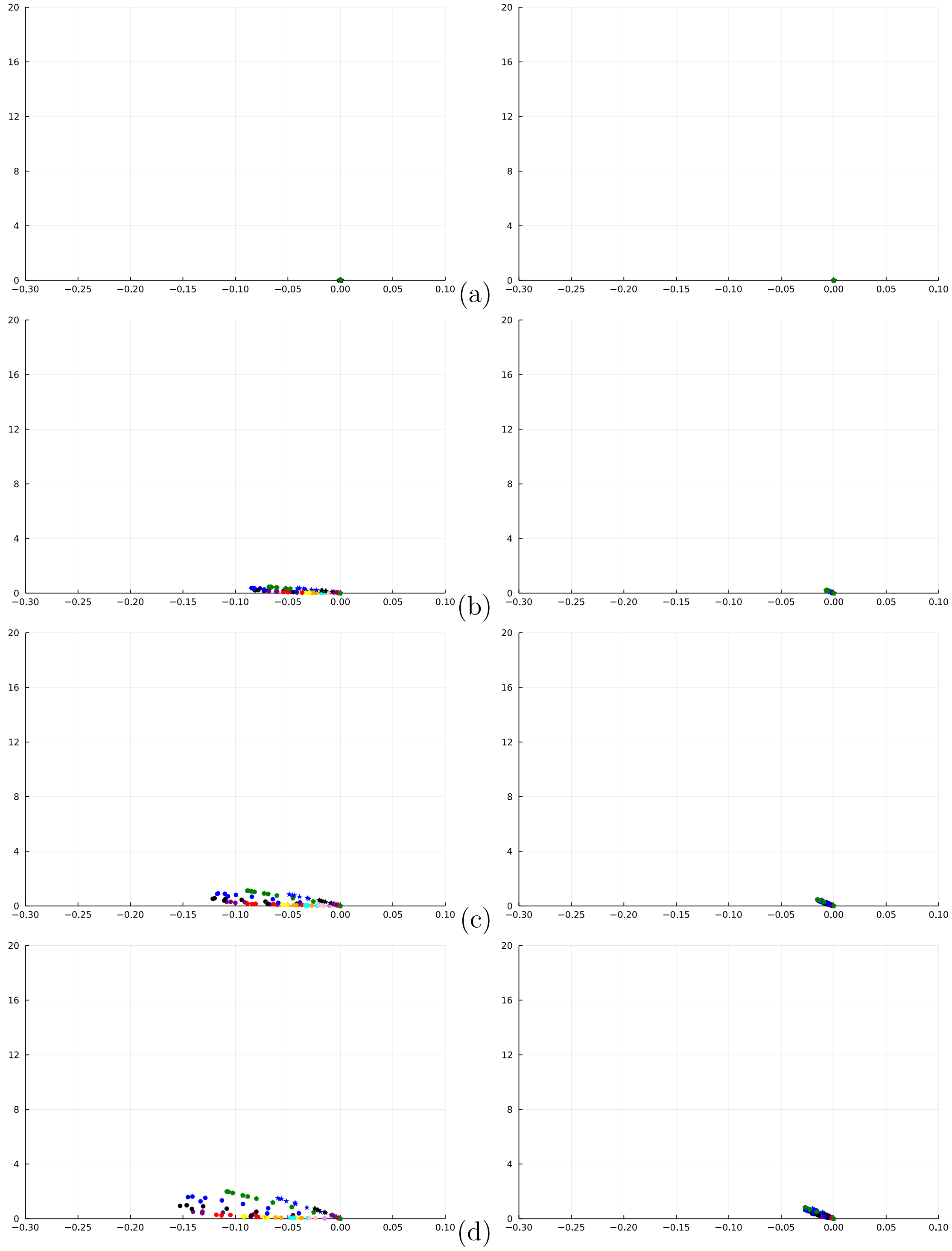


Figure 5.5: RDP procedure (bi-level randomisation) with different degrees of randomisation for $T = 60$: adaptive branch vs. pooled data: (a) $P = 0.5$ (b) $P = 0.6$ (c) $P = 0.7$ (d) $P = 0.8$. x-axis: Bias of MLE, y-axis: Covariance (MLE, Sample Size).

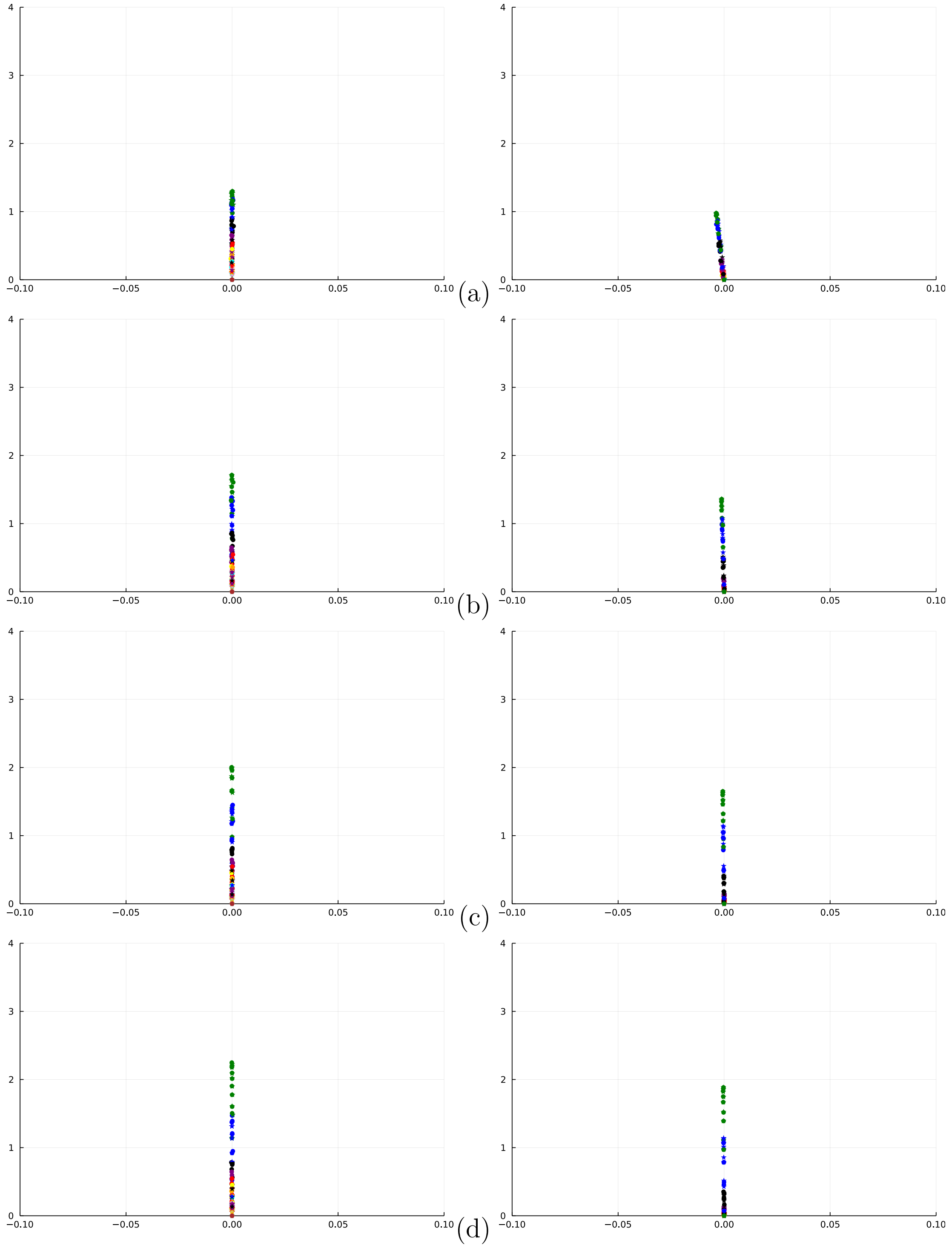


Figure 5.6: IPW vs. nIPW estimated in the designs using RDP procedure where $p = 0.9$: (a) $T = 60$ (b) $T = 120$ (c) $T = 180$ (d) $T = 240$. x-axis: Bias of the estimator, y-axis: Covariance (Estimator, Sample Size).

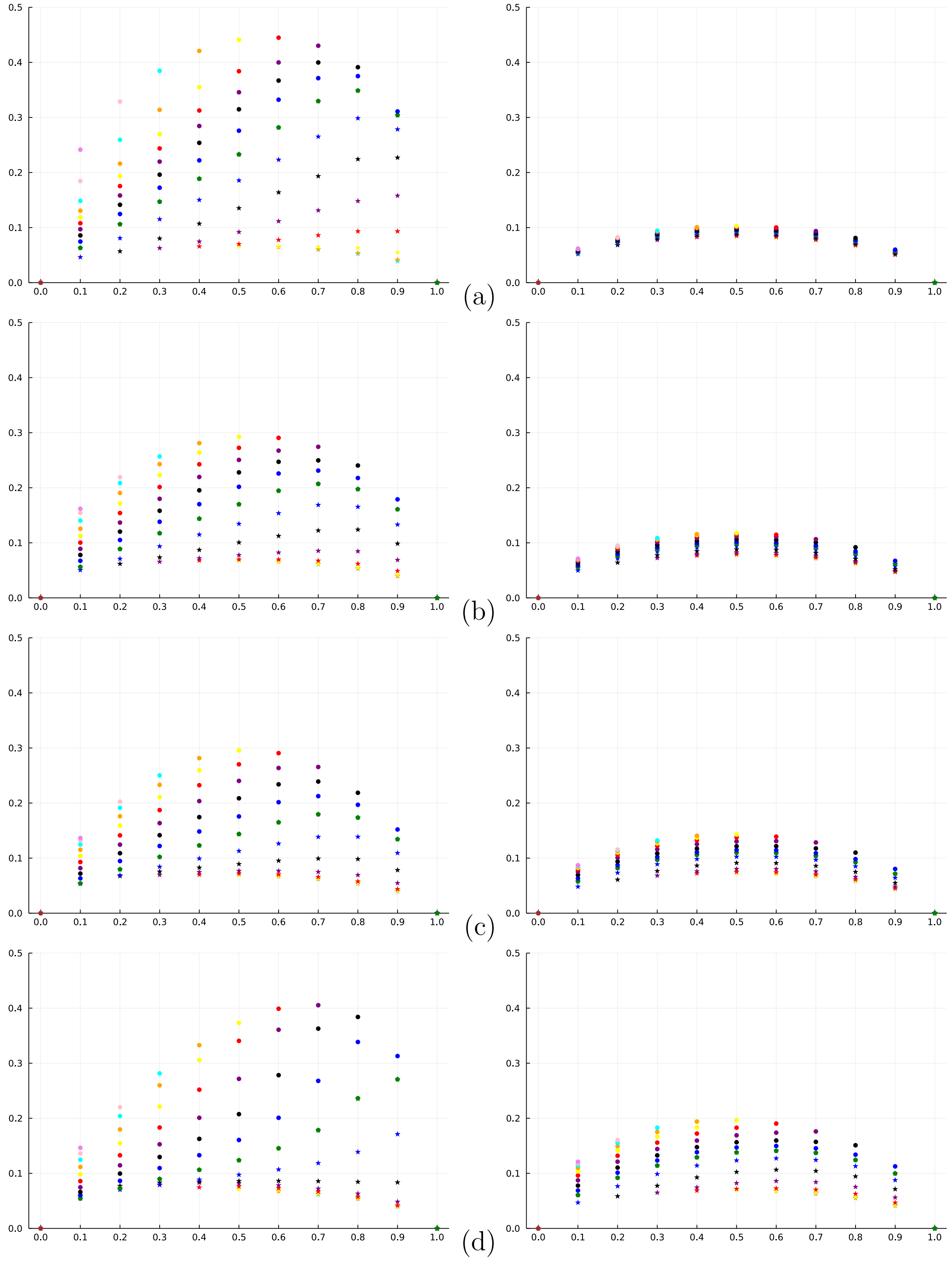


Figure 5.7: RMSE comparison for $T = 60$: OIDP designs with modifications vs. RDP designs with different degrees of randomisations. (a) $\dot{s} = 0$ vs $P = 0.6$ (b) $\dot{s} = 1$ vs $P = 0.7$ (c) $\dot{s} = 2$ vs $P = 0.8$ (d) $\dot{s} = ln(\tau + 1)$ vs $P = 0.9$. x-axis: Scenario, y-axis: RSME.

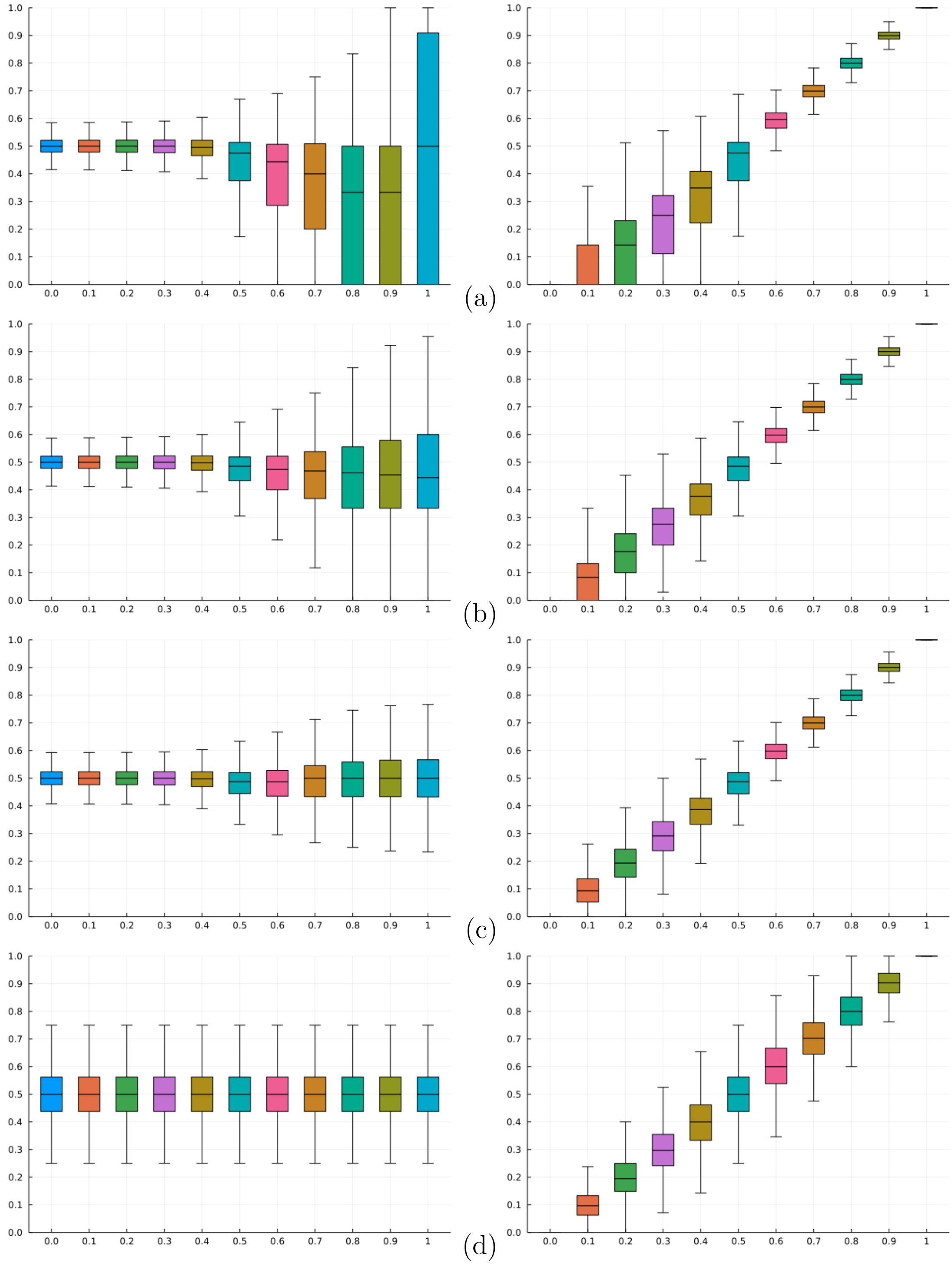


Figure 5.8: Box Plots comparison for $T = 240$: (a) Classical DP (b) OIDP with $\dot{s} = 2$ (c) RDP with $p = 0.9$ (d) EFR. x-axis: Scenario.

