Using re-randomisation to increase the recruitment rate in clinical trials

Brennan Chaim Kahan

Queen Mary University of London

submitted in partial fulfillment
of the requirements of the degree of
doctor *of* philosophy

I, Brennan Chaim Kahan, confirm that the research included within this thesis is my own work or that where it has been carried out in collaboration with, or supported by others, that this is duly acknowledged below and my contribution indicated. Previously published material is also acknowledged below.

I attest that I have exercised reasonable care to ensure that the work is original, and does not to the best of my knowledge break any UK law, infringe any third party's copyright or other Intellectual Property Right, or contain any confidential material.

I accept that the College has the right to use plagiarism detection software to check the electronic version of the thesis.

I confirm that this thesis has not been previously submitted for the award of a degree by this or any other university.

Signature:

Date: 28/10/2019

Details of collaboration and publications:

The work in chapter six was published as an article in the *Journal of Clinical Epidemiology* (*Kahan, B.C., et al., Re-randomization increased recruitment and provided similar treatment estimates as parallel designs in trials of febrile neutropenia. J Clin Epidemiol, 2018. 97: p. 14-19*).

The author contributions to this paper are:

Brennan Kahan	Study concept, developed review, performed data extraction, and wrote the manuscript
Tim Morris	Developed review, performed data extraction, and contributed to the manuscript
Erica Harris	Developed review and contributed to the manuscript
Rupert Pearse	Contributed to the manuscript
Richard Hooper	Contributed to the manuscript
Sandra Eldridge	Contributed to the manuscript

Acknowledgements

I am incredibly grateful to many people and groups who have helped me turn this PhD from a vague idea to an actual finished thesis. First and foremost, I would like to thank the National Institute for Health Research for funding this thesis. I would like to thank my primary supervisor Sandra Eldridge, and my secondary supervisors Rupert Pearse and Richard Hooper for their help during the thesis, and for their support and encouragement in developing and persevering with the funding application.

I would like to offer my sincere gratitude to Ian White for generously donating his time to help me make sense of estimands and all sorts of other statistical concepts that eluded my understanding. I would also like to thank Shaun Seaman for helpful discussions around estimands.

Tim Morris' help was invaluable throughout this thesis. This includes (but is not limited to) helping with the work in chapter 6, helping me figure out how to use Latex (and BibTex), and for useful discussions about re-randomisation. He also taught me what an em-dash is, though I haven't decided whether I'm grateful for this or not.

I would like to thank Jan Daru for being a fantastic PhD buddy and making life as a PhD student much more enjoyable.

I