	Classical DP				OIDP ($\dot{s} = 1$)			
(θ_C, θ_D)	**T=60**	**T=120**	**T=180**	**T=240**	**T=60**	**T=120**	**T=180**	**T=240**
(0 , 0)	0.00 ± 0.00	0.00 ± 0.00	0.00 ± 0.00	0.00 ± 0.00	0.00 ± 0.00	0.00 ± 0.00	0.00 ± 0.00	0.00 ± 0.00
(0 , 0.1)	4.57 ± 2.46	10.14 ± 3.49	15.91 ± 4.23	21.74 ± 4.84	3.86 ± 2.47	9.17 ± 3.66	14.80 ± 4.39	20.53 ± 5.00
(0 , 0.2)	10.31 ± 3.34	21.94 ± 4.60	33.73 ± 5.56	45.56 ± 6.37	9.37 ± 3.52	20.78 ± 4.76	32.41 ± 5.73	44.15 ± 6.55
(0 , 0.3)	16.29 ± 3.80	33.96 ± 5.22	51.74 ± 6.33	69.59 ± 7.26	15.28 ± 3.95	32.68 ± 5.39	50.32 ± 6.50	68.06 ± 7.43
(0 , 0.4)	22.37 ± 4.04	46.05 ± 5.57	69.85 ± 6.75	93.72 ± 7.75	21.28 ± 4.19	44.70 ± 5.73	68.33 ± 6.91	92.08 ± 7.91
(0 , 0.5)	28.51 ± 4.11	58.22 ± 5.68	88.03 ± 6.88	117.90 ± 7.90	27.34 ± 4.27	56.77 ± 5.85	86.42 ± 7.04	116.18 ± 8.06
(0 , 0.6)	34.68 ± 4.02	70.42 ± 5.55	106.27 ± 6.73	142.16 ± 7.74	33.46 ± 4.20	68.94 ± 5.76	104.58 ± 6.93	140.34 ± 7.95
(0 , 0.7)	40.85 ± 3.73	82.68 ± 5.21	124.53 ± 6.31	166.42 ± 7.23	39.61 ± 3.94	81.19 ± 5.43	122.85 ± 6.54	164.65 ± 7.50
(0 , 0.8)	46.99 ± 3.20	94.92 ± 4.50	142.87 ± 5.49	190.82 ± 6.33	45.74 ± 3.43	93.46 ± 4.74	141.20 ± 5.73	189.09 ± 6.59
(0 , 0.9)	53.03 ± 2.36	107.02 ± 3.31	161.01 ± 4.05	215.01 ± 4.68	51.86 ± 2.58	105.71 ± 3.57	159.58 ± 4.33	213.53 ± 4.96
(0 , 1)	59.00 ± 0.00	119.00 ± 0.00	179.00 ± 0.00	239.00 ± 0.00	58.00 ± 0.00	118.00 ± 0.00	178.00 ± 0.00	238.00 ± 0.00
(0.1 , 0.1)	6.00 ± 2.33	12.00 ± 3.29	18.00 ± 4.02	24.00 ± 4.65	6.00 ± 2.32	12.00 ± 3.28	18.00 ± 4.02	24.00 ± 4.65
(0.1 , 0.2)	10.26 ± 3.36	21.39 ± 5.07	32.87 ± 6.39	44.51 ± 7.45	9.96 ± 3.22	20.90 ± 4.81	32.22 ± 6.01	43.76 ± 6.98
(0.1 , 0.3)	15.99 ± 4.12	33.49 ± 5.75	51.22 ± 6.87	69.04 ± 7.78	15.33 ± 3.98	32.52 ± 5.60	50.04 ± 6.73	67.64 ± 7.70
(0.1 , 0.4)	22.11 ± 4.33	45.77 ± 5.84	69.57 ± 6.97	93.40 ± 7.95	21.23 ± 4.29	44.54 ± 5.88	68.08 ± 7.09	91.69 ± 8.12
(0.1 , 0.5)	28.33 ± 4.30	58.05 ± 5.80	87.85 ± 6.98	117.71 ± 7.98	27.31 ± 4.36	56.65 ± 5.96	86.21 ± 7.21	115.85 ± 8.27
(0.1 , 0.6)	34.56 ± 4.13	70.31 ± 5.62	106.14 ± 6.79	142.02 ± 7.79	33.45 ± 4.27	68.85 ± 5.86	104.44 ± 7.06	140.11 ± 8.11
(0.1 , 0.7)	40.78 ± 3.80	82.59 ± 5.24	124.45 ± 6.35	166.34 ± 7.27	39.62 ± 3.98	81.13 ± 5.51	122.76 ± 6.65	164.49 ± 7.63
(0.1 , 0.8)	46.94 ± 3.25	94.89 ± 4.53	142.82 ± 5.51	190.76 ± 6.35	45.78 ± 3.47	93.43 ± 4.81	141.12 ± 5.81	188.96 ± 6.69
(0.1 , 0.9)	53.03 ± 2.39	107.02 ± 3.34	161.00 ± 4.07	214.99 ± 4.69	51.92 ± 2.60	105.71 ± 3.61	159.48 ± 4.37	213.39 ± 5.03
(0.1 , 1)	59.05 ± 0.22	119.05 ± 0.22	179.05 ± 0.22	239.05 ± 0.22	58.04 ± 0.24	117.99 ± 0.16	177.86 ± 0.36	237.84 ± 0.43
(0.2 , 0.2)	12.00 ± 3.09	24.00 ± 4.38	36.01 ± 5.37	48.00 ± 6.19	12.01 ± 3.10	24.00 ± 4.38	36.00 ± 5.36	48.00 ± 6.19
(0.2 , 0.3)	16.13 ± 3.89	33.01 ± 5.92	50.25 ± 7.58	67.70 ± 8.99	15.95 ± 3.73	32.75 ± 5.57	49.93 ± 7.05	67.30 ± 8.28
(0.2 , 0.4)	21.78 ± 4.59	45.12 ± 6.63	68.80 ± 8.00	92.59 ± 9.08	21.32 ± 4.32	44.44 ± 6.15	67.89 ± 7.42	91.50 ± 8.46
(0.2 , 0.5)	28.00 ± 4.71	57.66 ± 6.36	87.46 ± 7.51	117.31 ± 8.45	27.30 ± 4.47	56.60 ± 6.10	86.12 ± 7.32	115.74 ± 8.35
(0.2 , 0.6)	34.35 ± 4.46	70.10 ± 5.89	105.93 ± 7.02	141.79 ± 7.97	33.46 ± 4.34	68.84 ± 5.91	104.38 ± 7.11	140.03 ± 8.16
(0.2 , 0.7)	40.65 ± 3.99	82.47 ± 5.36	124.33 ± 6.43	166.21 ± 7.33	39.65 ± 4.02	81.13 ± 5.53	122.72 ± 6.67	164.40 ± 7.67
(0.2 , 0.8)	46.89 ± 3.35	94.82 ± 4.58	142.76 ± 5.56	190.69 ± 6.38	45.83 ± 3.48	93.44 ± 4.83	141.08 ± 5.83	188.84 ± 6.72
(0.2 , 0.9)	53.03 ± 2.44	107.00 ± 3.36	160.99 ± 4.10	214.97 ± 4.70	51.98 ± 2.60	105.70 ± 3.63	159.40 ± 4.39	213.27 ± 5.07
(0.2 , 1)	59.10 ± 0.30	119.10 ± 0.30	179.10 ± 0.30	239.10 ± 0.30	58.07 ± 0.36	117.96 ± 0.31	177.74 ± 0.52	237.67 ± 0.65
(0.3 , 0.3)	17.99 ± 3.55	36.00 ± 5.02	54.00 ± 6.15	72.00 ± 7.10	18.00 ± 3.55	36.01 ± 5.01	54.00 ± 6.14	72.00 ± 7.10
(0.3 , 0.4)	22.05 ± 4.20	44.80 ± 6.41	67.89 ± 8.26	91.23 ± 9.85	21.95 ± 4.04	44.67 ± 6.04	67.76 ± 7.68	91.08 ± 9.05
(0.3 , 0.5)	27.67 ± 4.87	56.91 ± 7.15	86.51 ± 8.76	116.32 ± 9.99	27.36 ± 4.51	56.42 ± 6.47	85.92 ± 7.83	115.52 ± 8.90
(0.3 , 0.6)	33.96 ± 4.90	69.62 ± 6.65	105.43 ± 7.85	141.33 ± 8.77	33.44 ± 4.48	68.78 ± 6.12	104.34 ± 7.28	139.98 ± 8.28
(0.3 , 0.7)	40.41 ± 4.40	82.23 ± 5.78	124.11 ± 6.77	166.00 ± 7.62	39.67 ± 4.11	81.12 ± 5.58	122.71 ± 6.69	164.38 ± 7.67
(0.3 , 0.8)	46.77 ± 3.60	94.71 ± 4.75	142.65 ± 5.65	190.58 ± 6.46	45.88 ± 3.51	93.44 ± 4.82	141.05 ± 5.83	188.80 ± 6.71
(0.3 , 0.9)	52.99 ± 2.56	106.98 ± 3.42	160.97 ± 4.13	214.95 ± 4.74	52.04 ± 2.60	105.71 ± 3.62	159.35 ± 4.40	213.17 ± 5.06
(0.3 , 1)	59.15 ± 0.36	119.15 ± 0.36	179.15 ± 0.36	239.15 ± 0.36	58.12 ± 0.44	117.94 ± 0.45	177.63 ± 0.67	237.51 ± 0.81
(0.4 , 0.4)	24.00 ± 3.79	48.00 ± 5.37	72.01 ± 6.57	95.99 ± 7.59	24.00 ± 3.80	48.00 ± 5.38	72.00 ± 6.57	95.99 ± 7.58
(0.4 , 0.5)	28.02 ± 4.35	56.72 ± 6.66	85.73 ± 8.60	114.96 ± 10.32	27.96 ± 4.18	56.68 ± 6.27	85.72 ± 7.98	114.98 ± 9.45
(0.4 , 0.6)	33.65 ± 4.96	68.83 ± 7.39	104.44 ± 9.12	140.18 ± 10.46	33.45 ± 4.53	68.57 ± 6.54	104.06 ± 7.90	139.70 ± 8.98
(0.4 , 0.7)	40.02 ± 4.86	81.73 ± 6.66	123.58 ± 7.84	165.47 ± 8.76	39.63 ± 4.29	81.07 ± 5.82	122.66 ± 6.90	164.35 ± 7.82
(0.4 , 0.8)	46.53 ± 4.09	94.47 ± 5.30	142.42 ± 6.13	190.38 ± 6.89	45.90 ± 3.63	93.46 ± 4.88	141.07 ± 5.85	188.81 ± 6.70
(0.4 , 0.9)	52.92 ± 2.84	106.92 ± 3.63	160.90 ± 4.26	214.89 ± 4.84	52.11 ± 2.63	105.73 ± 3.62	159.34 ± 4.39	213.13 ± 5.04
(0.4 , 1)	59.20 ± 0.40	119.20 ± 0.40	179.20 ± 0.40	239.20 ± 0.40	58.18 ± 0.51	117.93 ± 0.56	177.56 ± 0.82	237.39 ± 0.95
(0.5 , 0.5)	29.99 ± 3.87	60.00 ± 5.48	89.99 ± 6.72	119.99 ± 7.74	29.99 ± 3.88	60.01 ± 5.47	90.00 ± 6.71	120.01 ± 7.75
(0.5 , 0.6)	34.03 ± 4.35	68.72 ± 6.70	103.71 ± 8.69	138.91 ± 10.44	34.01 ± 4.18	68.73 ± 6.28	103.79 ± 8.01	139.07 ± 9.50
(0.5 , 0.7)	39.70 ± 4.88	80.92 ± 7.35	122.53 ± 9.10	164.31 ± 10.45	39.59 ± 4.37	80.81 ± 6.30	122.34 ± 7.63	164.02 ± 8.64
(0.5 , 0.8)	46.14 ± 4.57	93.96 ± 6.27	141.92 ± 7.33	189.87 ± 8.17	45.86 ± 3.85	93.43 ± 5.16	141.05 ± 6.08	188.79 ± 6.90
(0.5 , 0.9)	52.73 ± 3.38	106.75 ± 4.19	160.74 ± 4.77	214.75 ± 5.25	52.18 ± 2.73	105.80 ± 3.67	159.40 ± 4.41	213.16 ± 5.04
(0.5 , 1)	59.25 ± 0.44	119.25 ± 0.43	179.25 ± 0.43	239.25 ± 0.43	58.28 ± 0.56	117.96 ± 0.65	177.55 ± 0.95	237.34 ± 1.06
(0.6 , 0.6)	36.00 ± 3.80	72.00 ± 5.36	108.00 ± 6.57	143.99 ± 7.58	36.00 ± 3.79	71.99 ± 5.36	107.99 ± 6.58	143.99 ± 7.59
(0.6 , 0.7)	40.05 ± 4.21	80.76 ± 6.54	121.78 ± 8.52	163.04 ± 10.24	40.07 ± 4.00	80.85 ± 6.07	121.98 ± 7.74	163.32 ± 9.20
(0.6 , 0.8)	45.79 ± 4.61	93.16 ± 6.98	140.87 ± 8.67	188.73 ± 9.98	45.82 ± 3.98	93.18 ± 5.72	140.81 ± 6.83	188.56 ± 7.71
(0.6 , 0.9)	52.40 ± 3.96	106.33 ± 5.23	160.36 ± 5.98	214.37 ± 6.61	52.20 ± 2.96	105.84 ± 3.91	159.47 ± 4.60	213.26 ± 5.17
(0.6 , 1)	59.29 ± 0.48	119.29 ± 0.47	179.30 ± 0.46	239.30 ± 0.46	58.41 ± 0.60	118.06 ± 0.72	177.65 ± 1.02	237.40 ± 1.12
(0.7 , 0.7)	42.00 ± 3.55	84.00 ± 5.02	125.99 ± 6.15	168.01 ± 7.10	41.99 ± 3.56	84.00 ± 5.02	125.99 ± 6.15	168.00 ± 7.10
(0.7 , 0.8)	46.10 ± 3.90	92.88 ± 6.14	140.03 ± 8.06	187.39 ± 9.73	46.18 ± 3.65	93.11 ± 5.56	140.35 ± 7.11	187.82 ± 8.42
(0.7 , 0.9)	52.10 ± 4.05	105.55 ± 6.22	159.38 ± 7.63	213.36 ± 8.57	52.18 ± 3.16	105.74 ± 4.45	159.46 ± 5.25	213.26 ± 5.88
(0.7 , 1)	59.30 ± 0.62	119.33 ± 0.52	179.33 ± 0.51	239.34 ± 0.50	58.59 ± 0.63	118.24 ± 0.74	177.88 ± 1.01	237.63 ± 1.11
(0.8 , 0.8)	48.00 ± 3.10	96.01 ± 4.38	144.01 ± 5.36	192.00 ± 6.19	48.01 ± 3.10	96.01 ± 4.38	144.00 ± 5.37	192.00 ± 6.20
(0.8 , 0.9)	52.31 ± 3.32	105.23 ± 5.43	158.43 ± 7.27	211.84 ± 8.89	52.39 ± 2.96	105.61 ± 4.54	159.13 ± 5.75	212.77 ± 6.77
(0.8 , 1)	59.18 ± 1.20	119.28 ± 0.91	179.33 ± 0.68	239.34 ± 0.61	58.84 ± 0.64	118.53 ± 0.73	178.26 ± 0.90	238.04 ± 1.00
(0.9 , 0.9)	54.00 ± 2.32	107.99 ± 3.29	162.01 ± 4.02	216.00 ± 4.65	54.00 ± 2.32	107.99 ± 3.29	162.00 ± 4.02	216.00 ± 4.64
(0.9 , 1)	58.83 ± 1.70	118.55 ± 2.66	178.62 ± 2.93	238.79 ± 2.79	59.11 ± 0.68	118.94 ± 0.67	178.80 ± 0.72	238.66 ± 0.80
(1 , 1)	60.00 ± 0.00	120.00 ± 0.00	180.00 ± 0.00	240.00 ± 0.00	60.00 ± 0.00	120.00 ± 0.00	180.00 ± 0.00	240.00 ± 0.00

Table 5.2: The numerical results of comparing classical DP vs OIDP in which ($\dot{s} = 1$). Each cell is composed of the average number of success responses (first component) added to/subtracted from the corresponding standard deviation (second component) for each scenario (θ_C, θ_D) in all trial sizes $T = 60, 120, 180$, and 240

	OIDP [$\dot{s} = 2$]				**OIDP** [$\dot{s} = \ln(t)$]			
(θ_C, θ_D)	**T=60**	**T=120**	**T=180**	**T=240**	**T=60**	**T=120**	**T=180**	**T=240**
(0 , 0)	0.00 ± 0.00	0.00 ± 0.00	0.00 ± 0.00	0.00 ± 0.00	0.00 ± 0.00	0.00 ± 0.00	0.00 ± 0.00	0.00 ± 0.00
(0 , 0.1)	3.39 ± 2.32	8.27 ± 3.76	13.79 ± 4.58	19.50 ± 5.18	2.86 ± 2.09	6.56 ± 3.68	11.27 ± 4.93	16.58 ± 5.74
(0 , 0.2)	8.50 ± 3.64	19.80 ± 4.94	31.38 ± 5.88	43.06 ± 6.69	7.16 ± 3.79	17.78 ± 5.45	29.12 ± 6.32	40.62 ± 7.07
(0 , 0.3)	14.37 ± 4.16	31.70 ± 5.55	49.24 ± 6.65	66.94 ± 7.57	13.07 ± 4.61	30.11 ± 5.92	47.49 ± 6.97	65.00 ± 7.89
(0 , 0.4)	20.40 ± 4.39	43.70 ± 5.87	67.26 ± 7.07	90.93 ± 8.06	19.44 ± 4.77	42.54 ± 6.18	65.94 ± 7.34	89.48 ± 8.35
(0 , 0.5)	26.53 ± 4.47	55.77 ± 6.01	85.38 ± 7.23	115.05 ± 8.23	25.87 ± 4.73	55.01 ± 6.23	84.44 ± 7.45	114.01 ± 8.46
(0 , 0.6)	32.74 ± 4.42	67.98 ± 5.95	103.61 ± 7.11	139.31 ± 8.13	32.34 ± 4.61	67.51 ± 6.11	103.00 ± 7.30	138.57 ± 8.30
(0 , 0.7)	39.03 ± 4.18	80.28 ± 5.61	121.95 ± 6.68	163.67 ± 7.66	38.92 ± 4.41	80.18 ± 5.82	121.70 ± 6.93	163.27 ± 7.84
(0 , 0.8)	45.37 ± 3.66	92.57 ± 4.89	140.29 ± 5.84	188.10 ± 6.72	45.70 ± 3.97	92.97 ± 5.13	140.58 ± 6.16	188.20 ± 6.99
(0 , 0.9)	51.69 ± 2.74	104.81 ± 3.65	158.62 ± 4.39	212.54 ± 5.05	52.53 ± 2.96	105.57 ± 3.79	159.41 ± 4.58	213.22 ± 5.27
(0 , 1)	58.00 ± 0.00	117.00 ± 0.00	177.00 ± 0.00	237.00 ± 0.00	59.00 ± 0.00	118.00 ± 0.00	178.00 ± 0.00	238.00 ± 0.00
(0.1 , 0.1)	6.00 ± 2.32	12.00 ± 3.29	18.00 ± 4.03	24.00 ± 4.64	6.00 ± 2.32	11.99 ± 3.29	18.00 ± 4.02	24.00 ± 4.65
(0.1 , 0.2)	9.60 ± 3.15	20.26 ± 4.76	31.42 ± 5.98	42.82 ± 6.98	9.11 ± 3.10	18.97 ± 4.79	29.43 ± 6.16	40.35 ± 7.31
(0.1 , 0.3)	14.58 ± 4.05	31.53 ± 5.77	48.90 ± 6.94	66.45 ± 7.94	13.49 ± 4.21	29.70 ± 6.25	46.75 ± 7.56	64.02 ± 8.60
(0.1 , 0.4)	20.37 ± 4.48	43.50 ± 6.12	66.93 ± 7.36	90.48 ± 8.42	19.28 ± 4.89	42.04 ± 6.63	65.22 ± 7.91	88.58 ± 9.07
(0.1 , 0.5)	26.47 ± 4.59	55.63 ± 6.20	85.08 ± 7.51	114.66 ± 8.59	25.64 ± 4.98	54.58 ± 6.64	83.84 ± 7.99	113.23 ± 9.14
(0.1 , 0.6)	32.69 ± 4.52	67.88 ± 6.12	103.37 ± 7.35	139.00 ± 8.42	32.15 ± 4.84	67.20 ± 6.46	102.52 ± 7.78	137.95 ± 8.91
(0.1 , 0.7)	38.98 ± 4.26	80.21 ± 5.74	121.76 ± 6.91	163.47 ± 7.89	38.80 ± 4.57	79.94 ± 6.09	121.34 ± 7.31	162.86 ± 8.33
(0.1 , 0.8)	45.31 ± 3.71	92.51 ± 5.00	140.18 ± 6.03	187.93 ± 6.88	45.58 ± 4.09	92.78 ± 5.32	140.32 ± 6.44	187.88 ± 7.35
(0.1 , 0.9)	51.61 ± 2.77	104.78 ± 3.74	158.60 ± 4.53	212.39 ± 5.16	52.45 ± 3.04	105.41 ± 3.90	159.20 ± 4.78	212.98 ± 5.51
(0.1 , 1)	57.87 ± 0.36	117.02 ± 0.30	177.02 ± 0.30	236.82 ± 0.51	58.94 ± 0.26	117.84 ± 0.43	177.83 ± 0.47	237.82 ± 0.48
(0.2 , 0.2)	12.00 ± 3.10	24.00 ± 4.38	36.00 ± 5.37	48.00 ± 6.20	12.00 ± 3.10	24.00 ± 4.39	36.01 ± 5.36	48.00 ± 6.20
(0.2 , 0.3)	15.70 ± 3.66	32.30 ± 5.44	49.33 ± 6.87	66.64 ± 8.07	15.31 ± 3.63	31.33 ± 5.41	47.84 ± 6.88	64.69 ± 8.11
(0.2 , 0.4)	20.72 ± 4.31	43.60 ± 6.17	66.90 ± 7.47	90.39 ± 8.53	19.86 ± 4.44	42.03 ± 6.56	65.00 ± 7.98	88.22 ± 9.12
(0.2 , 0.5)	26.56 ± 4.59	55.63 ± 6.26	85.01 ± 7.55	114.55 ± 8.63	25.66 ± 4.91	54.37 ± 6.75	83.49 ± 8.09	112.78 ± 9.27
(0.2 , 0.6)	32.72 ± 4.54	67.87 ± 6.15	103.29 ± 7.41	138.86 ± 8.51	32.07 ± 4.89	67.00 ± 6.57	102.18 ± 7.94	137.53 ± 9.15
(0.2 , 0.7)	38.99 ± 4.26	80.18 ± 5.78	121.67 ± 7.00	163.31 ± 8.01	38.70 ± 4.63	79.76 ± 6.19	121.06 ± 7.49	162.46 ± 8.64
(0.2 , 0.8)	45.30 ± 3.71	92.50 ± 5.05	140.12 ± 6.15	187.78 ± 7.00	45.51 ± 4.12	92.60 ± 5.41	140.06 ± 6.63	187.59 ± 7.65
(0.2 , 0.9)	51.56 ± 2.78	104.78 ± 3.79	158.54 ± 4.63	212.23 ± 5.27	52.37 ± 3.08	105.24 ± 4.00	158.96 ± 4.94	212.71 ± 5.75
(0.2 , 1)	57.77 ± 0.50	117.00 ± 0.51	176.99 ± 0.53	236.65 ± 0.74	58.88 ± 0.40	117.67 ± 0.65	177.62 ± 0.77	237.58 ± 0.82
(0.3 , 0.3)	18.00 ± 3.55	36.00 ± 5.02	54.00 ± 6.15	71.99 ± 7.11	18.00 ± 3.55	36.00 ± 5.01	54.01 ± 6.15	71.99 ± 7.09
(0.3 , 0.4)	21.76 ± 3.97	44.36 ± 5.87	67.35 ± 7.42	90.58 ± 8.74	21.46 ± 3.94	43.62 ± 5.80	66.20 ± 7.32	89.07 ± 8.62
(0.3 , 0.5)	26.89 ± 4.44	55.75 ± 6.37	85.07 ± 7.72	114.57 ± 8.81	26.21 ± 4.54	54.49 ± 6.63	83.46 ± 8.10	112.70 ± 9.25
(0.3 , 0.6)	32.83 ± 4.53	67.94 ± 6.19	103.32 ± 7.42	138.88 ± 8.48	32.13 ± 4.82	66.92 ± 6.59	102.07 ± 7.93	137.38 ± 9.06
(0.3 , 0.7)	39.05 ± 4.26	80.26 ± 5.75	121.67 ± 6.97	163.27 ± 7.98	38.66 ± 4.63	79.67 ± 6.19	120.89 ± 7.48	162.27 ± 8.60
(0.3 , 0.8)	45.32 ± 3.69	92.53 ± 5.03	140.09 ± 6.14	187.70 ± 7.01	45.44 ± 4.11	92.47 ± 5.41	139.87 ± 6.63	187.34 ± 7.67
(0.3 , 0.9)	51.55 ± 2.77	104.78 ± 3.79	158.50 ± 4.65	212.10 ± 5.29	52.29 ± 3.08	105.10 ± 4.02	158.73 ± 5.00	212.44 ± 5.82
(0.3 , 1)	57.71 ± 0.61	116.97 ± 0.70	176.91 ± 0.77	236.47 ± 1.00	58.81 ± 0.53	117.50 ± 0.81	177.38 ± 1.07	237.29 ± 1.15
(0.4 , 0.4)	24.00 ± 3.79	48.01 ± 5.37	72.01 ± 6.57	96.01 ± 7.58	24.00 ± 3.79	48.00 ± 5.37	71.99 ± 6.58	96.01 ± 7.59
(0.4 , 0.5)	27.83 ± 4.11	56.44 ± 6.09	85.43 ± 7.70	114.66 ± 9.09	27.60 ± 4.07	55.85 ± 5.99	84.54 ± 7.53	113.51 ± 8.85
(0.4 , 0.6)	33.09 ± 4.43	68.01 ± 6.35	103.36 ± 7.67	138.91 ± 8.73	32.58 ± 4.51	67.06 ± 6.51	102.10 ± 7.93	137.40 ± 9.02
(0.4 , 0.7)	39.13 ± 4.27	80.33 ± 5.80	121.76 ± 6.97	163.36 ± 7.92	38.70 ± 4.56	79.62 ± 6.16	120.86 ± 7.40	162.20 ± 8.45
(0.4 , 0.8)	45.38 ± 3.69	92.63 ± 5.00	140.14 ± 6.08	187.70 ± 6.93	45.39 ± 4.10	92.39 ± 5.37	139.74 ± 6.53	187.19 ± 7.53
(0.4 , 0.9)	51.58 ± 2.76	104.84 ± 3.77	158.49 ± 4.64	212.02 ± 5.29	52.22 ± 3.07	105.01 ± 4.01	158.55 ± 4.96	212.20 ± 5.78
(0.4 , 1)	57.70 ± 0.70	116.96 ± 0.88	176.82 ± 1.00	236.29 ± 1.28	58.74 ± 0.63	117.36 ± 0.95	177.11 ± 1.36	236.97 ± 1.47
(0.5 , 0.5)	30.00 ± 3.87	60.00 ± 5.48	90.00 ± 6.71	119.98 ± 7.75	30.00 ± 3.87	60.00 ± 5.47	90.00 ± 6.72	120.00 ± 7.74
(0.5 , 0.6)	33.90 ± 4.09	68.56 ± 6.08	103.58 ± 7.70	138.84 ± 9.10	33.72 ± 4.07	68.13 ± 5.96	102.94 ± 7.50	137.98 ± 8.79
(0.5 , 0.7)	39.31 ± 4.23	80.36 ± 6.04	121.77 ± 7.30	163.36 ± 8.28	38.98 ± 4.33	79.68 ± 6.14	120.88 ± 7.41	162.25 ± 8.43
(0.5 , 0.8)	45.45 ± 3.77	92.73 ± 5.07	140.26 ± 6.10	187.81 ± 6.92	45.38 ± 4.06	92.35 ± 5.36	139.72 ± 6.44	187.18 ± 7.39
(0.5 , 0.9)	51.67 ± 2.77	104.97 ± 3.75	158.55 ± 4.60	212.06 ± 5.24	52.16 ± 3.06	104.98 ± 3.98	158.45 ± 4.90	212.05 ± 5.70
(0.5 , 1)	57.76 ± 0.78	117.02 ± 1.03	176.75 ± 1.22	236.17 ± 1.57	58.68 ± 0.68	117.28 ± 1.07	176.87 ± 1.61	236.67 ± 1.76
(0.6 , 0.6)	36.00 ± 3.80	72.01 ± 5.36	108.00 ± 6.57	144.01 ± 7.59	36.00 ± 3.79	72.01 ± 5.37	108.00 ± 6.57	144.00 ± 7.59
(0.6 , 0.7)	39.99 ± 3.92	80.74 ± 5.85	121.85 ± 7.42	163.17 ± 8.76	39.88 ± 3.91	80.45 ± 5.73	121.40 ± 7.19	162.58 ± 8.42
(0.6 , 0.8)	45.58 ± 3.81	92.80 ± 5.38	140.32 ± 6.45	187.93 ± 7.28	45.50 ± 3.93	92.41 ± 5.42	139.76 ± 6.49	187.27 ± 7.37
(0.6 , 0.9)	51.81 ± 2.86	105.16 ± 3.81	158.75 ± 4.59	212.29 ± 5.23	52.12 ± 3.07	105.04 ± 3.97	158.47 ± 4.82	212.10 ± 5.56
(0.6 , 1)	57.90 ± 0.84	117.18 ± 1.14	176.80 ± 1.36	236.23 ± 1.75	58.65 ± 0.69	117.30 ± 1.14	176.77 ± 1.74	236.50 ± 1.93
(0.7 , 0.7)	42.00 ± 3.55	83.99 ± 5.02	126.00 ± 6.14	167.99 ± 7.10	42.01 ± 3.55	84.00 ± 5.01	125.99 ± 6.15	168.00 ± 7.10
(0.7 , 0.8)	46.11 ± 3.54	93.03 ± 5.31	140.28 ± 6.75	187.72 ± 7.93	46.10 ± 3.56	92.85 ± 5.20	140.00 ± 6.50	187.39 ± 7.59
(0.7 , 0.9)	51.96 ± 2.97	105.38 ± 4.07	159.00 ± 4.84	212.64 ± 5.45	52.12 ± 3.08	105.18 ± 4.08	158.67 ± 4.86	212.31 ± 5.52
(0.7 , 1)	58.15 ± 0.85	117.50 ± 1.15	177.08 ± 1.37	236.61 ± 1.69	58.67 ± 0.65	117.49 ± 1.13	176.93 ± 1.66	236.62 ± 1.81
(0.8 , 0.8)	48.00 ± 3.10	96.00 ± 4.38	144.00 ± 5.37	192.00 ± 6.21	48.01 ± 3.10	96.00 ± 4.38	143.99 ± 5.36	192.01 ± 6.19
(0.8 , 0.9)	52.34 ± 2.84	105.54 ± 4.25	159.03 ± 5.35	212.67 ± 6.21	52.37 ± 2.88	105.43 ± 4.16	158.86 ± 5.15	212.48 ± 5.94
(0.8 , 1)	58.49 ± 0.81	117.99 ± 1.05	177.62 ± 1.21	237.30 ± 1.41	58.77 ± 0.58	117.88 ± 1.01	177.41 ± 1.37	237.09 ± 1.51
(0.9 , 0.9)	54.00 ± 2.32	108.00 ± 3.29	162.00 ± 4.03	216.01 ± 4.65	54.00 ± 2.32	108.00 ± 3.29	162.00 ± 4.03	216.00 ± 4.65
(0.9 , 1)	58.96 ± 0.70	118.66 ± 0.83	178.43 ± 0.92	238.24 ± 1.02	59.01 ± 0.55	118.49 ± 0.81	178.21 ± 0.99	237.98 ± 1.11
(1 , 1)	60.00 ± 0.00	120.00 ± 0.00	180.00 ± 0.00	240.00 ± 0.00	60.00 ± 0.00	120.00 ± 0.00	180.00 ± 0.00	240.00 ± 0.00

Table 5.3: The numerical results of comparing OIDP in which ($\dot{s} = 2$) vs ($\dot{s} = \ln(t+1)$). Each cell is composed of the average number of success responses (first component) added to/subtracted from the corresponding standard deviation (second component) for each scenario (θ_C, θ_D) in all trial sizes $T = 60, 120, 180,$ and 240

	RDP with $p = 0.9$				**RDP with different p values for $T = 60$**			
(θ_C, θ_D)	**T=60**	**T=120**	**T=180**	**T=240**	**p=0.5**	**p=0.6**	**p=0.7**	**p=0.8**
(0 , 0)	0.00 ± 0.00	0.00 ± 0.00	0.00 ± 0.00	0.00 ± 0.00	0.00 ± 0.00	0.00 ± 0.00	0.00 ± 0.00	0.00 ± 0.00
(0 , 0.1)	4.39 ± 2.36	9.62 ± 3.33	14.95 ± 4.03	20.31 ± 4.60	3.00 ± 1.69	3.41 ± 1.87	3.80 ± 2.05	4.14 ± 2.22
(0 , 0.2)	9.71 ± 3.17	20.45 ± 4.38	31.20 ± 5.29	41.97 ± 6.09	6.00 ± 2.32	7.02 ± 2.55	8.00 ± 2.76	8.92 ± 2.97
(0 , 0.3)	15.16 ± 3.61	31.33 ± 5.00	47.49 ± 6.07	63.68 ± 6.99	9.00 ± 2.76	10.62 ± 2.99	12.22 ± 3.21	13.75 ± 3.41
(0 , 0.4)	20.63 ± 3.87	42.21 ± 5.37	63.78 ± 6.54	85.40 ± 7.53	12.00 ± 3.08	14.22 ± 3.31	16.43 ± 3.51	18.60 ± 3.70
(0 , 0.5)	26.10 ± 3.97	53.08 ± 5.53	80.08 ± 6.75	107.07 ± 7.77	15.00 ± 3.33	17.83 ± 3.54	20.64 ± 3.71	23.43 ± 3.86
(0 , 0.6)	31.57 ± 3.95	63.96 ± 5.52	96.35 ± 6.74	128.76 ± 7.77	18.00 ± 3.53	21.43 ± 3.70	24.85 ± 3.82	28.25 ± 3.90
(0 , 0.7)	37.02 ± 3.80	74.80 ± 5.33	112.60 ± 6.52	150.42 ± 7.50	21.00 ± 3.66	25.02 ± 3.80	29.03 ± 3.86	33.05 ± 3.85
(0 , 0.8)	42.43 ± 3.50	85.63 ± 4.93	128.83 ± 6.04	172.04 ± 6.96	23.99 ± 3.75	28.62 ± 3.84	33.23 ± 3.82	37.84 ± 3.71
(0 , 0.9)	47.83 ± 3.03	96.44 ± 4.30	145.02 ± 5.26	193.63 ± 6.07	27.00 ± 3.80	32.22 ± 3.81	37.41 ± 3.69	42.63 ± 3.44
(0 , 1)	53.20 ± 2.29	107.20 ± 3.26	161.20 ± 4.00	215.20 ± 4.62	30.00 ± 3.81	35.80 ± 3.73	41.60 ± 3.49	47.40 ± 3.04
(0.1 , 0.1)	6.00 ± 2.32	12.00 ± 3.28	17.99 ± 4.03	24.01 ± 4.65	6.00 ± 2.32	6.00 ± 2.32	6.00 ± 2.32	6.00 ± 2.32
(0.1 , 0.2)	10.07 ± 3.23	20.86 ± 4.79	31.92 ± 5.96	43.10 ± 6.89	9.00 ± 2.77	9.30 ± 2.85	9.58 ± 2.96	9.83 ± 3.09
(0.1 , 0.3)	15.34 ± 3.89	31.90 ± 5.43	48.62 ± 6.51	65.37 ± 7.39	12.00 ± 3.10	12.89 ± 3.25	13.76 ± 3.44	14.59 ± 3.66
(0.1 , 0.4)	20.89 ± 4.11	43.02 ± 5.61	65.18 ± 6.74	87.34 ± 7.70	15.00 ± 3.35	16.54 ± 3.51	18.04 ± 3.68	19.52 ± 3.89
(0.1 , 0.5)	26.47 ± 4.13	54.04 ± 5.66	81.61 ± 6.85	109.20 ± 7.87	17.99 ± 3.54	20.17 ± 3.68	22.32 ± 3.83	24.44 ± 3.98
(0.1 , 0.6)	32.03 ± 4.03	65.01 ± 5.58	98.00 ± 6.78	130.98 ± 7.79	21.00 ± 3.67	23.79 ± 3.79	26.57 ± 3.89	29.32 ± 3.96
(0.1 , 0.7)	37.54 ± 3.83	75.91 ± 5.34	114.32 ± 6.50	152.70 ± 7.50	24.00 ± 3.77	27.40 ± 3.85	30.80 ± 3.88	34.19 ± 3.87
(0.1 , 0.8)	43.00 ± 3.49	86.79 ± 4.90	130.60 ± 5.98	174.39 ± 6.91	27.00 ± 3.82	31.01 ± 3.84	35.01 ± 3.80	39.01 ± 3.69
(0.1 , 0.9)	48.43 ± 2.99	97.63 ± 4.22	146.84 ± 5.17	196.02 ± 5.96	30.00 ± 3.83	34.61 ± 3.78	39.23 ± 3.64	43.84 ± 3.38
(0.1 , 1)	53.84 ± 2.19	108.45 ± 3.12	163.03 ± 3.82	217.63 ± 4.41	33.00 ± 3.80	38.21 ± 3.67	43.42 ± 3.39	48.64 ± 2.94
(0.2 , 0.2)	12.00 ± 3.10	24.01 ± 4.38	36.00 ± 5.37	48.01 ± 6.20	12.00 ± 3.10	12.00 ± 3.10	12.00 ± 3.10	12.00 ± 3.10
(0.2 , 0.3)	15.94 ± 3.74	32.52 ± 5.55	49.40 ± 6.98	66.43 ± 8.14	15.00 ± 3.35	15.25 ± 3.40	15.51 ± 3.49	15.73 ± 3.60
(0.2 , 0.4)	21.13 ± 4.29	43.55 ± 6.09	66.21 ± 7.30	88.93 ± 8.26	18.00 ± 3.55	18.82 ± 3.65	19.63 ± 3.81	20.41 ± 4.03
(0.2 , 0.5)	26.74 ± 4.40	54.82 ± 5.97	82.97 ± 7.11	111.14 ± 8.11	21.00 ± 3.69	22.48 ± 3.79	23.94 ± 3.94	25.38 ± 4.13
(0.2 , 0.6)	32.40 ± 4.22	65.96 ± 5.73	99.53 ± 6.89	133.11 ± 7.89	24.00 ± 3.79	26.13 ± 3.86	28.25 ± 3.94	30.36 ± 4.06
(0.2 , 0.7)	38.01 ± 3.93	76.98 ± 5.38	115.98 ± 6.53	154.93 ± 7.51	27.00 ± 3.84	29.77 ± 3.87	32.53 ± 3.89	35.29 ± 3.91
(0.2 , 0.8)	43.55 ± 3.52	87.93 ± 4.89	132.32 ± 5.95	176.70 ± 6.86	30.00 ± 3.85	33.40 ± 3.83	36.79 ± 3.76	40.18 ± 3.66
(0.2 , 0.9)	49.03 ± 2.95	98.82 ± 4.14	148.62 ± 5.06	198.41 ± 5.85	33.00 ± 3.82	37.01 ± 3.73	41.02 ± 3.56	45.02 ± 3.32
(0.2 , 1)	54.48 ± 2.09	109.68 ± 2.96	164.87 ± 3.63	220.07 ± 4.19	36.00 ± 3.75	40.62 ± 3.57	45.24 ± 3.27	49.85 ± 2.81
(0.3 , 0.3)	18.00 ± 3.55	36.00 ± 5.02	54.00 ± 6.16	72.00 ± 7.09	18.00 ± 3.55	18.00 ± 3.55	18.00 ± 3.55	18.01 ± 3.55
(0.3 , 0.4)	21.87 ± 4.03	44.35 ± 5.98	67.11 ± 7.52	90.04 ± 8.81	21.00 ± 3.69	21.24 ± 3.73	21.47 ± 3.80	21.67 ± 3.90
(0.3 , 0.5)	27.03 ± 4.49	55.34 ± 6.41	83.97 ± 7.69	112.64 ± 8.71	24.01 ± 3.79	24.80 ± 3.86	25.57 ± 3.99	26.32 ± 4.20
(0.3 , 0.6)	32.67 ± 4.47	66.74 ± 6.06	100.89 ± 7.17	135.03 ± 8.14	27.00 ± 3.84	28.46 ± 3.90	29.91 ± 4.01	31.31 ± 4.18
(0.3 , 0.7)	38.40 ± 4.11	77.95 ± 5.53	117.52 ± 6.62	157.08 ± 7.56	30.00 ± 3.87	32.13 ± 3.87	34.25 ± 3.91	36.34 ± 3.97
(0.3 , 0.8)	44.04 ± 3.60	89.02 ± 4.91	134.02 ± 5.95	178.97 ± 6.83	33.01 ± 3.84	35.78 ± 3.79	38.54 ± 3.73	41.30 ± 3.66
(0.3 , 0.9)	49.60 ± 2.94	100.00 ± 4.09	150.39 ± 4.97	200.78 ± 5.73	35.99 ± 3.77	39.41 ± 3.65	42.80 ± 3.48	46.21 ± 3.25
(0.3 , 1)	55.12 ± 1.97	110.91 ± 2.79	166.72 ± 3.42	222.50 ± 3.95	39.00 ± 3.66	43.03 ± 3.44	47.06 ± 3.12	51.08 ± 2.67
(0.4 , 0.4)	23.99 ± 3.79	48.01 ± 5.37	72.00 ± 6.58	96.00 ± 7.59	24.01 ± 3.79	24.00 ± 3.79	24.01 ± 3.80	24.01 ± 3.80
(0.4 , 0.5)	27.86 ± 4.16	56.27 ± 6.16	84.99 ± 7.76	113.90 ± 9.08	27.00 ± 3.85	27.23 ± 3.87	27.45 ± 3.93	27.66 ± 4.02
(0.4 , 0.6)	33.00 ± 4.50	67.30 ± 6.45	101.90 ± 7.72	136.58 ± 8.72	30.00 ± 3.87	30.79 ± 3.91	31.56 ± 4.02	32.30 ± 4.20
(0.4 , 0.7)	38.70 ± 4.33	78.77 ± 5.84	118.93 ± 6.88	159.07 ± 7.78	33.00 ± 3.85	34.46 ± 3.85	35.91 ± 3.91	37.33 ± 4.05
(0.4 , 0.8)	44.48 ± 3.76	90.05 ± 5.02	135.61 ± 6.00	181.17 ± 6.84	36.00 ± 3.78	38.14 ± 3.72	40.28 ± 3.68	42.39 ± 3.68
(0.4 , 0.9)	50.16 ± 2.96	101.14 ± 4.04	152.13 ± 4.90	203.11 ± 5.62	39.01 ± 3.68	41.80 ± 3.53	44.59 ± 3.37	47.37 ± 3.17
(0.4 , 1)	55.75 ± 1.85	112.14 ± 2.61	168.55 ± 3.19	224.93 ± 3.69	42.00 ± 3.52	45.43 ± 3.28	48.87 ± 2.95	52.31 ± 2.50
(0.5 , 0.5)	30.00 ± 3.87	60.00 ± 5.48	90.00 ± 6.71	120.00 ± 7.75	30.00 ± 3.87	30.00 ± 3.88	30.00 ± 3.87	30.00 ± 3.87
(0.5 , 0.6)	33.86 ± 4.14	68.28 ± 6.13	103.02 ± 7.70	137.89 ± 9.02	33.00 ± 3.85	33.23 ± 3.87	33.45 ± 3.92	33.65 ± 4.01
(0.5 , 0.7)	39.05 ± 4.35	79.38 ± 6.20	119.99 ± 7.40	160.67 ± 8.29	36.00 ± 3.79	36.80 ± 3.79	37.58 ± 3.88	38.34 ± 4.04
(0.5 , 0.8)	44.83 ± 3.95	90.93 ± 5.26	137.09 ± 6.18	183.21 ± 6.96	38.99 ± 3.69	40.49 ± 3.63	41.97 ± 3.63	43.43 ± 3.70
(0.5 , 0.9)	50.66 ± 3.06	102.25 ± 4.06	153.82 ± 4.86	205.36 ± 5.55	42.00 ± 3.54	44.17 ± 3.39	46.35 ± 3.25	48.51 ± 3.12
(0.5 , 1)	56.39 ± 1.71	113.38 ± 2.40	170.38 ± 2.93	227.36 ± 3.39	45.00 ± 3.34	47.83 ± 3.07	50.68 ± 2.75	53.53 ± 2.32
(0.6 , 0.6)	36.00 ± 3.80	72.00 ± 5.36	108.00 ± 6.56	144.01 ± 7.60	36.00 ± 3.80	36.00 ± 3.79	36.00 ± 3.79	36.00 ± 3.79
(0.6 , 0.7)	39.89 ± 3.96	80.37 ± 5.87	121.11 ± 7.35	162.03 ± 8.57	39.00 ± 3.69	39.24 ± 3.70	39.47 ± 3.74	39.69 ± 3.82
(0.6 , 0.8)	45.19 ± 3.99	91.64 ± 5.58	138.27 ± 6.62	184.96 ± 7.36	42.00 ± 3.55	42.83 ± 3.51	43.66 ± 3.54	44.45 ± 3.68
(0.6 , 0.9)	51.09 ± 3.21	103.26 ± 4.17	155.42 ± 4.90	207.52 ± 5.55	45.01 ± 3.34	46.55 ± 3.21	48.09 ± 3.11	49.61 ± 3.08
(0.6 , 1)	57.02 ± 1.55	114.60 ± 2.17	172.20 ± 2.64	229.77 ± 3.06	48.00 ± 3.08	50.24 ± 2.82	52.49 ± 2.50	54.75 ± 2.10
(0.7 , 0.7)	42.00 ± 3.54	84.00 ± 5.02	126.00 ± 6.15	168.00 ± 7.09	42.00 ± 3.54	42.01 ± 3.55	42.00 ± 3.55	42.00 ± 3.55
(0.7 , 0.8)	45.97 ± 3.61	92.56 ± 5.34	139.42 ± 6.65	186.39 ± 7.68	45.00 ± 3.35	45.26 ± 3.34	45.51 ± 3.37	45.75 ± 3.45
(0.7 , 0.9)	51.49 ± 3.27	104.12 ± 4.43	156.83 ± 5.12	209.49 ± 5.70	48.00 ± 3.10	48.91 ± 2.99	49.80 ± 2.95	50.67 ± 3.01
(0.7 , 1)	57.63 ± 1.38	115.81 ± 1.91	173.99 ± 2.32	232.16 ± 2.68	50.99 ± 2.76	52.64 ± 2.50	54.29 ± 2.20	55.95 ± 1.85
(0.8 , 0.8)	48.00 ± 3.10	96.00 ± 4.38	143.99 ± 5.36	192.00 ± 6.19	48.00 ± 3.10	48.00 ± 3.10	48.00 ± 3.09	48.00 ± 3.10
(0.8 , 0.9)	52.14 ± 2.96	104.95 ± 4.35	158.00 ± 5.31	211.14 ± 6.01	51.00 ± 2.77	51.31 ± 2.71	51.61 ± 2.72	51.89 ± 2.78
(0.8 , 1)	58.18 ± 1.25	116.97 ± 1.64	175.74 ± 1.97	234.49 ± 2.27	54.00 ± 2.32	55.03 ± 2.08	56.08 ± 1.83	57.13 ± 1.55
(0.9 , 0.9)	54.00 ± 2.33	108.00 ± 3.29	162.00 ± 4.02	216.00 ± 4.65	54.00 ± 2.32	54.01 ± 2.32	54.00 ± 2.32	54.00 ± 2.32
(0.9 , 1)	58.64 ± 1.22	118.02 ± 1.43	177.39 ± 1.60	236.68 ± 1.90	57.00 ± 1.69	57.43 ± 1.50	57.88 ± 1.33	58.29 ± 1.19
(1 , 1)	60.00 ± 0.00	120.00 ± 0.00	180.00 ± 0.00	240.00 ± 0.00	60.00 ± 0.00	60.00 ± 0.00	60.00 ± 0.00	60.00 ± 0.00

Table 5.4: The numerical results for RDP (bi-level randomisation) in which $p = 0.9$ for all trial sizes (right-hand side column) and RDP (bi-level randomisation) in which $T = 60$ for different degrees of randomisation (left-hand side column). Each cell is composed of the average number of success responses (first component) added to/subtracted from the corresponding standard deviation (second component) for each scenario (θ_C, θ_D)

Chapter 6

Extension to Trials with Early Stopping

6.1 Introduction

Temporarily stopping a trial to perform planned interim analyses for safety, efficacy, or futility is relatively common in the RAR procedures literature. The term 'futility' is used to refer to the inability of a clinical trial to achieve its objectives. Frequentist hypothesis testing, where the efficacy of an experimental arm is compared with the control one, is a crucial factor in formulating an interim analysis. Bayesian testing approaches can also be allowed for interim studies with more exact precision (Berry, 2005). However, providing a test statistic using the Bayes factor is not as straightforward as in the frequentist counterpart (Pham-Gia et al., 2017). Another essential element in the designs with interim analysis is setting up appropriate stopping criteria by which a statistical inference can be firmly obtained.

From a more general point of view, sequential designs with interim inspections are not only convenient to implement in practice but also raise the possibility that a trial needs fewer subjects to reach a statistically significance test result (Jennison and Turnbull, 1999; Wassmer and Brannath, 2016). However, recruiting less subjects may result in efficacy estimation with either positive or negative bias. Since the lack of observations gives rise to estimation with bias (even in RCT designs with potential early stopping) Bauer et al. (2010) broadly discussed the trade-off between the selection scheme determined in advance and the reporting bias to overcome this sparsity of observed data in the trial designs with a planned interim analysis. On the other side of the related literature, the sign and magnitude of the bias, together with some certain natural monotonicity properties of conditional bias of the rewards, e.g. the sample mean of each arm in MAB experiments, have been thoroughly examined in Shin et al. (2019a,b), and Shin et al. (2020), respectively. The authors of all three contributions show that the sign of the bias can potentially depend on stopping rules as well as sampling and choosing rules that are defined adaptively. Long before this, Starr and Woodroofe (1968) also showed that under specific stopping circumstances, the sample mean can be positively biased regardless of the actual efficacy. Last but not least, simulation results in the study by Jiang et al. (2017) show that in Bayesian RAR settings, the Type I error rate can be inflated depending on the frequency of implementing an interim analysis. Hence, the critical boundaries should be carefully chosen to preserve the Type I error rate for a given efficacy.

In this chapter, we first formulate a one-sided Frequentist hypothesis testing where up-to-date arms' MLEs are compared in our Bayesian-Bernoulli two-armed problem. Then we develop some stopping criteria upon which a trial can be interrupted for a non-trivial interim analysis in the middle of the time horizon. It is worth mentioning that we assume two different model cases: (i) **Dynamic Programming without Interim (*DPWOI*)**: setting an interim analysis through the simulation step whilst the classical DP solutions are used for estimation, (ii) **Dynamic Programming with Interim (*DPWI*)**: setting an identical interim inspection in both the DP and simulation step. Estimation results in section 6.5 are presented in three different trial categories for each model case. Also, we show that the frequency of the false early stopping is inflated above the significance level of the test in section 6.6 followed by some notes on subject benefits acquired from model cases mentioned above.

6.2 Model

Recalling the backward induction algorithm used for solving the Bayesian-Bernoulli two-armed model, see section 2.3.3, together with the fact that the model converts to the probability tree when the allocation procedure implements with only one arm, we run a hypothesis testing upon satisfying early stopping criteria once the trial reaches the middle, i.e. $t = T/2$. The early stopping criteria and the pertinent developmental process by which the intended requirements are getting satisfied will be described in section 6.4.3. Now, suppose the trial reaches the state $\boldsymbol{x}(t)$ when $t = T/2$ so that all possible combinations can be enumerated by $s_C(t) + f_C(t) + s_D(t) + f_D(t) = t$. For those combinations satisfying the early stopping criteria, the trial enters the hypothesis-testing phase, where it can be terminated upon rejecting the null hypothesis or continue otherwise. It is evident that rejecting the null hypothesis depends on the predetermined significance level

α and, subsequently, the critical value z_α for the one-sided test as we compare up-to-date arms' MLE. After entering the hypothesis-testing phase, the trial can be directed towards two possible routes depending on the calculated z-statistic value formulated in section 6.4.1.

Route 1: the null hypothesis fails to reject; therefore, the trial continues to the end, and the classical DP procedure can be applied accordingly.

Route 2: the null hypothesis is rejected, and the trial stops early.

Now, we need to set up well-defined quantities for the corresponding value functions at $t = T/2$. To do so, one can assume that stopping a trial early in favour of the superior arm is equivalent to dropping the inferior arm in the middle of the trial and continuing to allocate the superior one to the remaining subjects. Hence, based on the probability tree drawing from $t = T/2$ to $t = T$ for each arm, the appropriate quantity for the value functions can be the multiplication of the corresponding current beliefs, i.e. the Bayes-expected number of successes by the number of remaining subjects, i.e. half of the trial size. The whole procedure can be formulated as below:

For any given time epoch $t < T/2$ in all designs, and for all time epochs for those designs that fail to stop early, under an optimal policy, the expected total reward, i.e. the Bayes-expected number of successes, see section 3.1, can be calculated as follows:

$$\begin{aligned} F_t^C\Big(s_C, f_C, s_D, f_D\Big) =& q_{C,(x,i),1}\cdot\Big(1 + F_{t+1}\Big(s_C + 1, f_C, s_D, f_D\Big)\Big) \\ & + q_{C,(x,i),0}\cdot\Big(0 + F_{t+1}\Big(s_C, f_C + 1, s_D, f_D\Big)\Big) \\ F_t^D\Big(s_C, f_C, s_D, f_D\Big) =& q_{D,(x,i),1}\cdot\Big(1 + F_{t+1}\Big(s_C, f_C, s_D + 1, f_D\Big)\Big) \\ & + q_{D,(x,i),0}\cdot\Big(0 + F_{t+1}\Big(s_C, f_C, s_D, f_D + 1\Big)\Big)\ . \end{aligned} \tag{6.1}$$

However, for the designs with early stopping at time epoch $t = T/2$, we have:

$$\begin{aligned} F_t^C\Big(s_C, f_C, s_D, f_D\Big) =& q_{C,(x,i),1}\cdot\Big(T/2\Big) \\ F_t^D\Big(s_C, f_C, s_D, f_D\Big) =& q_{D,(x,i),1}\cdot\Big(T/2\Big) \end{aligned} \tag{6.2}$$

where $q_{k,(x,i),o}$ represents the posterior probability of observing response $o \in \mathcal{O}$ for arm k at time epoch t. See equation (3.1). Therefore, based on the principle of optimality if a trial stops early we have:

$$\begin{aligned} & F_t\Big(s_C, f_C, s_D, f_D\Big) = max\Big\{F_t^C\Big(s_C, f_C, s_D, f_D\Big), F_t^D\Big(s_C, f_C, s_D, f_D\Big)\Big\} \quad \text{for } 0 \leq t \leq T/2, \\ & F_t\Big(s_C, f_C, s_D, f_D\Big) = 0, \text{otherwise}, \end{aligned} \tag{6.3}$$

and if not we have

$$\begin{aligned} & F_t\Big(s_C, f_C, s_D, f_D\Big) = max\Big\{F_t^C\Big(s_C, f_C, s_D, f_D\Big), F_t^D\Big(s_C, f_C, s_D, f_D\Big)\Big\} \quad \text{for } 0 \leq t \leq T, \\ & F_t\Big(s_C, f_C, s_D, f_D\Big) = 0, \text{otherwise}. \end{aligned} \tag{6.4}$$

6.3 Simulation Set-up

The simulation set-ups and criteria used for this chapter are the same as for the previous two. Also, we are evaluating the frequentist Maximum Likelihood Estimator (MLE) and its bias formulated in the previous chapters. For details, please see section 3.2. Note that we mainly focus on two types of model cases in this study: (i)$\boldsymbol{DPWOI}$: a design in which an interim analysis is performed in the simulation step only whilst the classical DP results are used, (ii) $\boldsymbol{DPWI}$: a design in which identical interim analysis conditions are applied in the DP and simulation procedures. It is worth mentioning that MLE estimation results are given in three different trial categories: (i) MLE results obtained from trials without stopping early, i.e. the null hypothesis fails to reject, (ii) those estimated from trials with early stopping owing to the rejection of the null hypothesis at interim, (iii) pooled MLE results which are made up of the mixture of categories (i) and (ii).

6.4 Interim Analysis Characteristics

6.4.1 z-test

To begin with, we borrow the z statistic (pooled version) for comparing the differences in two proportions using the normal approximation offered by (Pham-Gia et al., 2017). Using the observed responses obtained from the arms up to the middle of the trial, we implement a frequentist hypothesis test where the control arm C is compared with the research arm D. In addition, we consider the following hypotheses:

$$\begin{cases} H_0 : \theta_C \geq \theta_D \\ H_a : \theta_C < \theta_D \end{cases} \tag{6.5}$$

As above, the null hypothesis is that the control arm C has the same or higher efficacy, i.e. Bernoulli mean, than the research arm D. Hence, the z statistic, which is a suitable test statistic for large samples binomial distribution approximated by normal distribution, can be formulated as below:

$$Z_{st} = \frac{\widehat{\theta}_C(\tau) - \widehat{\theta}_D(\tau)}{\sqrt{\left(\dfrac{s_C(\tau) + s_D(\tau)}{n_C(\tau) + n_D(\tau)}\right)\left(1 - \dfrac{s_C(\tau) + s_D(\tau)}{n_C(\tau) + n_D(\tau)}\right)\left(\frac{1}{n_C(\tau)} + \frac{1}{n_D(\tau)}\right)}} \tag{6.6}$$

where $\tau = T/2$ in this study. This test statistic could also be used when multiple interim analyses are proposed by adapting the value of τ. Note that we run the one-tailed hypothesis test above by comparing the z statistic with four different standard Normal critical values corresponding to significance levels: $\alpha = 0.0025$, $\alpha = 0.005$, $\alpha = 0.01$, and $\alpha = 0.02$, respectively. Furthermore, in this study, we focus on RAR procedures where the trial size is set to $T = 120$ and $T = 240$ as they are large enough to return a decent amount of observations to carry out the z-test in the middle of the trial. In addition, these values underpin the use of the Normal approximation, and also the results of those trials with early stopping can be compared with the RAR procedures presented in the first chapter, where $T = 60$ and $T = 120$.

6.4.2 Stopping Criteria: Primary Set-up

Stopping a trial for an interim analysis by running a statistical hypothesis test requires establishing criteria so that dividing by zeros is excluded in the z statistic calculation. To do so, aside from the fixed initial allocations (applying in the simulation step) assumed in our proposed designs, we let the necessary conditions for stopping a trial at the middle have at least one observation on the summation of successes and the summation of failures for both arms, i.e. $s_C(\tau) + s_D(\tau) \geq 1$ and $f_C(\tau) + f_D(\tau) \geq 1$ when $\tau = T/2$. Typically, considering the stopping rule together with fixed initial allocations guarantees that the denominator of equation (6.6) will not be zero. By taking the defined stopping criteria into account, the trial can be stopped early for some scenarios in which one can firmly conclude about arms' efficacies based on observed responses up to the interim inspection point.

6.4.3 Stopping Criteria: Developmental Process

We begin with the model case *DPWOI*, and fix the trial size at 120 and the significance level at $\alpha = 0.01$. Figure 6.1 part (a) illustrates estimation results corresponding to the trial that does not stop early (left-hand side) and the one that terminates at the interim (right-hand side). For those stopping early, it is evident that all stars show positive bias since we assume θ_C to be the inferior arm and θ_D the superior one. In contrast, circles are placed in the negative part of the bias axis with a relatively wide range. Focusing on the green cases (the null scenarios, see 2.5), which stand for the scenarios with equal efficacies, i.e. $\theta_C = \theta_D$, larger values of θ_C give rise to more negative bias as choosing larger values for efficacies causes the observation of more success responses in Bernoulli trials. In turn, these heavily-negative bias estimations are happening because of the passive-aggressive performance of DP when it continues sampling from a seemingly better arm which returns some initial appealing responses. Thus, a lack of observations on the other arm leads to even more negative bias estimations when the trial stops in the interim.

On the other hand, in the left-hand side column of figure 6.1 part (a) representing completed trials, one can see that for some scenarios, the MLE of the inferior arm is estimated with highly positive bias. Note that the MLE for most scenarios are negatively estimated as it is typical in our assumed response-adaptive procedure. Numerical investigations revealed that our proposed primary stopping criteria estimate the MLE corresponding to the inferior arm to have heavily-positive bias for the family of scenarios in which $\theta_D = 1$ and a few of which $\theta_D = 0.9$ (except for extreme cases, i.e. $\theta_C = 1$ and $\theta_C = 0$). For example, the violet circle on the far right represents arm C in scenario $(0.1, 1)$. Based on the bias equation presented in the first chapter, it can be concluded that the MLE has been estimated at 1 as the bias is valued at 0.9. Recalling the fixed initial allocations in which arm C and arm D are deterministically allocated to the first and the second time epochs, respectively, along with allocating either arm with a probability of 0.5 at the third time epoch, we expect $\theta_C = 0.1$ to be estimated at 1 with a chance of approximately 5% and 0 with 45%. That is, as arm C is allocated at the third time epoch with a probability of 0.5, we expect to observe a success response in 10% of the time ($\theta_C = 0.1$). This results in estimating MLE at 1 with a chance of approximately 5%, i.e. $0.5 \times 0.1 = 0.05$. A similar explanation can be applied to those 45% cases. For those scenarios where θ_C is close to 1 (black: 0.8 and blue: 0.9), this chance will be even more than 5% as the probability of observing a success response increases. Less severe but similar circumstances can be applied to the scenarios in which $\theta_D = 0.9$ and $\theta_C \leq 0.7$. For example, the second pink circle from the far right representing arm C in scenario $(0.1, 0.9)$ tends to show a bias of 0.4, resulting in estimating MLE at 0.5. Again, a similar reasoning mentioned above can be used for justifying why the estimation bias is large and positive in this case.

To fix the issue with heavily-positive biased observations, we implemented the model *DPWI* by defining the new value functions and transitional probabilities according to the equation (6.2). Figure 6.1 part (b) illustrates the estimation results acquired from this type of design. Although some scenarios have been slightly improved, those heavily-positive biased observations still appear in their places. Comparing the information of the success and failure observations on both arms in the interim with the end-of-the-trial counterparts for a given scenario, we noticed that for most simulation replications where the number of success observations on the inferior arm (arm C) is zero, the trial stops early. For those replications without stopping early (mainly because of the smaller mean for the other arm), the number of success observations remains zero at the end of the trial.

On the other hand, highly-positive biased estimations belong to the replications where the number of success observations on the inferior arm (arm C) is not zero. Hence, the trial does not stop early in these cases as the z-score is not as extreme as the critical value. As a final alteration in formulating appropriate stopping criteria, we force a trial to continue even if no success response has been observed up to the interim point instead of stopping early. By doing so, we have quite a few simulation replications where the MLE is estimated by 0 to decrease the average

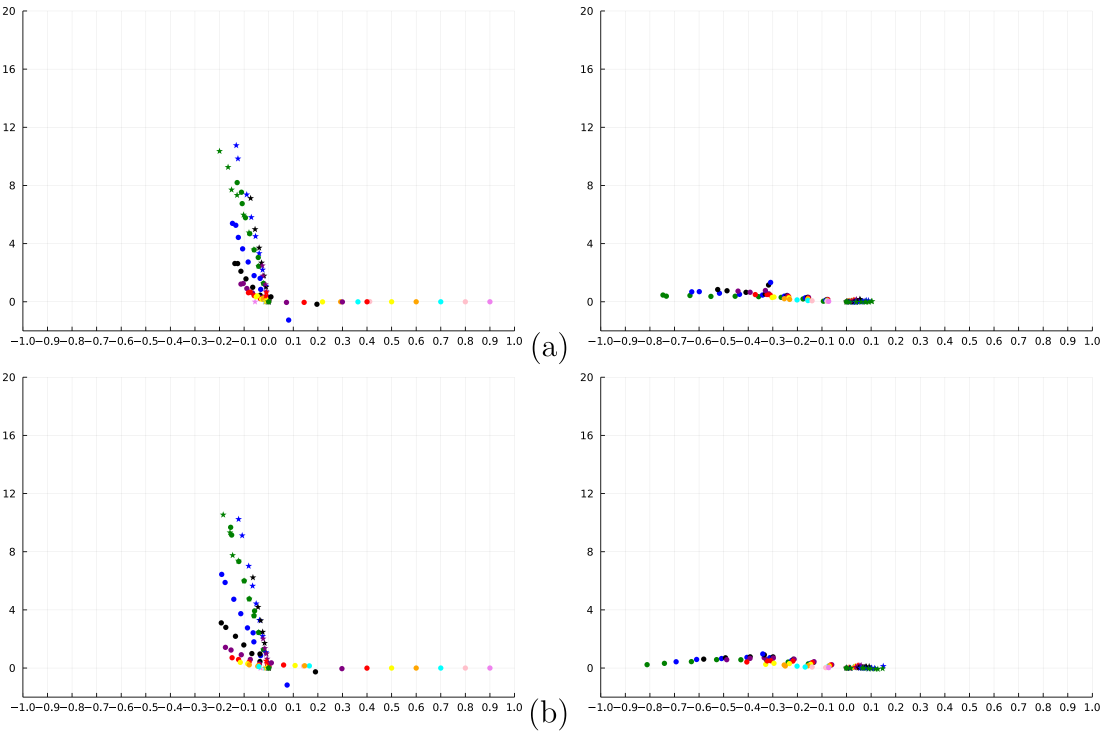


Figure 6.1: The estimation bias and covariance reduction comparison in the model case (a) *DPWOI*, and (b) *DPWI*. Note that $T = 120$ and $\alpha = 0.01$, and completed trials (left-hand side) are compared with those stopped at the interim (right-hand side). x-axis: Bias of the estimator, y-axis: Covariance (Estimator, Sample Size).

bias at the end of the trial. This alteration can ultimately remove all scenarios with heavily-positive bias estimation (those with a 5% chance of MLE estimated at 1) by preventing the other 45% simulation replications from entering the interim stage.

6.4.4 Stopping Criteria: Final Necessary Conditions

The final stopping criteria for the two-armed Bayesian Beta-Bernoulli model with a planned interim analysis in the middle of the trial ($\tau = T/2$) can be formulated as follows:

$$\begin{cases} s_C(\tau) + s_D(\tau) \geq 1 \\ f_C(\tau) + f_D(\tau) \geq 1 \\ s_C(\tau) \neq 0 \quad \& \quad f_C(\tau) \neq 1 \\ s_D(\tau) \neq 0 \quad \& \quad f_D(\tau) \neq 1 \end{cases} \tag{6.7}$$

Note that the above final conditions guarantee that not only do trials without success observation fail to enter the interim analysis phase, but those where MLE is estimated at 0 up to the interim inspection point are prevented from entering as well. Considering the ultimate stopping criteria, we set up the one-tailed hypothesis testing (6.5) with four different significance levels: $\alpha = 0.0025$, $\alpha = 0.005$, $\alpha = 0.01$, and $\alpha = 0.02$, in both model cases *DPWOI* and *DPWI* with the sizes of 120 and 240. The estimation results are presented in the following section.

6.5 Average Estimation Bias

In this section the estimation results are generally presented in three different trial categories for $T = 120$ and $T = 240$:

- **Trial category (i)**: completed trials, i.e. the null hypothesis fails to reject, figures 6.2 and 6.5.
- **Trial category (ii)**: trials terminated at interim, i.e. the alternative hypothesis is statistically significant as the null is rejected, figures 6.3 and 6.6.
- **Trial category (iii)**: pooling the first and the second categories, which is called *standard design*, 6.4 and 6.7

We also assumed two different model cases:

- **(i)** ***DPWOI***: figures 6.2, 6.3, and 6.4.
- **(ii)** ***DPWI***: figures 6.5, 6.6, and 6.7.

Simulation results presented in figures 6.2 to 6.7 generally show that the differences in like-for-like designs for model cases (i) *DPWOI* and (ii) *DPWI* are relatively small. For most scenarios, there is no change, or the changes are infinitesimal in a way that one might ignore them. On the other hand, comparing trial categories (i) and (iii) in both *DPWOI* and *DPWI* reveals that the results in the standard designs are slightly better than the equivalent completed trials (trial category (i)) as there are some reductions in both bias and covariance values (compare the blue colour scenarios). This mainly occurs because of the fact that adding the second category results, which are usually associated with positive bias values, can compensate for negative counterparts in the first category.

The effect of increasing the significance level of the hypothesis test in the interim for trial category (i) is the inverse of trial category (ii). By comparing part (a) with part (d) in figure 6.2, one can realise that using a larger significance level α in the first category causes more trials to be rejected and consequently makes the estimations slightly more negative in bias when it averages out over the number of simulation replications. However, this increment in α has an almost reverse effect (except for some extreme scenarios) in the second trial category, see figure 6.3. In fact, assuming higher significance levels results in rejecting the null hypothesis more frequently, which makes the estimation range tighter for trial category (ii), and in both *DPWOI* and *DPWI* when it averages out over the number of simulation replications. Note that the reason for having almost all circles in the negative part and stars in the positive part of the bias axis, along with those extreme scenarios, i.e. circles appearing in the positive part, has been explained in section 6.4.3.

Another interesting observation in the data illustrated in trial category (i) and both *DPWOI* and *DPWI* is the null scenarios, i.e. scenarios in which arms' efficacies are identical. Comparing plots through part (a) to part (d) ascertains that the higher the significance level, the larger the gap between the star and circle in several null scenarios. Consequently, these trials are now plotted in the trial category (ii) instead. Although in the null scenarios in which $\theta_k \leq 0.5$, the gap can be ignored as it is not appreciable, for those where θ_k is close to 1 the frequency of false early stopping slightly inflates and goes beyond the significance level of the test. The reason for the larger gaps observed in those null scenarios can be attributed to the passive-aggressive property of the DP procedure. This property, together with the frequency of false early stopping, is explained in section 6.6.

6.6 Percentage of False Early Stopping

In this section, we report some key points about the frequencies of rejecting the null hypothesis in (6.5) and stopping the trial in the interim when the arm's efficacies are equal. Note that all numbers are obtained by averaging over the total number of simulation replications which is set to 1 million. We call this phenomenon the *percentage of false early stopping*. In other words, and equivalent to the type I error rate, the percentage of false early stopping is the percentage of erroneously rejecting the null hypothesis when it is statistically significant. Figure 6.8 illustrates the percentage of false early stopping in the model cases *DPWOI*

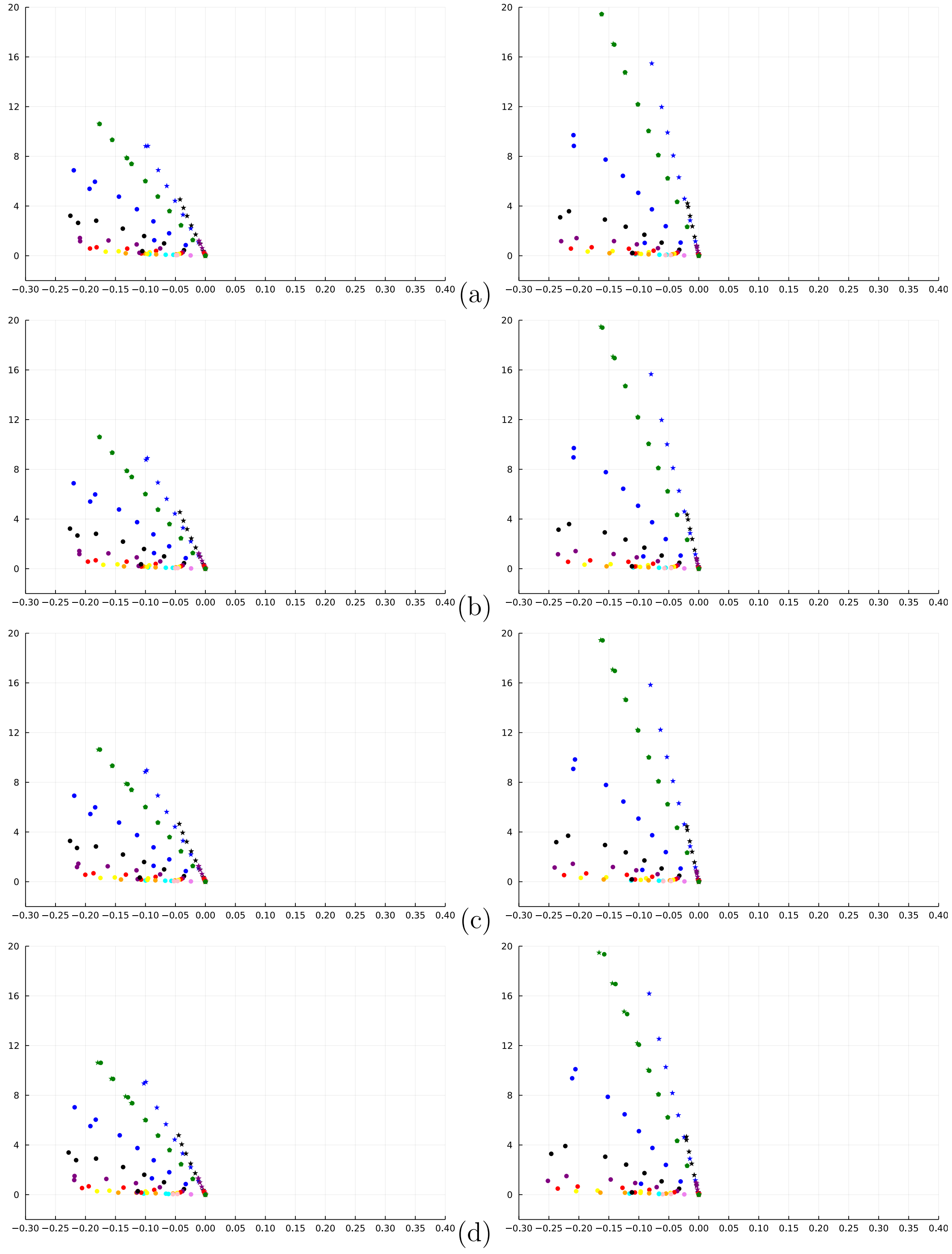


Figure 6.2: The estimation bias and covariance reduction comparison in the model case *DPWOI*, trial category (i), $T = 120$ vs $T = 240$: (a) $\alpha = 0.0025$ (b) $\alpha = 0.005$ (c) $\alpha = 0.01$ (d) $\alpha = 0.02$. x-axis: Bias of the estimator, y-axis: Covariance (Estimator, Sample Size).

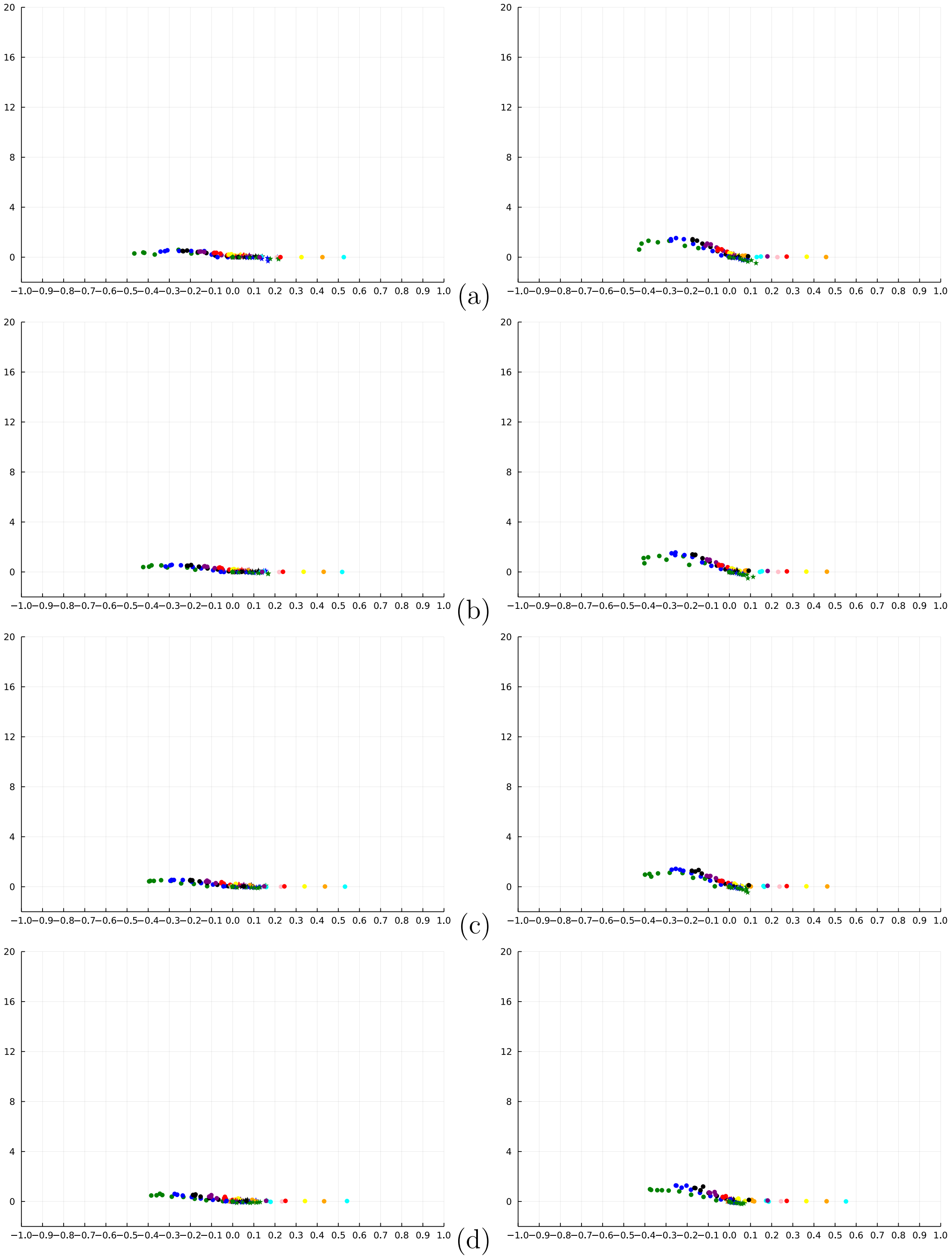


Figure 6.3: The estimation bias and covariance reduction comparison in the model case *DPWOI*, trial category (ii), $T = 120$ vs $T = 240$: (a) $\alpha = 0.0025$ (b) $\alpha = 0.005$ (c) $\alpha = 0.01$ (d) $\alpha = 0.02$. x-axis: Bias of the estimator, y-axis: Covariance (Estimator, Sample Size).

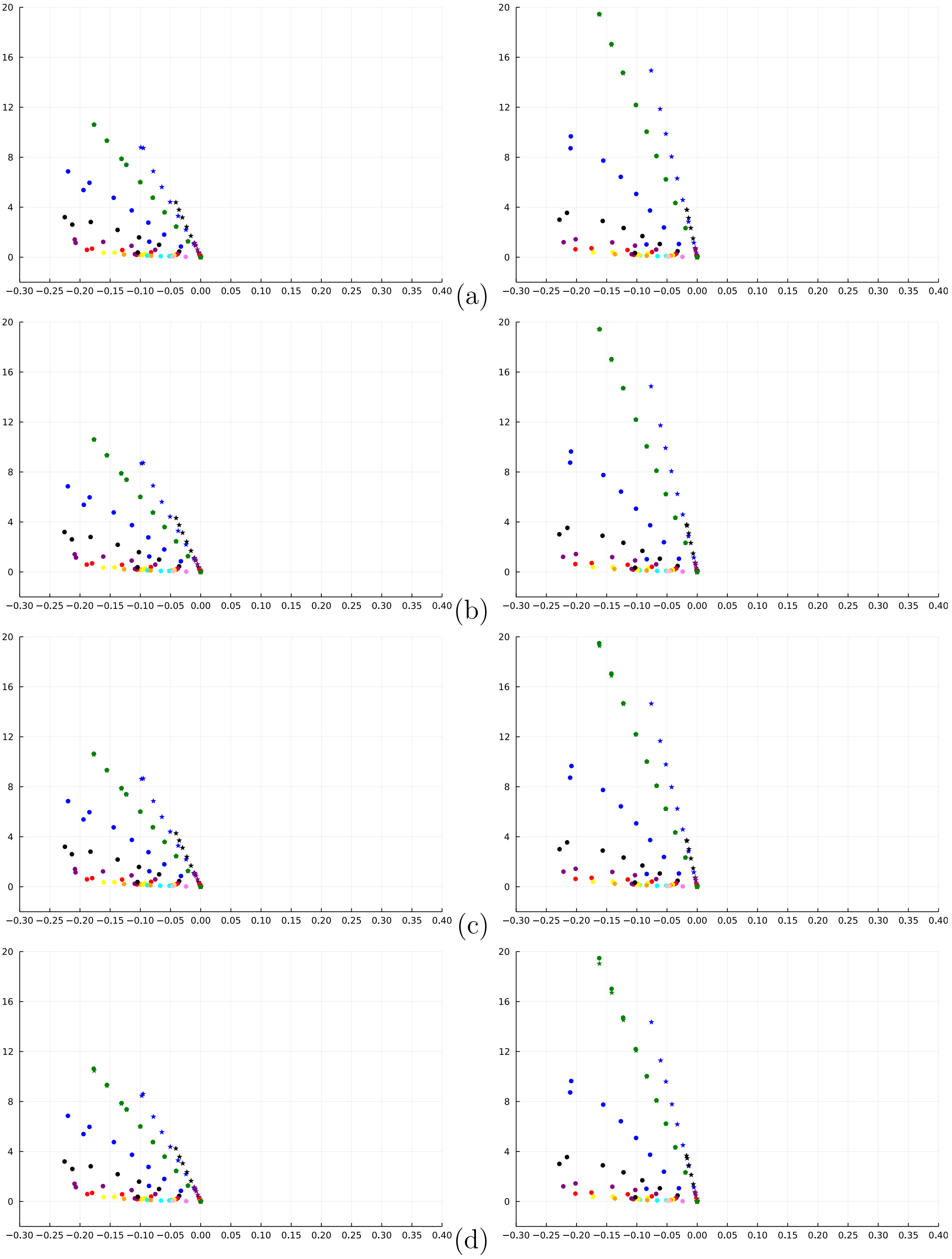


Figure 6.4: The estimation bias and covariance reduction comparison in the model case *DPWOI*, trial category (iii), $T = 120$ vs $T = 240$: (a) $\alpha = 0.0025$ (b) $\alpha = 0.005$ (c) $\alpha = 0.01$ (d) $\alpha = 0.02$. x-axis: Bias of the estimator, y-axis: Covariance (Estimator, Sample Size).

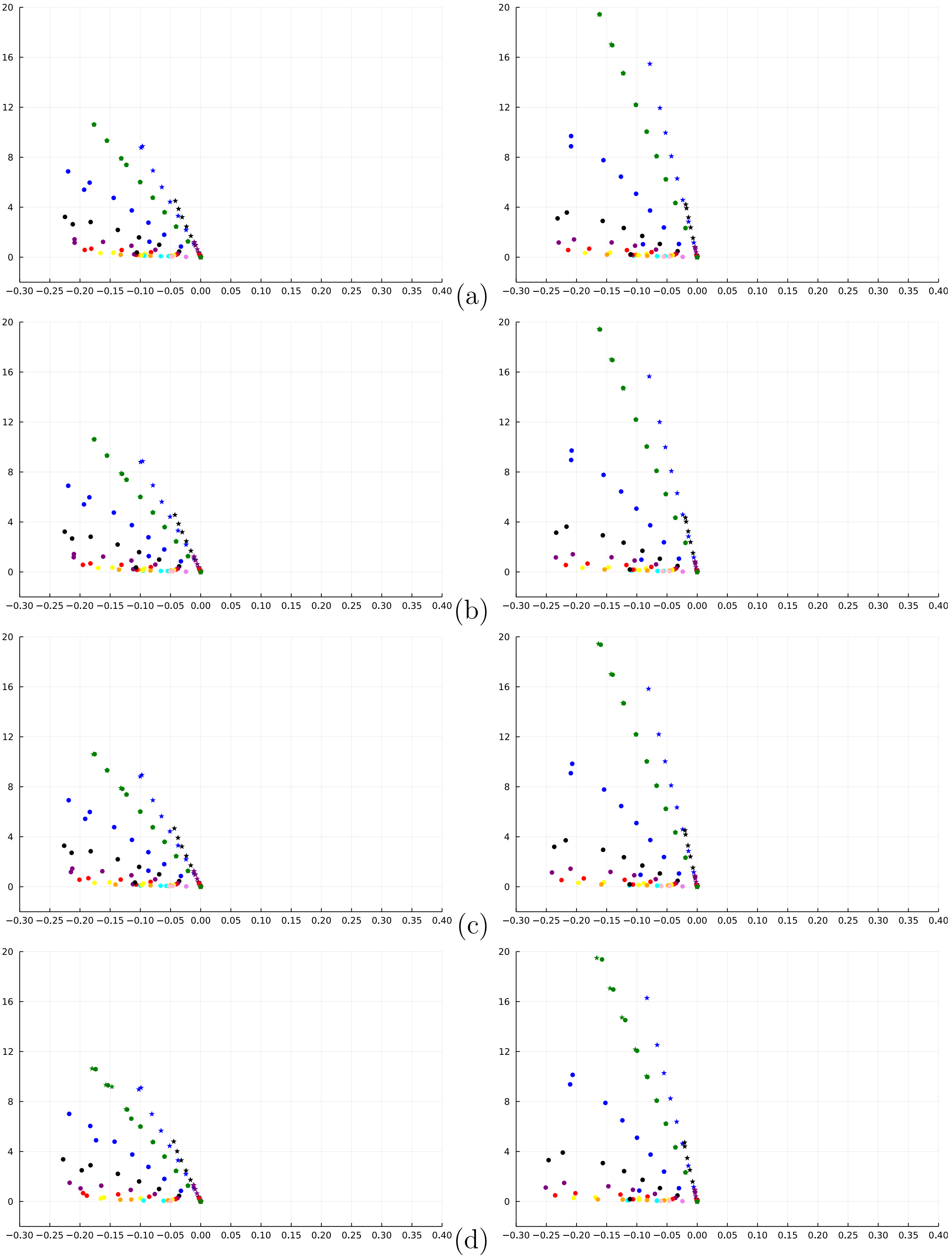


Figure 6.5: The estimation bias and covariance reduction comparison in the model case *DPWI*, trial category (i), $T = 120$ vs $T = 240$: (a) $\alpha = 0.0025$ (b) $\alpha = 0.005$ (c) $\alpha = 0.01$ (d) $\alpha = 0.02$. x-axis: Bias of the estimator, y-axis: Covariance (Estimator, Sample Size).

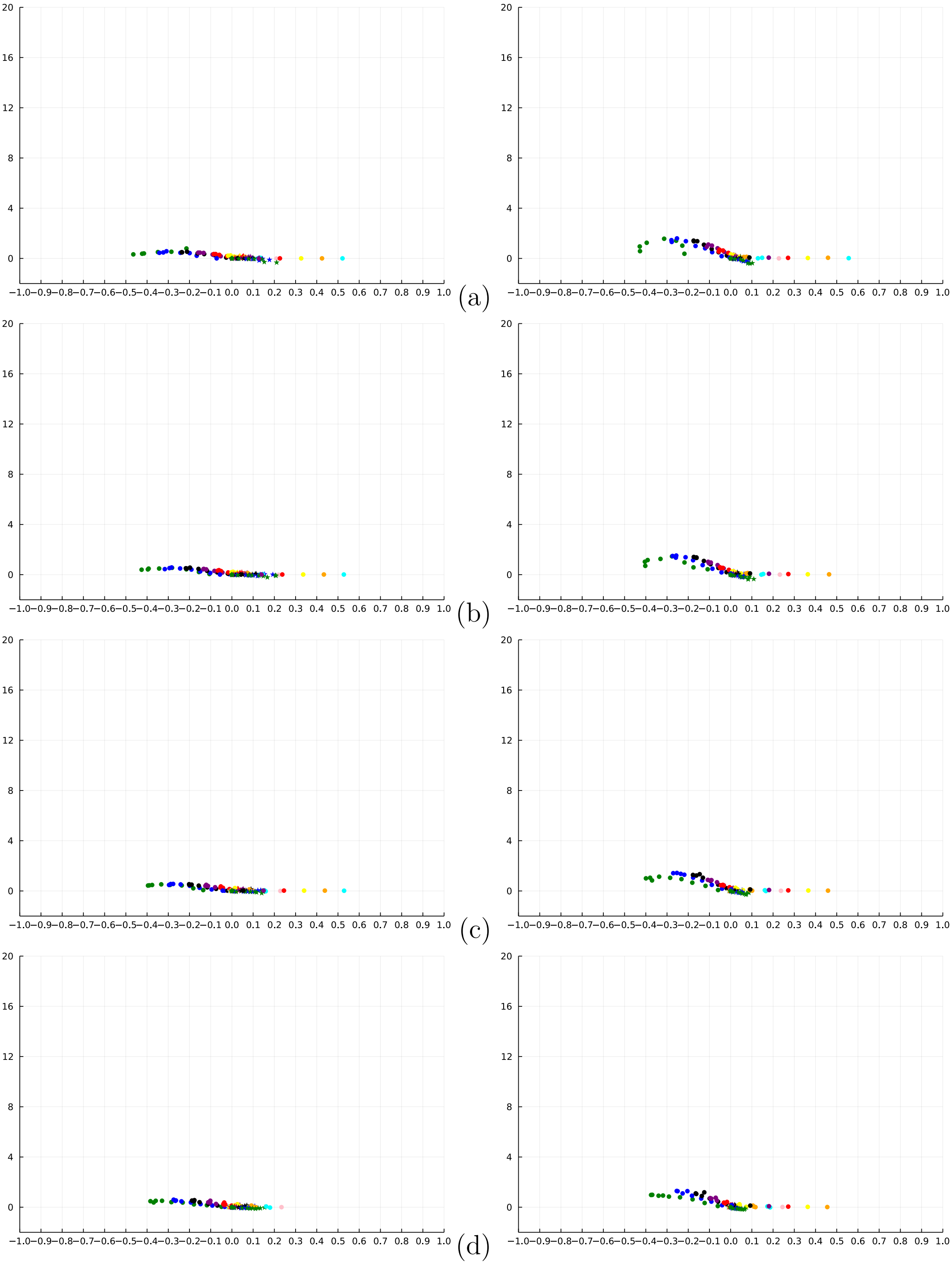


Figure 6.6: The estimation bias and covariance reduction comparison in the model case *DPWI*, trial category (ii), $T = 120$ vs $T = 240$: (a) $\alpha = 0.0025$ (b) $\alpha = 0.005$ (c) $\alpha = 0.01$ (d) $\alpha = 0.02$. x-axis: Bias of the estimator, y-axis: Covariance (Estimator, Sample Size).

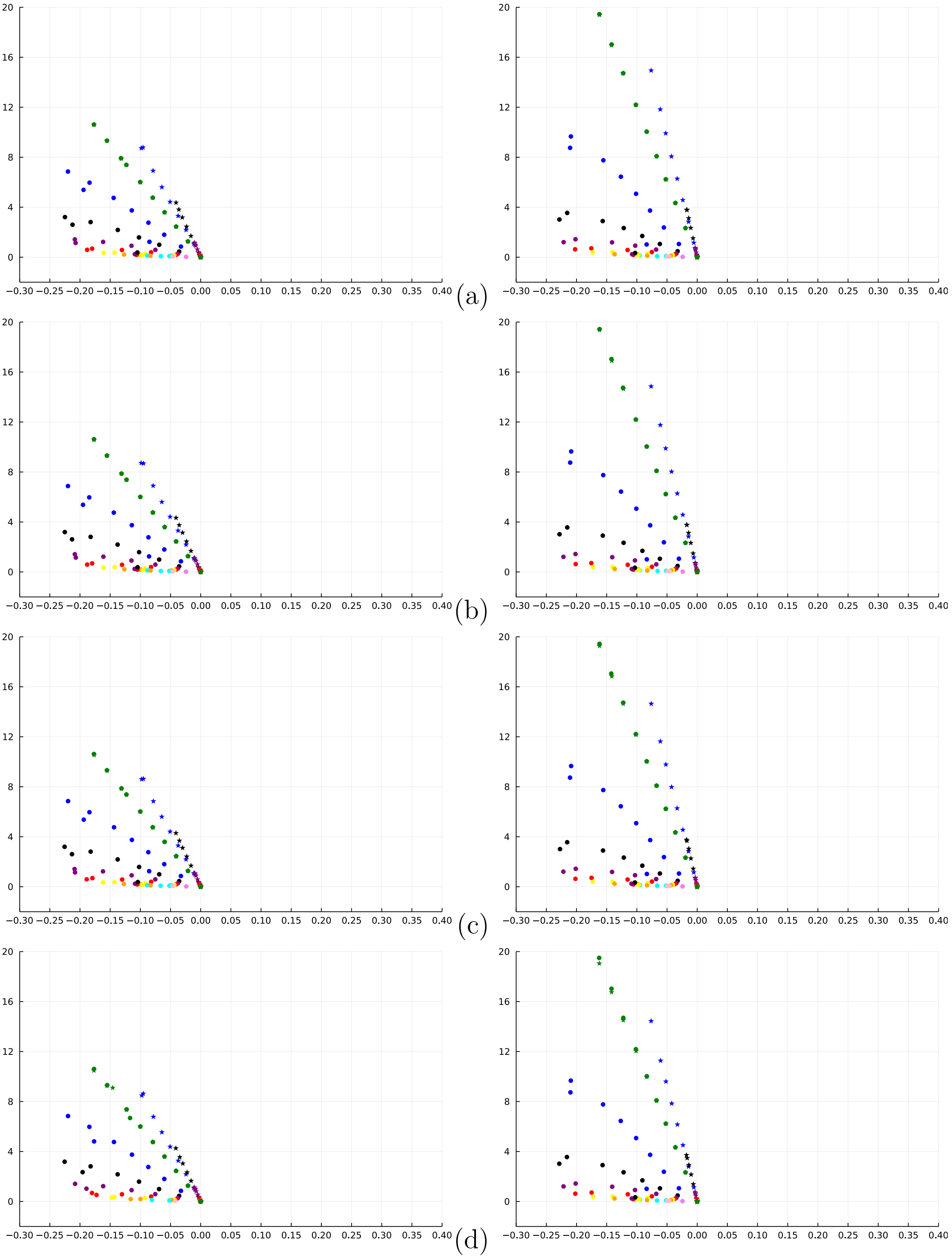
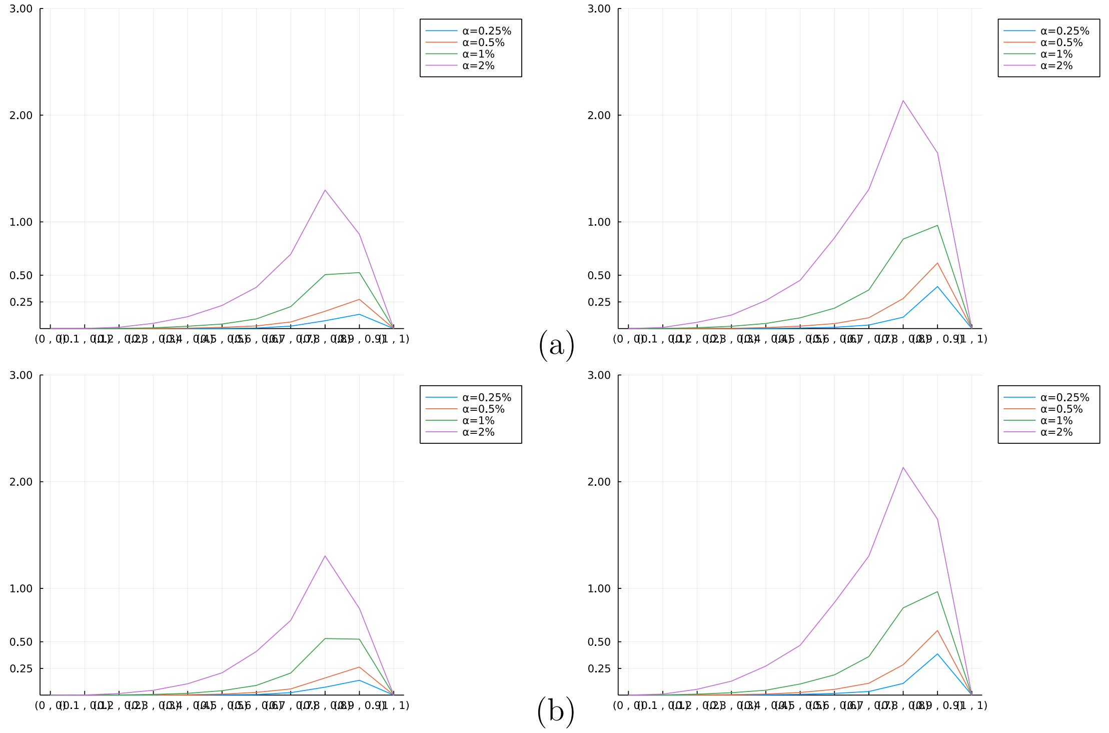


Figure 6.7: The estimation bias and covariance reduction comparison in the model case *DPWI*, trial category (iii), $T = 120$ vs $T = 240$: (a) $\alpha = 0.0025$ (b) $\alpha = 0.005$ (c) $\alpha = 0.01$ (d) $\alpha = 0.02$. x-axis: Bias of the estimator, y-axis: Covariance (Estimator, Sample Size).

(part a) and *DPWI* (part b) for various significance levels. It can be seen that the alteration patterns in the percentage of false early stopping plots are almost identical in *DPWOI* and *DPWI* for both trial sizes.

For the trial size of 120, and both parts (a) and (b), the percentage of false early stopping in all scenarios is below the significance level of the test. Whilst for the scenarios where $0.5 \leq \theta_k \leq 0.9$ the alteration trend dramatically inflates with a maximum at $\theta_k = 0.8$ for the significance levels $\alpha = 0.01, 0.02$. However, the story is slightly different for the designs with a trial size of 240. For the significance level $\alpha = 0.02$, the climax is still at $\theta_k = 0.8$, whilst the percentage of false early stopping goes above the significance level for this scenario. Furthermore, the percentage of false early stopping also inflates beyond the significance levels $\alpha = 0.0025, 0.005$, and 0.01 in the scenario with $\theta_k = 0.9$. This issue can be justified by the possibility that the trial size of 240 can potentially provide a decent number of observations for running hypothesis testing (6.5). On the other hand, sticking the DP procedure to one arm in the null scenarios when the efficacy is close to 1 particularly gives rise to rejecting the null hypothesis more frequently than the allowed boundary, i.e. significance test level. To soften this passive-aggressive behaviour of the DP procedure and subsequently correct the inflation of the percentage of false early stopping, one may be interested in applying techniques introduced in chapter 5 of this dissertation.

Table 6.1 reports the percentage of trials stopping early for all scenarios in both model cases *DPWOI* and *DPWI* and for each significance level. All reported numbers are rounded to three decimal places since the simulation error for our



(a)

(b)

Figure 6.8: The percentages of false early stopping for various significance levels: left-hand side $T = 120$ vs right-hand side $T = 240$. Part (a): *DPWOI*, and part (b): *DPWI*. x-axis: Scenario, y-axis: Percentage.

Bayesian-Bernoulli two-armed model is fixed at 0.5%. As a general rule, numbers reported for trial size 240 are larger than the like-for-like counterpart in trial size 120. For example, for scenario $(0.2, 0.8)$, and significance levels at $\alpha = 0.005$ and 0.02, the designs with the trial size 120 show that 2.08% and 6.963% in the model case *DPWOI* , and 2.101% and 7.009% in *DPWI* of the trials stop early at interim inspection. On the other hand, the like-for-like numbers for the trial size 240 are 3.471% and 10.819% in the model case *DPWOI*, and 3.494% and 10.896% in *DPWI*, respectively.

	DPWOI $[T = 120]$				***DPWI*** $[T = 120]$				***DPWOI*** $[T = 240]$				***DPWI*** $[T = 240]$			
α / (θ_C, θ_D)	0.0025	0.005	0.01	0.02	0.0025	0.005	0.01	0.02	0.0025	0.005	0.01	0.02	0.0025	0.005	0.01	0.02
(0 , 0)	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
(0 , 0.1)	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
(0 , 0.2)	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
(0 , 0.3)	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
(0 , 0.4)	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
(0 , 0.5)	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
(0 , 0.6)	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
(0 , 0.7)	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
(0 , 0.8)	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
(0 , 0.9)	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
(0 , 1)	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
(0.1 , 0.1)	0.000	0.000	0.000	0.002	0.000	0.000	0.000	0.001	0.000	0.000	0.001	0.011	0.000	0.000	0.002	0.010
(0.1 , 0.2)	0.000	0.001	0.008	0.048	0.000	0.001	0.009	0.053	0.005	0.023	0.084	0.393	0.004	0.020	0.083	0.401
(0.1 , 0.3)	0.003	0.012	0.078	0.345	0.003	0.012	0.072	0.350	0.052	0.169	0.498	1.718	0.053	0.160	0.506	1.702
(0.1 , 0.4)	0.020	0.067	0.288	1.011	0.020	0.065	0.286	1.009	0.176	0.424	1.105	3.293	0.169	0.444	1.103	3.314
(0.1 , 0.5)	0.076	0.204	0.621	1.833	0.070	0.207	0.610	1.849	0.304	0.715	1.693	4.521	0.300	0.719	1.730	4.579
(0.1 , 0.6)	0.176	0.431	1.005	2.724	0.173	0.425	1.000	2.740	0.504	1.038	2.331	5.944	0.495	1.035	2.337	5.926
(0.1 , 0.7)	0.421	0.789	1.625	3.520	0.422	0.796	1.629	3.551	0.772	1.527	3.088	6.619	0.759	1.534	3.091	6.605
(0.1 , 0.8)	0.857	1.250	2.345	4.035	0.864	1.258	2.337	3.996	1.193	2.049	3.762	6.087	1.191	2.061	3.774	6.092
(0.1 , 0.9)	1.180	1.642	2.394	2.824	1.178	1.642	2.375	2.806	2.232	2.826	3.318	3.619	2.228	2.824	3.345	3.613
(0.1 , 1)	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
(0.2 , 0.2)	0.000	0.000	0.002	0.013	0.000	0.000	0.002	0.015	0.000	0.002	0.008	0.059	0.000	0.002	0.009	0.055
(0.2 , 0.3)	0.001	0.005	0.032	0.167	0.001	0.004	0.027	0.170	0.013	0.049	0.165	0.723	0.013	0.048	0.174	0.718
(0.2 , 0.4)	0.012	0.046	0.181	0.770	0.013	0.045	0.190	0.768	0.116	0.307	0.856	2.803	0.117	0.312	0.869	2.813
(0.2 , 0.5)	0.060	0.204	0.605	1.994	0.064	0.198	0.601	1.974	0.372	0.868	2.057	5.570	0.371	0.877	2.051	5.579
(0.2 , 0.6)	0.210	0.557	1.317	3.776	0.210	0.566	1.323	3.775	0.755	1.531	3.400	8.592	0.757	1.530	3.396	8.515
(0.2 , 0.7)	0.602	1.190	2.511	5.494	0.605	1.193	2.512	5.491	1.206	2.441	4.929	10.491	1.253	2.443	4.920	10.454
(0.2 , 0.8)	1.371	2.080	4.006	6.963	1.356	2.101	4.001	7.009	2.008	3.471	6.545	10.819	2.032	3.494	6.559	10.896
(0.2 , 0.9)	2.164	3.027	4.507	5.354	2.187	3.034	4.495	5.334	4.121	5.260	6.330	7.070	4.112	5.217	6.361	7.116
(0.2 , 1)	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
(0.3 , 0.3)	0.001	0.001	0.007	0.050	0.000	0.001	0.007	0.045	0.001	0.003	0.022	0.127	0.001	0.005	0.022	0.131
(0.3 , 0.4)	0.004	0.014	0.076	0.355	0.005	0.017	0.073	0.347	0.023	0.082	0.290	1.159	0.027	0.079	0.280	1.138
(0.3 , 0.5)	0.034	0.109	0.360	1.315	0.032	0.107	0.362	1.322	0.198	0.501	1.336	3.980	0.198	0.509	1.322	3.978
(0.3 , 0.6)	0.164	0.436	1.123	3.363	0.157	0.436	1.121	3.358	0.687	1.427	3.262	8.457	0.673	1.436	3.287	8.432
(0.3 , 0.7)	0.583	1.210	2.644	6.048	0.582	1.212	2.669	6.040	1.451	2.812	5.630	12.082	1.424	2.761	5.590	12.073
(0.3 , 0.8)	1.555	2.525	4.961	8.898	1.553	2.545	4.954	8.920	2.603	4.439	8.321	14.008	2.580	4.451	8.319	14.125
(0.3 , 0.9)	2.949	4.219	6.202	7.669	2.946	4.259	6.230	7.642	5.647	7.127	8.825	10.219	5.643	7.209	8.833	10.246
(0.3 , 1)	0.000	0.000	0.001	0.000	0.001	0.001	0.001	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
(0.4 , 0.4)	0.001	0.004	0.022	0.111	0.001	0.004	0.017	0.106	0.002	0.009	0.048	0.263	0.002	0.010	0.046	0.272
(0.4 , 0.5)	0.012	0.045	0.142	0.621	0.011	0.040	0.144	0.623	0.045	0.143	0.472	1.688	0.049	0.142	0.469	1.716
(0.4 , 0.6)	0.083	0.247	0.661	2.225	0.077	0.224	0.666	2.222	0.332	0.795	2.074	5.895	0.339	0.800	2.074	5.892
(0.4 , 0.7)	0.419	0.921	2.118	5.222	0.418	0.923	2.141	5.249	1.244	2.475	5.139	11.310	1.211	2.468	5.137	11.326
(0.4 , 0.8)	1.440	2.490	5.134	9.540	1.436	2.501	5.085	9.545	2.817	4.813	9.081	15.734	2.825	4.836	9.056	15.795
(0.4 , 0.9)	3.395	5.018	7.503	9.524	3.417	5.027	7.467	9.494	6.784	8.683	10.872	13.050	6.715	8.690	10.835	13.130
(0.4 , 1)	0.008	0.011	0.013	0.014	0.009	0.013	0.011	0.000	0.003	0.002	0.002	0.002	0.002	0.002	0.002	0.001
(0.5 , 0.5)	0.002	0.011	0.043	0.217	0.003	0.009	0.041	0.209	0.006	0.023	0.101	0.453	0.007	0.025	0.104	0.467
(0.5 , 0.6)	0.026	0.095	0.286	1.084	0.027	0.090	0.296	1.086	0.088	0.259	0.801	2.734	0.090	0.260	0.778	2.727
(0.5 , 0.7)	0.214	0.529	1.304	3.549	0.221	0.515	1.317	3.543	0.649	1.444	3.326	8.122	0.652	1.433	3.330	8.063
(0.5 , 0.8)	1.048	1.987	4.290	8.560	1.052	1.988	4.290	8.565	2.489	4.410	8.422	15.298	2.475	4.471	8.459	15.301
(0.5 , 0.9)	3.432	5.247	8.031	10.554	3.440	5.233	8.063	10.570	7.287	9.569	12.295	15.310	7.322	9.610	12.282	15.295
(0.5 , 1)	0.050	0.079	0.094	0.103	0.047	0.077	0.096	0.000	0.025	0.023	0.028	0.024	0.021	0.026	0.025	0.024
(0.6 , 0.6)	0.005	0.025	0.091	0.388	0.006	0.026	0.091	0.409	0.012	0.047	0.191	0.851	0.016	0.054	0.190	0.868
(0.6 , 0.7)	0.081	0.208	0.616	1.820	0.085	0.222	0.607	1.824	0.195	0.491	1.398	4.000	0.195	0.496	1.396	3.972
(0.6 , 0.8)	0.601	1.196	2.887	6.195	0.591	1.212	2.872	6.159	1.495	2.925	6.033	11.899	1.529	2.945	6.067	11.802
(0.6 , 0.9)	2.785	4.556	7.249	10.058	2.781	4.552	7.277	10.049	6.964	9.542	12.546	16.337	6.989	9.525	12.552	16.302
(0.6 , 1)	0.180	0.331	0.414	0.481	0.178	0.337	0.428	0.000	0.181	0.180	0.185	0.185	0.180	0.172	0.180	0.180
(0.7 , 0.7)	0.023	0.062	0.206	0.697	0.023	0.058	0.207	0.701	0.033	0.101	0.362	1.304	0.034	0.111	0.361	1.303
(0.7 , 0.8)	0.254	0.556	1.461	3.453	0.258	0.537	1.466	3.436	0.532	1.175	2.814	6.375	0.531	1.163	2.840	6.399
(0.7 , 0.9)	1.743	3.005	5.046	7.422	1.744	2.982	5.027	7.389	5.190	7.565	10.454	14.537	5.271	7.556	10.504	14.545
(0.7 , 1)	0.440	0.968	1.382	1.815	0.439	0.954	1.358	0.000	0.933	0.995	0.978	1.005	0.939	0.966	0.979	0.987
(0.8 , 0.8)	0.074	0.162	0.506	1.298	0.074	0.161	0.530	1.304	0.107	0.282	0.839	2.137	0.110	0.285	0.817	2.134
(0.8 , 0.9)	0.719	1.265	2.295	3.504	0.714	1.279	2.285	3.505	2.244	3.368	5.017	7.699	2.204	3.342	5.043	7.731
(0.8 , 1)	0.696	1.754	2.893	4.653	0.700	1.763	2.912	0.000	3.642	4.177	4.231	4.323	3.655	4.130	4.229	4.283
(0.9 , 0.9)	0.134	0.274	0.525	0.884	0.139	0.263	0.524	0.812	0.394	0.615	0.967	1.646	0.386	0.606	0.970	1.649
(0.9 , 1)	0.539	1.582	2.984	5.772	0.530	1.578	2.971	0.000	7.141	9.632	10.883	12.073	7.117	9.637	10.862	12.146
(1 , 1)	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000

Table 6.1: The percentages of the trials that the null hypothesis in (6.5) rejects at interim inspection. Numbers rounded to three decimal places as the simulation error fixed at 0.005.

6.7 Subject benefit

Tables 6.2 and 6.3 report the numerical results of the subject benefit obtained from the model cases *DPWOI* and *DPWI*, respectively. Each component in the tables comprises the average number of success responses on both arms and the corresponding standard deviation at the end of the trial. Comparing the like-for-like results in cases *DPWOI* and *DPWI*, the differences are so small that one can consider them almost identical. The estimation results confirmed this identical resemblance in the data in section 6.5. On the other hand, setting higher significance levels for the test generally results in losing subject benefit in both cases *DPWOI* and *DPWI* and for both trial sizes 120 and 240. However, the loss of the subject benefit may not be appreciable as the reductions in estimation bias, discussed in section 6.5, are also not substantial. For instance, for scenario $(0.2, 0.8)$ and significance levels at $\alpha = 0.005$ and 0.02, the designs with the trial size 120 give a subject benefit of 93.82 ± 8.41 and 91.48 ± 13.28 in the model case *DPWOI*, and 93.81 ± 8.43 and 91.46 ± 13.33 in *DPWI*. On the other hand, the like-for-like results for the trial size 240 are 187.35 ± 18.88 and 180.31 ± 30.75 in the model case *DPWOI*, and 187.73 ± 18.93 and 180.24 ± 30.84 in *DPWI*, respectively.

6.8 Discussion

In this chapter, we implemented an interim study on the two-armed Bayesian Bernoulli bandit model for the first time in the literature. Apart from the technical complexity associated with the implementation process, numerical results confirmed improvements in the efficacy estimation bias when an interim analysis is implemented in the model compared with the original model without interim studies. Moreover, by introducing a concept called *percentage of false early stopping*, we made some points about the inflation of the percentage of false early stopping, together with the acquired subject benefit results.

In the following, it is worth mentioning some areas of improvement and future works for this novel family of RAR designs. The test statistic we used in this study is the pooled version of the z statistic with the approximation of standard Normal distribution. Please note that using another frequentist or Bayesian test statistic might be of interest, see Pham-Gia et al. (2017). Furthermore, comparing the Bayesian posterior distribution at interim inspections can be a good substitution for the frequentist hypothesis testing. Raineri et al. (2014) introduced a known algorithm for computing exactly inequalities between Beta distributions by which, in turn, a statistically significant inference about arm's efficacies can be achieved at interim analysis points.

While we considered a single interim inspection look at the middle of the trial, setting up a sequence of interim analyses can be another potential area to make a more accurate estimate. However, implementing multiple interim inspection looks in our model will be associated with greater statistical complexity as each point will depend on all previous interim points.

Finally, setting up an interim analysis in the modified models discussed in chapter 2 can also give rise to having less biased estimation. For instance, the relationship between the $\dot{s}$ in OIDP, the degree of randomisation in RDP, and the stopping criteria are other areas that potentially need improvement.

	model case *DPWOI* $[T = 120]$				**model case *DPWOI*** $[T = 240]$			
(θ_C, θ_D)	$\alpha = 0.0025$	$\alpha = 0.005$	$\alpha = 0.01$	$\alpha = 0.02$	$\alpha = 0.0025$	$\alpha = 0.005$	$\alpha = 0.01$	$\alpha = 0.02$
(0 , 0)	0.00 ± 0.00	0.00 ± 0.00	0.00 ± 0.00	0.00 ± 0.00	0.00 ± 0.00	0.00 ± 0.00	0.00 ± 0.00	0.00 ± 0.00
(0 , 0.1)	10.15 ± 3.50	10.15 ± 3.49	10.14 ± 3.50	10.15 ± 3.50	21.74 ± 4.84	21.74 ± 4.84	21.73 ± 4.84	21.73 ± 4.83
(0 , 0.2)	21.94 ± 4.59	21.95 ± 4.59	21.95 ± 4.59	21.94 ± 4.60	45.55 ± 6.37	45.56 ± 6.38	45.56 ± 6.37	45.55 ± 6.38
(0 , 0.3)	33.95 ± 5.23	33.96 ± 5.22	33.96 ± 5.22	33.96 ± 5.23	69.58 ± 7.28	69.59 ± 7.26	69.58 ± 7.26	69.58 ± 7.27
(0 , 0.4)	46.06 ± 5.57	46.04 ± 5.56	46.06 ± 5.57	46.05 ± 5.57	93.72 ± 7.75	93.71 ± 7.75	93.71 ± 7.75	93.72 ± 7.75
(0 , 0.5)	58.21 ± 5.68	58.22 ± 5.68	58.21 ± 5.67	58.22 ± 5.68	117.91 ± 7.89	117.91 ± 7.91	117.90 ± 7.89	117.93 ± 7.91
(0 , 0.6)	70.42 ± 5.55	70.42 ± 5.54	70.42 ± 5.55	70.42 ± 5.54	142.16 ± 7.74	142.17 ± 7.74	142.15 ± 7.74	142.15 ± 7.73
(0 , 0.7)	82.67 ± 5.21	82.67 ± 5.21	82.66 ± 5.21	82.67 ± 5.20	166.41 ± 7.23	166.44 ± 7.22	166.44 ± 7.23	166.43 ± 7.23
(0 , 0.8)	94.93 ± 4.50	94.92 ± 4.50	94.93 ± 4.49	94.93 ± 4.50	190.82 ± 6.33	190.82 ± 6.33	190.81 ± 6.33	190.80 ± 6.33
(0 , 0.9)	107.02 ± 3.32	107.02 ± 3.32	107.02 ± 3.32	107.02 ± 3.32	215.01 ± 4.68	215.01 ± 4.68	215.01 ± 4.68	215.01 ± 4.68
(0 , 1)	119.00 ± 0.00	119.00 ± 0.00	119.00 ± 0.00	119.00 ± 0.00	239.00 ± 0.00	239.00 ± 0.00	239.00 ± 0.00	239.00 ± 0.00
(0.1 , 0.1)	12.00 ± 3.28	12.00 ± 3.29	12.00 ± 3.29	12.00 ± 3.29	24.00 ± 4.65	24.01 ± 4.65	23.99 ± 4.65	24.00 ± 4.65
(0.1 , 0.2)	21.40 ± 5.06	21.40 ± 5.07	21.41 ± 5.07	21.40 ± 5.07	44.52 ± 7.45	44.50 ± 7.44	44.49 ± 7.48	44.41 ± 7.56
(0.1 , 0.3)	33.49 ± 5.75	33.49 ± 5.75	33.47 ± 5.77	33.42 ± 5.83	69.03 ± 7.83	68.99 ± 7.92	68.87 ± 8.20	68.43 ± 9.09
(0.1 , 0.4)	45.76 ± 5.85	45.76 ± 5.87	45.71 ± 5.97	45.53 ± 6.31	93.31 ± 8.23	93.21 ± 8.61	92.87 ± 9.50	91.82 ± 11.83
(0.1 , 0.5)	58.02 ± 5.87	57.98 ± 5.98	57.84 ± 6.29	57.49 ± 7.12	117.52 ± 8.72	117.27 ± 9.56	116.70 ± 11.30	114.98 ± 15.02
(0.1 , 0.6)	70.24 ± 5.83	70.15 ± 6.13	69.94 ± 6.74	69.32 ± 8.23	141.67 ± 9.40	141.28 ± 10.82	140.34 ± 13.54	137.73 ± 18.87
(0.1 , 0.7)	82.41 ± 5.93	82.26 ± 6.49	81.91 ± 7.55	81.12 ± 9.53	165.70 ± 10.43	165.06 ± 12.74	163.75 ± 16.41	160.77 ± 22.40
(0.1 , 0.8)	94.47 ± 6.42	94.28 ± 7.12	93.76 ± 8.70	92.94 ± 10.66	189.61 ± 12.32	188.79 ± 15.15	187.15 ± 19.52	184.91 ± 24.02
(0.1 , 0.9)	106.38 ± 6.83	106.13 ± 7.78	105.72 ± 9.15	105.49 ± 9.87	212.58 ± 16.76	211.94 ± 18.74	211.42 ± 20.27	211.09 ± 21.13
(0.1 , 1)	119.05 ± 0.22	119.05 ± 0.22	119.05 ± 0.22	119.05 ± 0.22	239.05 ± 0.22	239.05 ± 0.22	239.05 ± 0.22	239.05 ± 0.22
(0.2 , 0.2)	24.00 ± 4.38	24.00 ± 4.38	24.00 ± 4.38	24.01 ± 4.38	48.01 ± 6.20	48.00 ± 6.20	47.98 ± 6.19	47.98 ± 6.22
(0.2 , 0.3)	33.01 ± 5.93	33.01 ± 5.93	33.01 ± 5.93	32.98 ± 5.96	67.69 ± 8.99	67.68 ± 9.02	67.65 ± 9.08	67.42 ± 9.41
(0.2 , 0.4)	45.12 ± 6.64	45.10 ± 6.66	45.08 ± 6.70	44.93 ± 6.90	92.54 ± 9.23	92.43 ± 9.49	92.18 ± 10.12	91.24 ± 11.99
(0.2 , 0.5)	57.64 ± 6.39	57.60 ± 6.49	57.47 ± 6.77	57.04 ± 7.59	117.10 ± 9.31	116.78 ± 10.27	116.06 ± 12.20	113.97 ± 16.32
(0.2 , 0.6)	70.02 ± 6.13	69.89 ± 6.49	69.62 ± 7.23	68.73 ± 9.13	141.25 ± 10.32	140.69 ± 12.13	139.35 ± 15.56	135.60 ± 21.89
(0.2 , 0.7)	82.22 ± 6.31	81.98 ± 7.13	81.43 ± 8.61	80.17 ± 11.20	165.21 ± 11.92	164.17 ± 15.13	162.08 ± 19.86	157.40 ± 27.12
(0.2 , 0.8)	94.16 ± 7.36	93.82 ± 8.41	92.91 ± 10.65	91.48 ± 13.28	188.76 ± 15.07	187.35 ± 18.88	184.41 ± 24.82	180.31 ± 30.75
(0.2 , 0.9)	105.84 ± 8.71	105.37 ± 10.06	104.57 ± 12.04	104.12 ± 13.07	210.53 ± 22.17	209.29 ± 24.89	208.13 ± 27.20	207.35 ± 28.66
(0.2 , 1)	119.10 ± 0.30	119.10 ± 0.30	119.10 ± 0.30	119.10 ± 0.30	239.10 ± 0.30	239.10 ± 0.30	239.10 ± 0.30	239.10 ± 0.30
(0.3 , 0.3)	36.00 ± 5.02	36.00 ± 5.02	36.00 ± 5.01	35.99 ± 5.03	72.00 ± 7.10	71.99 ± 7.10	71.99 ± 7.12	71.96 ± 7.20
(0.3 , 0.4)	44.81 ± 6.42	44.80 ± 6.41	44.79 ± 6.43	44.73 ± 6.53	91.19 ± 9.89	91.17 ± 9.94	91.06 ± 10.14	90.66 ± 10.91
(0.3 , 0.5)	56.89 ± 7.18	56.88 ± 7.22	56.80 ± 7.35	56.49 ± 7.83	116.15 ± 10.36	115.98 ± 10.85	115.48 ± 12.13	113.89 ± 15.23
(0.3 , 0.6)	69.56 ± 6.80	69.46 ± 7.07	69.22 ± 7.64	68.41 ± 9.23	140.82 ± 10.78	140.29 ± 12.43	138.96 ± 15.73	135.22 ± 21.96
(0.3 , 0.7)	82.00 ± 6.62	81.73 ± 7.44	81.14 ± 8.95	79.69 ± 11.71	164.78 ± 12.90	163.64 ± 16.17	161.27 ± 21.14	155.85 ± 28.78
(0.3 , 0.8)	93.96 ± 7.74	93.49 ± 9.10	92.33 ± 11.64	90.44 ± 14.73	188.08 ± 16.85	186.32 ± 21.05	182.59 ± 27.57	177.14 ± 34.22
(0.3 , 0.9)	105.39 ± 9.96	104.70 ± 11.65	103.63 ± 13.88	102.83 ± 15.33	208.85 ± 25.61	207.26 ± 28.58	205.41 ± 31.57	203.91 ± 33.75
(0.3 , 1)	119.15 ± 0.38	119.15 ± 0.37	119.15 ± 0.41	119.15 ± 0.38	239.15 ± 0.38	239.15 ± 0.36	239.15 ± 0.36	239.15 ± 0.40
(0.4 , 0.4)	47.99 ± 5.37	48.01 ± 5.36	48.00 ± 5.37	47.97 ± 5.42	95.99 ± 7.60	96.01 ± 7.60	95.97 ± 7.66	95.87 ± 7.96
(0.4 , 0.5)	56.72 ± 6.66	56.70 ± 6.69	56.68 ± 6.73	56.54 ± 6.95	114.96 ± 10.39	114.92 ± 10.51	114.70 ± 10.99	113.96 ± 12.52
(0.4 , 0.6)	68.80 ± 7.45	68.75 ± 7.58	68.60 ± 7.88	68.03 ± 8.91	139.94 ± 11.26	139.60 ± 12.25	138.68 ± 14.55	135.95 ± 19.56
(0.4 , 0.7)	81.55 ± 7.18	81.32 ± 7.77	80.83 ± 8.96	79.52 ± 11.41	164.42 ± 12.97	163.38 ± 15.97	161.16 ± 20.72	155.97 ± 28.17
(0.4 , 0.8)	93.78 ± 7.87	93.28 ± 9.29	92.00 ± 11.95	89.90 ± 15.23	187.66 ± 17.65	185.75 ± 21.99	181.65 ± 28.76	175.27 ± 35.94
(0.4 , 0.9)	105.08 ± 10.67	104.20 ± 12.64	102.86 ± 15.16	101.78 ± 16.89	207.56 ± 27.89	205.51 ± 31.25	203.15 ± 34.61	200.80 ± 37.49
(0.4 , 1)	119.20 ± 0.69	119.19 ± 0.76	119.19 ± 0.80	119.19 ± 0.82	239.20 ± 0.79	239.20 ± 0.69	239.20 ± 0.71	239.20 ± 0.64
(0.5 , 0.5)	60.00 ± 5.47	60.00 ± 5.48	59.99 ± 5.51	59.93 ± 5.64	120.00 ± 7.75	120.00 ± 7.78	119.93 ± 7.96	119.74 ± 8.68
(0.5 , 0.6)	68.70 ± 6.72	68.67 ± 6.77	68.60 ± 6.91	68.30 ± 7.47	138.88 ± 10.64	138.72 ± 11.00	138.33 ± 12.04	136.95 ± 15.10
(0.5 , 0.7)	80.83 ± 7.59	80.71 ± 7.89	80.37 ± 8.61	79.43 ± 10.42	163.74 ± 12.46	163.08 ± 14.44	161.52 ± 18.17	157.48 ± 24.85
(0.5 , 0.8)	93.47 ± 7.93	93.02 ± 9.15	91.90 ± 11.49	89.86 ± 14.74	187.49 ± 17.27	185.65 ± 21.56	181.77 ± 28.08	175.17 ± 35.65
(0.5 , 0.9)	104.89 ± 10.89	103.91 ± 13.03	102.41 ± 15.72	101.05 ± 17.75	206.88 ± 28.93	204.42 ± 32.73	201.47 ± 36.56	198.22 ± 40.11
(0.5 , 1)	119.22 ± 1.44	119.20 ± 1.78	119.19 ± 1.93	119.19 ± 2.01	239.22 ± 1.96	239.22 ± 1.88	239.22 ± 2.06	239.22 ± 1.92
(0.6 , 0.6)	72.00 ± 5.36	71.99 ± 5.39	71.96 ± 5.47	71.86 ± 5.78	143.98 ± 7.63	143.97 ± 7.74	143.87 ± 8.19	143.39 ± 9.97
(0.6 , 0.7)	80.73 ± 6.64	80.68 ± 6.77	80.51 ± 7.19	80.00 ± 8.31	162.89 ± 10.84	162.60 ± 11.67	161.85 ± 13.85	159.65 ± 18.64
(0.6 , 0.8)	92.87 ± 7.83	92.61 ± 8.59	91.78 ± 10.36	90.18 ± 13.10	187.27 ± 15.29	185.92 ± 18.87	182.90 ± 24.69	177.27 ± 32.16
(0.6 , 0.9)	104.83 ± 10.38	103.87 ± 12.56	102.42 ± 15.22	100.90 ± 17.51	206.85 ± 28.57	204.05 ± 32.89	200.82 ± 37.04	196.72 ± 41.30
(0.6 , 1)	119.19 ± 2.64	119.10 ± 3.55	119.05 ± 3.96	119.01 ± 4.27	239.08 ± 5.18	239.08 ± 5.16	239.08 ± 5.23	239.08 ± 5.24
(0.7 , 0.7)	83.98 ± 5.06	83.97 ± 5.12	83.91 ± 5.33	83.71 ± 6.04	167.97 ± 7.26	167.92 ± 7.56	167.70 ± 8.65	166.91 ± 11.79
(0.7 , 0.8)	92.75 ± 6.53	92.61 ± 6.97	92.18 ± 8.13	91.23 ± 10.24	186.86 ± 11.80	186.27 ± 13.85	184.68 ± 17.98	181.25 ± 24.35
(0.7 , 0.9)	104.61 ± 9.35	103.93 ± 11.05	102.83 ± 13.34	101.53 ± 15.54	207.76 ± 25.40	205.19 ± 29.90	202.06 ± 34.35	197.66 ± 39.36
(0.7 , 1)	119.06 ± 4.09	118.75 ± 6.03	118.50 ± 7.18	118.24 ± 8.21	238.22 ± 11.68	238.15 ± 12.06	238.17 ± 11.96	238.13 ± 12.12
(0.8 , 0.8)	95.97 ± 4.56	95.93 ± 4.76	95.76 ± 5.49	95.38 ± 6.92	191.89 ± 6.91	191.73 ± 7.95	191.19 ± 10.60	189.94 ± 15.05
(0.8 , 0.9)	104.85 ± 7.00	104.55 ± 8.00	103.99 ± 9.60	103.33 ± 11.18	209.43 ± 17.93	208.21 ± 21.02	206.44 ± 24.77	203.54 ± 29.64
(0.8 , 1)	118.86 ± 5.17	118.23 ± 8.10	117.54 ± 10.32	116.49 ± 12.97	234.98 ± 22.77	234.33 ± 24.33	234.27 ± 24.49	234.16 ± 24.75
(0.9 , 0.9)	107.93 ± 3.84	107.85 ± 4.35	107.72 ± 5.14	107.52 ± 6.10	215.58 ± 8.21	215.33 ± 9.67	214.95 ± 11.62	214.22 ± 14.60
(0.9 , 1)	118.22 ± 5.17	117.60 ± 8.01	116.75 ± 10.65	115.08 ± 14.40	230.22 ± 31.29	227.22 ± 35.89	225.72 ± 37.94	224.30 ± 39.77
(1 , 1)	120.00 ± 0.00	120.00 ± 0.00	120.00 ± 0.00	120.00 ± 0.00	240.00 ± 0.00	240.00 ± 0.00	240.00 ± 0.00	240.00 ± 0.00

Table 6.2: The numerical subject benefit results for different significance levels in the model case *DPWOI*. Each cell is composed of the average number of success responses (first component) added to/subtracted from the corresponding standard deviation (second component) for each scenario (θ_C, θ_D).

	model case *DPWI* $[T = 120]$				**model case *DPWI*** $[T = 240]$			
(θ_C, θ_D)	$\alpha = 0.0025$	$\alpha = 0.005$	$\alpha = 0.01$	$\alpha = 0.02$	$\alpha = 0.0025$	$\alpha = 0.005$	$\alpha = 0.01$	$\alpha = 0.02$
(0 , 0)	0.00 ± 0.00	0.00 ± 0.00	0.00 ± 0.00	0.00 ± 0.00	0.00 ± 0.00	0.00 ± 0.00	0.00 ± 0.00	0.00 ± 0.00
(0 , 0.1)	10.15 ± 3.49	10.15 ± 3.49	10.14 ± 3.50	10.14 ± 3.49	21.73 ± 4.84	21.73 ± 4.84	21.74 ± 4.84	21.74 ± 4.84
(0 , 0.2)	21.95 ± 4.59	21.95 ± 4.60	21.94 ± 4.59	21.95 ± 4.60	45.55 ± 6.38	45.56 ± 6.37	45.55 ± 6.38	45.56 ± 6.37
(0 , 0.3)	33.96 ± 5.22	33.95 ± 5.23	33.96 ± 5.22	33.96 ± 5.23	69.59 ± 7.26	69.58 ± 7.26	69.59 ± 7.27	69.59 ± 7.26
(0 , 0.4)	46.05 ± 5.57	46.05 ± 5.56	46.05 ± 5.57	46.06 ± 5.56	93.72 ± 7.76	93.72 ± 7.75	93.71 ± 7.76	93.70 ± 7.74
(0 , 0.5)	58.22 ± 5.67	58.21 ± 5.67	58.21 ± 5.66	58.22 ± 5.67	117.91 ± 7.89	117.91 ± 7.91	117.91 ± 7.91	117.91 ± 7.90
(0 , 0.6)	70.42 ± 5.55	70.42 ± 5.55	70.42 ± 5.55	70.42 ± 5.55	142.17 ± 7.74	142.16 ± 7.74	142.16 ± 7.74	142.15 ± 7.74
(0 , 0.7)	82.67 ± 5.21	82.68 ± 5.21	82.67 ± 5.20	82.66 ± 5.20	166.43 ± 7.23	166.45 ± 7.24	166.45 ± 7.23	166.42 ± 7.23
(0 , 0.8)	94.93 ± 4.50	94.93 ± 4.50	94.92 ± 4.50	94.93 ± 4.50	190.80 ± 6.34	190.81 ± 6.33	190.81 ± 6.33	190.80 ± 6.33
(0 , 0.9)	107.02 ± 3.32	107.02 ± 3.32	107.02 ± 3.32	107.02 ± 3.32	215.01 ± 4.68	215.01 ± 4.68	215.01 ± 4.68	215.01 ± 4.67
(0 , 1)	119.00 ± 0.00	119.00 ± 0.00	119.00 ± 0.00	119.00 ± 0.00	239.00 ± 0.00	239.00 ± 0.00	239.00 ± 0.00	239.00 ± 0.00
(0.1 , 0.1)	11.99 ± 3.28	12.00 ± 3.29	12.00 ± 3.28	12.00 ± 3.29	24.00 ± 4.65	24.01 ± 4.65	24.00 ± 4.65	24.00 ± 4.65
(0.1 , 0.2)	21.40 ± 5.07	21.41 ± 5.07	21.40 ± 5.07	21.39 ± 5.07	44.51 ± 7.45	44.51 ± 7.45	44.48 ± 7.47	44.41 ± 7.56
(0.1 , 0.3)	33.50 ± 5.75	33.49 ± 5.74	33.47 ± 5.77	33.43 ± 5.83	69.02 ± 7.82	68.99 ± 7.92	68.86 ± 8.20	68.42 ± 9.09
(0.1 , 0.4)	45.77 ± 5.85	45.77 ± 5.87	45.71 ± 5.98	45.53 ± 6.31	93.32 ± 8.22	93.19 ± 8.63	92.88 ± 9.51	91.82 ± 11.85
(0.1 , 0.5)	58.02 ± 5.86	57.98 ± 5.96	57.86 ± 6.30	57.48 ± 7.14	117.53 ± 8.71	117.28 ± 9.57	116.68 ± 11.35	114.96 ± 15.08
(0.1 , 0.6)	70.24 ± 5.83	70.16 ± 6.12	69.95 ± 6.74	69.31 ± 8.24	141.66 ± 9.40	141.28 ± 10.81	140.33 ± 13.56	137.74 ± 18.84
(0.1 , 0.7)	82.41 ± 5.94	82.27 ± 6.49	81.91 ± 7.56	81.11 ± 9.57	165.70 ± 10.38	165.05 ± 12.76	163.75 ± 16.42	160.80 ± 22.38
(0.1 , 0.8)	94.47 ± 6.43	94.29 ± 7.13	93.77 ± 8.68	92.97 ± 10.62	189.62 ± 12.32	188.78 ± 15.18	187.13 ± 19.54	184.92 ± 24.02
(0.1 , 0.9)	106.38 ± 6.82	106.13 ± 7.77	105.73 ± 9.12	105.50 ± 9.84	212.59 ± 16.74	211.95 ± 18.73	211.38 ± 20.34	211.10 ± 21.11
(0.1 , 1)	119.05 ± 0.22	119.05 ± 0.22	119.05 ± 0.22	119.10 ± 0.30	239.05 ± 0.22	239.05 ± 0.22	239.05 ± 0.22	239.05 ± 0.22
(0.2 , 0.2)	24.00 ± 4.38	24.00 ± 4.38	24.01 ± 4.39	24.00 ± 4.38	48.01 ± 6.19	47.99 ± 6.20	48.01 ± 6.20	47.99 ± 6.22
(0.2 , 0.3)	33.01 ± 5.92	33.02 ± 5.92	33.01 ± 5.92	32.97 ± 5.94	67.70 ± 9.00	67.67 ± 9.01	67.63 ± 9.10	67.42 ± 9.39
(0.2 , 0.4)	45.12 ± 6.63	45.11 ± 6.65	45.07 ± 6.71	44.93 ± 6.91	92.54 ± 9.24	92.45 ± 9.48	92.19 ± 10.13	91.26 ± 11.99
(0.2 , 0.5)	57.65 ± 6.40	57.59 ± 6.50	57.48 ± 6.77	57.06 ± 7.57	117.11 ± 9.32	116.79 ± 10.30	116.07 ± 12.18	113.97 ± 16.34
(0.2 , 0.6)	70.02 ± 6.13	69.90 ± 6.51	69.60 ± 7.24	68.73 ± 9.13	141.26 ± 10.30	140.69 ± 12.13	139.35 ± 15.54	135.67 ± 21.79
(0.2 , 0.7)	82.22 ± 6.31	81.98 ± 7.13	81.43 ± 8.60	80.18 ± 11.19	165.18 ± 12.08	164.16 ± 15.14	162.08 ± 19.83	157.42 ± 27.07
(0.2 , 0.8)	94.17 ± 7.33	93.81 ± 8.43	92.90 ± 10.64	91.46 ± 13.33	188.73 ± 15.15	187.33 ± 18.93	184.39 ± 24.84	180.24 ± 30.84
(0.2 , 0.9)	105.83 ± 8.75	105.37 ± 10.07	104.58 ± 12.03	104.14 ± 13.06	210.54 ± 22.15	209.34 ± 24.80	208.11 ± 27.27	207.28 ± 28.76
(0.2 , 1)	119.10 ± 0.30	119.10 ± 0.30	119.10 ± 0.31	119.20 ± 0.40	239.10 ± 0.30	239.10 ± 0.30	239.10 ± 0.30	239.10 ± 0.30
(0.3 , 0.3)	36.00 ± 5.02	36.01 ± 5.02	35.99 ± 5.02	35.99 ± 5.03	72.00 ± 7.10	71.99 ± 7.11	71.99 ± 7.12	71.94 ± 7.20
(0.3 , 0.4)	44.82 ± 6.42	44.81 ± 6.42	44.80 ± 6.43	44.73 ± 6.52	91.20 ± 9.88	91.17 ± 9.94	91.08 ± 10.13	90.68 ± 10.88
(0.3 , 0.5)	56.89 ± 7.18	56.87 ± 7.21	56.79 ± 7.35	56.50 ± 7.84	116.16 ± 10.38	115.97 ± 10.87	115.49 ± 12.09	113.89 ± 15.24
(0.3 , 0.6)	69.55 ± 6.80	69.45 ± 7.07	69.21 ± 7.65	68.42 ± 9.22	140.82 ± 10.75	140.26 ± 12.47	138.96 ± 15.76	135.24 ± 21.93
(0.3 , 0.7)	81.99 ± 6.62	81.73 ± 7.44	81.11 ± 8.98	79.70 ± 11.69	164.79 ± 12.83	163.67 ± 16.06	161.29 ± 21.08	155.85 ± 28.77
(0.3 , 0.8)	93.97 ± 7.74	93.50 ± 9.12	92.33 ± 11.63	90.43 ± 14.74	188.11 ± 16.79	186.31 ± 21.08	182.60 ± 27.56	177.03 ± 34.35
(0.3 , 0.9)	105.39 ± 9.96	104.68 ± 11.70	103.61 ± 13.91	102.88 ± 15.32	208.85 ± 25.61	207.17 ± 28.73	205.41 ± 31.59	203.88 ± 33.80
(0.3 , 1)	119.15 ± 0.38	119.15 ± 0.40	119.15 ± 0.42	119.30 ± 0.46	239.15 ± 0.40	239.15 ± 0.36	239.15 ± 0.38	239.15 ± 0.36
(0.4 , 0.4)	47.99 ± 5.37	48.00 ± 5.36	47.98 ± 5.37	47.97 ± 5.41	96.01 ± 7.59	96.00 ± 7.60	95.98 ± 7.65	95.87 ± 7.97
(0.4 , 0.5)	56.71 ± 6.67	56.70 ± 6.67	56.67 ± 6.72	56.52 ± 6.96	114.96 ± 10.37	114.91 ± 10.53	114.69 ± 10.98	113.96 ± 12.56
(0.4 , 0.6)	68.80 ± 7.45	68.74 ± 7.57	68.60 ± 7.89	68.04 ± 8.90	139.93 ± 11.29	139.58 ± 12.27	138.67 ± 14.56	135.93 ± 19.56
(0.4 , 0.7)	81.55 ± 7.18	81.35 ± 7.75	80.82 ± 8.98	79.52 ± 11.44	164.46 ± 12.88	163.41 ± 15.96	161.16 ± 20.72	155.97 ± 28.19
(0.4 , 0.8)	93.79 ± 7.85	93.27 ± 9.29	92.02 ± 11.91	89.89 ± 15.24	187.67 ± 17.67	185.73 ± 22.04	181.69 ± 28.72	175.22 ± 35.98
(0.4 , 0.9)	105.07 ± 10.69	104.20 ± 12.65	102.88 ± 15.12	101.81 ± 16.89	207.64 ± 27.76	205.50 ± 31.27	203.18 ± 34.56	200.71 ± 37.59
(0.4 , 1)	119.19 ± 0.70	119.19 ± 0.79	119.19 ± 0.76	119.40 ± 0.49	239.20 ± 0.71	239.20 ± 0.65	239.20 ± 0.68	239.20 ± 0.59
(0.5 , 0.5)	60.00 ± 5.48	60.00 ± 5.49	59.99 ± 5.51	59.93 ± 5.63	120.00 ± 7.76	119.99 ± 7.80	119.94 ± 7.97	119.72 ± 8.71
(0.5 , 0.6)	68.70 ± 6.71	68.67 ± 6.76	68.60 ± 6.92	68.30 ± 7.48	138.85 ± 10.64	138.73 ± 10.99	138.34 ± 11.99	136.95 ± 15.11
(0.5 , 0.7)	80.82 ± 7.60	80.71 ± 7.89	80.36 ± 8.63	79.44 ± 10.41	163.76 ± 12.45	163.07 ± 14.44	161.49 ± 18.19	157.50 ± 24.78
(0.5 , 0.8)	93.45 ± 7.96	93.01 ± 9.16	91.90 ± 11.49	89.86 ± 14.74	187.49 ± 17.23	185.57 ± 21.68	181.73 ± 28.15	175.18 ± 35.65
(0.5 , 0.9)	104.89 ± 10.89	103.92 ± 13.03	102.39 ± 15.74	101.07 ± 17.77	206.84 ± 28.99	204.37 ± 32.79	201.48 ± 36.55	198.25 ± 40.10
(0.5 , 1)	119.22 ± 1.40	119.20 ± 1.75	119.19 ± 1.94	119.50 ± 0.50	239.22 ± 1.81	239.22 ± 2.02	239.22 ± 1.95	239.22 ± 1.94
(0.6 , 0.6)	71.99 ± 5.37	71.98 ± 5.39	71.97 ± 5.46	71.86 ± 5.80	143.99 ± 7.64	143.95 ± 7.77	143.88 ± 8.19	143.37 ± 10.03
(0.6 , 0.7)	80.73 ± 6.64	80.67 ± 6.78	80.51 ± 7.17	79.99 ± 8.32	162.87 ± 10.84	162.60 ± 11.68	161.86 ± 13.83	159.69 ± 18.58
(0.6 , 0.8)	92.87 ± 7.82	92.59 ± 8.59	91.79 ± 10.34	90.23 ± 13.08	187.24 ± 15.38	185.87 ± 18.94	182.87 ± 24.75	177.38 ± 32.07
(0.6 , 0.9)	104.84 ± 10.36	103.88 ± 12.55	102.42 ± 15.24	100.95 ± 17.53	206.82 ± 28.62	204.08 ± 32.86	200.81 ± 37.03	196.76 ± 41.26
(0.6 , 1)	119.19 ± 2.63	119.09 ± 3.58	119.04 ± 4.03	119.60 ± 0.49	239.08 ± 5.17	239.09 ± 5.05	239.08 ± 5.16	239.08 ± 5.16
(0.7 , 0.7)	83.99 ± 5.05	83.97 ± 5.12	83.92 ± 5.33	83.71 ± 6.05	167.97 ± 7.25	167.91 ± 7.62	167.71 ± 8.63	166.91 ± 11.78
(0.7 , 0.8)	92.76 ± 6.54	92.62 ± 6.95	92.18 ± 8.14	91.25 ± 10.23	186.88 ± 11.78	186.26 ± 13.82	184.66 ± 18.04	181.23 ± 24.39
(0.7 , 0.9)	104.62 ± 9.34	103.94 ± 11.02	102.84 ± 13.33	101.60 ± 15.54	207.67 ± 25.56	205.20 ± 29.88	202.02 ± 34.43	197.65 ± 39.38
(0.7 , 1)	119.06 ± 4.09	118.75 ± 5.98	118.51 ± 7.12	119.70 ± 0.46	238.21 ± 11.72	238.18 ± 11.88	238.16 ± 11.96	238.15 ± 12.01
(0.8 , 0.8)	95.97 ± 4.56	95.92 ± 4.76	95.74 ± 5.54	95.37 ± 6.94	191.89 ± 6.93	191.73 ± 7.96	191.22 ± 10.51	189.96 ± 15.04
(0.8 , 0.9)	104.84 ± 7.00	104.55 ± 8.02	104.00 ± 9.59	103.38 ± 11.22	209.47 ± 17.81	208.24 ± 20.96	206.40 ± 24.82	203.49 ± 29.70
(0.8 , 1)	118.86 ± 5.19	118.22 ± 8.12	117.53 ± 10.35	119.80 ± 0.40	234.96 ± 22.81	234.39 ± 24.20	234.27 ± 24.49	234.21 ± 24.64
(0.9 , 0.9)	107.92 ± 3.86	107.86 ± 4.31	107.71 ± 5.14	107.56 ± 5.95	215.59 ± 8.15	215.35 ± 9.61	214.95 ± 11.63	214.22 ± 14.62
(0.9 , 1)	118.23 ± 5.14	117.60 ± 8.00	116.76 ± 10.63	119.90 ± 0.30	230.24 ± 31.24	227.22 ± 35.90	225.75 ± 37.90	224.21 ± 39.88
(1 , 1)	120.00 ± 0.00	120.00 ± 0.00	120.00 ± 0.00	120.00 ± 0.00	240.00 ± 0.00	240.00 ± 0.00	240.00 ± 0.00	240.00 ± 0.00

Table 6.3: The numerical subject benefit results for different significance levels in the model case *DPWI*. Each cell is composed of the average number of success responses (first component) added to/subtracted from the corresponding standard deviation (second component) for each scenario (θ_C, θ_D) .

Chapter 7

Conclusion and Further Work

7.1 Summary and Contributions

This thesis examines the estimation problem in response-adaptive procedures, namely, two-armed bandit models. In particular, we have proposed a range of randomisation designs based on solutions to the multi-armed bandit problem (MABP) with the aim of balancing the trade-off between reducing the bias in the estimation process whilst maintaining subject benefits, i.e. the total number of success responses is maximised. It is worth mentioning that each design was associated with merits and demerits that inhibit the bandit-based design from being implemented in practice. We briefly summarise the main points made in each chapter and, in turn, highlight the corresponding contributions together with areas of further investigation.

7.1.1 Chapter 3, Model and After-trial Studies

In chapter 2, we formulated a Bayesian Beta-Bernoulli finite-horizon two-armed bandit problem with binary responses with the objective function of maximising the Bayes-expected total number of subject successes in the trial. The model's performance in using the MLE to estimate success probabilities while using a DP solution for subject allocation was assessed when using a traditional method of correcting the MLE estimation called after-trial studies, although the results were not good enough. We learnt that some computational issues must be considered when implementing the after-trial studies. As an area of improvement, we mention the relationship between the size of the entire population of the trial and available computational capacity. Moreover, applying proposed alternative designs and novel estimators instead of the classical DP design and MLE, respectively, can be another potential area of improvement.

Moreover, we believe that after-trial studies have much room to be improved, most of which with much scope here might be considered potential candidates for some serious future works. In order to gain a more profound insight into the relationship between the level of precision and the type of randomisation occurring in the course of the trial, (i) to investigate the frequency of 50 : 50 randomisation happening in the trial, we recommend comparing the optimal arm matrices obtained from the DP procedure to one another for the cases mentioned above, (ii) to show the trade-off between losing the subject benefit and unbiased estimation, we advise on calculating the patient benefit results in terms of the average total number of success observations at the end of the trial, for a different level of precision used for the DP procedure, i.e. for case (i) and (ii).

7.1.2 Chapter 4, Augmented Estimator

An augmented estimator with the aim of mitigating the bias of the MLE was introduced and mathematically investigated in this chapter. We showed it can potentially be in either the frequentist or Bayesian sense. Moreover, we showed that since DP satisfies the *exploit* property proposed in Nie et al. (2018), the derived bias of the MLE is always negative, but this is not the case with the Bayesian augmented estimator. We evaluated the performance of the augmented estimator performance with different carefully chosen augmentations and compared it to other estimators, such as the inverse probability weighted (IPW) estimator in the literature. Finding the optimal augmentation coefficient, which will be directly correlated with the assumed design's matrices, and applying the augmented estimator to previous adaptive designs, are possible areas of future work.

7.1.3 Chapter 5, OIDP and RDP

The optimistic on inferior dynamic programming (OIDP) design and the randomised dynamic programming (RDP) design were discussed in this chapter. Please note that RDP is equivalent to bi-level randomisation procedure which has been applied to the Bayesian Beta-Bernoulli two-armed bandit model for the first time in the literature. Both are novel allocation procedures that aim to mitigate the bias induced by the DP procedure within the trial by modifying the allocation decision at every time step. OIDP was evaluated by applying different pseudo success observations, whilst the RDP was examined by varying the degree of randomisation. Simulation results from both procedures illustrate the trade-off between reducing the bias and increasing the suboptimality in terms of cumulative reward. As an extension of this work, we can mention learning the optimal $\dot{s}$ in the OIDP design, which needs a meticulous mathematical investigation of the DP algorithm. Another extension can be applied to the RDP or bi-level randomisation procedure by filtering observations based on the branch they obtained. We mean defining an 8-dimensional vector of observations containing the success and failure responses on both arms for fixed randomisation (FR) and deterministic DP branches, i.e. $\boldsymbol{x} := \Big(s_{C,EFR}, f_{C,EFR}, s_{D,EFR}, f_{D,EFR}, s_{C,DP}, f_{C,DP}, s_{D,DP}, f_{D,DP}\Big)$. Then, based on the posterior distribution calculated by equation (3.1) allocation policies can be formed in a way that the MLE estimates with less bias.

7.1.4 Chapter 6, DPWOI, DPWI

Formulating response-adaptive designs where an interim analysis was implemented in the middle of the trial was another novel investigation in the classical Bayesian Beta-Bernoulli two-armed bandit model. After developing some stopping criteria for a non-trivial interim analysis in the middle of the time horizon, we assume two different circumstances: (i) setting an interim analysis through the simulation step whilst the classical DP solution is used for estimation, i.e. DPWOI (ii) considering an identical interim analysis condition in both DP and the simulation step, i.e. DPWI. The latter case showed slightly less biased results in comparison with the former. However, to draw a more decisive conclusion, one needs to set up multiple interim inspections of the data instead of a single one. Doing so will be associated with some statistical complexity between the interim analysis points. Hence, it might be considered a useful potential line of enquiry for this context.

7.2 Future Work

Apart from all the aforementioned potentials for future work, in this part, we briefly list some other general areas that need improvement. One can also consider expanding the practical framework of covariate-adaptive RAR design offered by Ji et al. (2019) in which randomisation probabilities depend on both covariates and patients' responses. Note that, in the commentary by Saville and Meurer (2019), some objections associated with the study of Ji et al. (2019) have been raised. The foremost ones that can serve as a motivation for future studies are the difficulty of extrapolating and poor understanding of statistical properties such as power, bias, etc., related to a multi-armed context. Another practical and useful problem for future study can be obtained by considering the time-trend with or without an early stopping assumption in a multi-armed bandit setting. The details on the two-armed case have been investigated in the work of (Jiang et al., 2020).

The practical implications of this study are countless. Generally, one can take advantage of the present study in many decision-making contexts. For instance, the first in the list belongs to clinical trials where the efficacy of an experimental treatment arm is being estimated within a trial compared with the control treatment arm. As a rule in the clinical trials setting, the treatment with the smallest true efficacy shows the largest bias, and this bias grows as the difference between the superior and inferior treatment increases (Bowden and Trippa, 2017). Benefiting from our proposed estimator, the bias can be substantially reduced, and the treatment efficacy can be estimated more accurately. Digital marketing and social networks are among the other practical settings in which issues discussed in this paper might arise. For example, an augmented estimator may give the decision maker more profound insight into the popularity of a post or product on Instagram when people, who follow the page, like or dislike the post. In summary, the RAR design is useful in many settings to improve the overall response in the trial. Although the arm effect estimation is usually associated with a relatively small and negative bias, we propose a framework within which the bias is mitigated significantly. In order to have this unavoidable bias converge to zero, more studies need to be carried out to adjust the augmentation appropriately. Note that, in this thesis, we only focus on DP designs, while one can apply the proposed contributions in other randomisation procedures to improve the performance of the estimators.

Another general potential direction for future work could be applying all mitigation bias techniques, aside from implementing them post or during the trial, to a multi-arm bandit setting with binary responses. It is evident that extension to a multi-arm setting requires dealing with extra dimensions in the DP procedure, which in turn gives rise to computational limitations and complexity in the algorithm (Powell, 2007). To overcome the curse of dimensionality, some coding and programming techniques, together with taking advantage of computer machines with larger memory capacity and more advanced configuration, could be potential alternatives.

Furthermore, we strongly believe that using the augmented estimators introduced in chapter 4 in the OIDP designs discussed in chapter 5 can produce the most effective estimation results in terms of having less bias. Both novel areas need more investigation in tuning the parameters involved separately. For example, the augmented estimators should be appropriately formulated by tuning the parameters in the equation (4.19), and the OIDP by choosing $\acute{s}$ in the equation (5.1).

Although both require serious research and time, thinking about mixing these two mitigation approaches can be an intriguing area for further studies. Note that the relationship in tuning parameters can become sophisticated upon mixing these two approaches, but it plays a crucial role in estimation with minimal bias. However, drawing such a relationship and formulating the correlation involved requires broad research work and commitment.

Finally, to define a potential PhD project for the future, we recommend designing a group sequential procedure utilising either RDP or OIDP designs so that an appropriate (identical or different) augmented estimator is used for modifying up-to-date effectiveness estimation at each interim inspection look before entering the following interim point. We actually mean that combining all novel findings of this PhD project can be a fruitful starting point for more investigation to produce unbiased estimation in this particular bandit context.

References

Agresti, A. and Caffo, B. (2000). Simple and effective confidence intervals for proportions and differences of proportions result from adding two successes and two failures. *The American Statistician*, 54(4):280–288.

Ahuja, V. and Birge, J. R. (2016). Response-adaptive designs for clinical trials: Simultaneous learning from multiple patients. *European Journal of Operational Research*, 248(2):619–633.

Amberson Jr, J. B., McMahon, B., and Pinner, M. (1931). A clinical trial of sanocrysin in pulmonary tuberculosis. *American Review of Tuberculosis*, 24(4):401–435.

Bauer, P., Koenig, F., Brannath, W., and Posch, M. (2010). Selection and bias—two hostile brothers. *Statistics in Medicine*, 29(1):1–13.

Bellman, R. (1956). A problem in the sequential design of experiments. *Sankhyā: The Indian Journal of Statistics (1933-1960)*, 16(3/4):221–229.

Bellman, R. (1957). *Dynamic Programming.* Princeton University Press, Princeton, NJ, USA.

Bellman, R. (1961). Adaptive control processes; a guided tour. *Princeton University Press.*

Bellman, R. (1966). Dynamic programming. *Science*, 153(3731):34–37.

Berry, D. A. (1978). Modified two-armed bandit strategies for certain clinical trials. *Journal of the American Statistical Association*, 73(362):339–345.

Berry, D. A. (2005). Introduction to Bayesian methods iii: use and interpretation of Bayesian tools in design and analysis. *Clinical Trials*, 2(4):295–300.

Berry, D. A. and Eick, S. G. (1995). Adaptive assignment versus balanced randomization in clinical trials: a decision analysis. *Statistics in Medicine*, 14(3):231–246.

Berry, D. A. and Fristedt, B. (1985). *Bandit Problems: Sequential Allocation of Experiments.* Springer.

Berry, D. A. and Stangl, D. K. (1996). *Bayesian methods in health-related research.* New York: Marcel Dekker Inc, USA.

Boos, D. and Stefanski, L. (2013). *Essential Statistical Inference: Theory and Methods.* Springer Texts in Statistics. Springer New York, USA.

Bowden, J. and Trippa, L. (2017). Unbiased estimation for response adaptive clinical trials. *Statistical Methods in Medical Research*, 26(5):2376–2388.

Bubeck, S., Cesa-Bianchi, N., et al. (2012). Regret analysis of stochastic and nonstochastic multi-armed bandit problems. *Foundations and Trends® in Machine Learning*, 5(1):1–122.

Burtini, G., Loeppky, J., and Lawrence, R. (2015). A survey of online experiment design with the stochastic multi-armed bandit. *arXiv preprint arXiv:1510.00757*.

Cheng, Y. and Berry, D. A. (2007). Optimal adaptive randomized designs for clinical trials. *Biometrika*, 94(3):673–689.

Chow, S.-C. and Chang, M. (2008). Adaptive design methods in clinical trials–a review. *Orphanet Journal of Rare Diseases*, 3(1):1–13.

Chow, S.-C. and Liu, J.-p. (2008). *Design and analysis of clinical trials: concepts and methodologies*, volume 507. John Wiley & Sons, USA.

Coad, D. S. and Ivanova, A. (2001). Bias calculations for adaptive urn designs. *Sequential Analysis*, 20(3):91–116.

Craft, A. et al. (1998). The first randomised controlled trial. *Archives of Disease in Childhood*, 79(5):410–410.

Fisher, R. A. (1926). The arrangement of field experiments. *Journal of the Ministry of Agriculture*, 33:503–515.

Gittins, J., Glazebrook, K., and Weber, R. (2011). *Multi-armed Bandit Allocation Indices*. Wiley, USA.

Gittins, J. C. (1979). Bandit processes and dynamic allocation indices. *Journal of the Royal Statistical Society: Series B (Methodological)*, 41(2):148–164.

Hadad, V., Hirshberg, D. A., Zhan, R., Wager, S., and Athey, S. (2021). Confidence intervals for policy evaluation in adaptive experiments. *Proceedings of the National Academy of Sciences*, 118(15):e2014602118.

Han, L., Arfe, A., and Trippa, L. (2022). Sensitivity analyses of clinical trial designs: Selecting scenarios and summarizing operating characteristics. *arXiv preprint arXiv:2208.03887*.

Hardwick, J., Stout, Q. F., et al. (1991). Bandit strategies for ethical sequential allocation. *Computing Science and Statistics*, 23(6.1):421–424.

Hu, F. and Rosenberger, W. (2006). *The Theory of Response-Adaptive Randomization in Clinical Trials*. Wiley Series in Probability and Statistics. Wiley, USA.

Ivanova, A. and Rosenberger, W. (2001). Adaptive designs for clinical trials with highly successful treatments. *Drug Information Journal - DRUG INF J*, 35:1087–1093.

Jacko, P. (2019a). Binarybandit: An efficient Julia package for optimization and evaluation of the finite-horizon bandit problem with binary responses, Lancaster University Management School.

Jacko, P. (2019b). The finite-horizon two-armed bandit problem with binary responses: A multidisciplinary survey of the history, state of the art, and myths. *arXiv preprint arXiv:1906.10173*.

Jennison, C. and Turnbull, B. (1999). *Group Sequential Methods with Applications to Clinical Trials*. Chapman & Hall/CRC Interdisciplinary Statistics. Taylor & Francis, USA.

Ji, L., McShane, L. M., Krailo, M., and Sposto, R. (2019). Bias in retrospective analyses of biomarker effect using data from an outcome-adaptive randomized trial. *Clinical Trials*, 16(6):599–609.

Jiang, Y., Zhao, W., and Durkalski-Mauldin, V. (2017). Impact of adaptation algorithm, timing, and stopping boundaries on the performance of Bayesian response adaptive randomization in confirmative trials with a binary endpoint. *Contemporary Clinical Trials*, 62:114–120.

Jiang, Y., Zhao, W., and Durkalski-Mauldin, V. (2020). Time-trend impact on treatment estimation in two-arm clinical trials with a binary outcome and Bayesian response adaptive randomization. *Journal of Biopharmaceutical Statistics*, 30(1):69–88.

Kaufmann, E., Korda, N., and Munos, R. (2012). Thompson sampling: An asymptotically optimal finite-time analysis. In Bshouty, N. H., Stoltz, G., Vayatis, N., and Zeugmann, T., editors, *Algorithmic Learning Theory*, pages 199–213, Berlin, Heidelberg. Springer Berlin Heidelberg.

Lipsky, A. M. and Lewis, R. J. (2013). Response-adaptive decision-theoretic trial design: operating characteristics and ethics. *Statistics in Medicine*, 32(21):3752–3765.

Liu, Y. and Chen, Y. (2016). A bandit framework for strategic regression. In Lee, D., Sugiyama, M., Luxburg, U., Guyon, I., and Garnett, R., editors, *Advances in Neural Information Processing Systems*, volume 29. Curran Associates, Inc.

Luedtke, A. R. and Van Der Laan, M. J. (2016). Statistical inference for the mean outcome under a possibly non-unique optimal treatment strategy. *Annals of Statistics*, 44(2):713–742.

Marschner, I. C. (2021). A general framework for the analysis of adaptive experiments. *Statistical Science*, 36(3):465–492.

Merrell, D., Chandereng, T., and Park, Y. (2021). A markov decision process for response-adaptive randomization in clinical trials. *arXiv preprint arXiv:2109.14642*.

Nie, X., Tian, X., Taylor, J., and Zou, J. (2018). Why adaptively collected data have negative bias and how to correct for it. In *International Conference on Artificial Intelligence and Statistics*, pages 1261–1269. PMLR.

Palmer, C. R. and Rosenberger, W. F. (1999). Ethics and practice: alternative designs for phase iii randomized clinical trials. *Controlled Clinical Trials*, 20(2):172–186.

Panigrahi, S., Taylor, J., and Weinstein, A. (2016). Bayesian post-selection inference in the linear model. *arXiv preprint arXiv:1605.08824*, 28.

Peace, K. E. and Chen, D.-G. D. (2010). *Clinical trial methodology.* CRC Press.

Pham-Gia, T., Van Thin, N., Doan, P. P., et al. (2017). Inferences on the difference of two proportions: A Bayesian approach. *Open Journal of Statistics*, 7:1–15.

Pocock, S. J. (2013). *Clinical trials: a practical approach.* John Wiley & Sons, USA.

Powell, W. B. (2007). *Approximate Dynamic Programming: Solving the Curses of Dimensionality (Wiley Series in Probability and Statistics).* Wiley-Interscience, USA.

Proschan, M. and Evans, S. (2020). Resist the temptation of response-adaptive randomization. *Clinical Infectious Diseases*, 71(11):3002–3004.

Puterman, M. L. (2014). *Markov decision processes: discrete stochastic dynamic programming*. John Wiley & Sons, USA.

Rafferty, A., Ying, H., Williams, J., et al. (2019). Statistical consequences of using multi-armed bandits to conduct adaptive educational experiments. *Journal of Educational Data Mining*, 11(1):47–79.

Raineri, E., Dabad, M., and Heath, S. (2014). A note on exact differences between beta distributions in genomic (methylation) studies. *PLoS One*, 9(5):e97349.

Robbins, H. (1952). Some aspects of the sequential design of experiments. *Bull. Amer. Math. Soc.*, 58(5):527–535.

Robertson, D. S., Lee, K. M., Lopez-Kolkovska, B. C., and Villar, S. S. (2020). Response-adaptive randomization in clinical trials: from myths to practical considerations. *arXiv preprint arXiv:2005.00564*.

Rosenberger, W. F. and Lachin, J. M. (2015). *Randomization in clinical trials: theory and practice*. John Wiley & Sons, USA.

Rosenberger, W. F., Stallard, N., Ivanova, A., Harper, C. N., and Ricks, M. L. (2001). Optimal adaptive designs for binary response trials. *Biometrics*, 57(3):909–913.

Rosenberger, W. F., Sverdlov, O., and Hu, F. (2012). Adaptive randomization for clinical trials. *Journal of Biopharmaceutical Statistics*, 22(4):719–736.

Rosenberger, W. F., Uschner, D., and Wang, Y. (2019). Randomization: The forgotten component of the randomized clinical trial. *Statistics in Medicine*, 38(1):1–12.

Sabo, R. T. (2014). Adaptive allocation for binary outcomes using decreasingly informative priors. *Journal of Biopharmaceutical Statistics*, 24(3):569–578.

Saville, B. R. and Meurer, W. (2019). Commentary on Ji et al: Sub-optimal illustration of response adaptive randomization. *Clinical Trials*, 16(6):610–612.

Shin, J., Ramdas, A., and Rinaldo, A. (2019a). Are sample means in multi-armed bandits positively or negatively biased? In *Advances in Neural Information Processing Systems*, pages 7102–7111.

Shin, J., Ramdas, A., and Rinaldo, A. (2019b). On the bias, risk and consistency of sample means in multi-armed bandits. *arXiv preprint arXiv:1902.00746*.

Shin, J., Ramdas, A., and Rinaldo, A. (2020). On conditional versus marginal bias in multi-armed bandits. In *International Conference on Machine Learning*, pages 8852–8861. PMLR.

Stallard, N., Hampson, L., Benda, N., Brannath, W., Burnett, T., Friede, T., Kimani, P. K., Koenig, F., Krisam, J., Mozgunov, P., et al. (2020). Efficient adaptive designs for clinical trials of interventions for covid-19. *Statistics in Biopharmaceutical Research*, 12(4):483–497.

Stallard, N. and Rosenberger, W. F. (2002). Exact group-sequential designs for clinical trials with randomized play-the-winner allocation. *Statistics in Medicine*, 21(4):467–480.

Starr, N. and Woodroofe, M. B. (1968). Remarks on a stopping time. *Proceedings of the National Academy of Sciences*, 61(4):1215–1218.

Sutton, R. S. and Barto, A. G. (2018). *Reinforcement learning: An introduction*. MIT press, USA.

Thall, P., Fox, P., and Wathen, J. (2015). Statistical controversies in clinical research: scientific and ethical problems with adaptive randomization in comparative clinical trials. *Annals of Oncology*, 26(8):1621–1628.

Thompson, W. R. (1933). On the likelihood that one unknown probability exceeds another in view of the evidence of two samples. *Biometrika*, 25(3/4):285–294.

U.S. Food and Drug Administration (2022). Center for Drug Evaluation and Research (CDER): Accelerating Rare disease Cures (ARC) Program.

Villar, S. S. (2018). Bandit strategies evaluated in the context of clinical trials in rare life-threatening diseases. *Probability in the Engineering and Informational Sciences*, 32:229–245.

Villar, S. S., Bowden, J., and Wason, J. (2015a). Multi-armed bandit models for the optimal design of clinical trials: benefits and challenges. *Statistical Science*, 30(2):199.

Villar, S. S., Bowden, J., and Wason, J. (2018). Response-adaptive designs for binary responses: How to offer patient benefit while being robust to time trends? *Pharmaceutical Statistics*, 17(2):182–197.

Villar, S. S., Wason, J., and Bowden, J. (2015b). Response-adaptive randomization for multi-arm clinical trials using the forward looking Gittins index rule. *Biometrics*, 71(4):969–978.

Wang, Y.-G. (1991). Sequential allocation in clinical trials. *Communications in Statistics-Theory and Methods*, 20(3):791–805.

Wassmer, G. and Brannath, W. (2016). *Group Sequential and Confirmatory Adaptive Designs in Clinical Trials*. Springer Series in Pharmaceutical Statistics. Springer International Publishing, Switzerland.

Wei, L. J. and Durham, S. (1978). The randomized play-the-winner rule in medical trials. *Journal of the American Statistical Association*, 73(364):840–843.

Whitehead, J. (1986). On the bias of maximum likelihood estimation following a sequential test. *Biometrika*, 73(3):573–581.

Williamson, S. F., Jacko, P., and Jaki, T. (2022). Generalisations of a Bayesian decision-theoretic randomisation procedure and the impact of delayed responses. *Computational Statistics & Data Analysis*, 174:107407.

Williamson, S. F., Jacko, P., Villar, S. S., and Jaki, T. (2017). A Bayesian adaptive design for clinical trials in rare diseases. *Computational Statistics & Data Analysis*, 113:136–153.

Williamson, S. F. and Villar, S. S. (2020). A response-adaptive randomization procedure for multi-armed clinical trials with normally distributed outcomes. *Biometrics*, 76(1):197–209.

Xu, Z., Bandos, A. I., Ma, T., Tang, L., Talisa, V. B., and Chang, C.-C. H. (2022).

Bayesian response adaptive randomization design with a composite endpoint of mortality and morbidity. *arXiv preprint arXiv:2208.08472*.

Zhang, K., Janson, L., and Murphy, S. (2020). Inference for batched bandits. *Advances in Neural Information Processing Systems*, 33:9818–9829.

Zhang, Y., Trippa, L., and Parmigiani, G. (2019). Frequentist operating characteristics of bayesian optimal designs via simulation. *Statistics in Medicine*, 38(21):4026–4039.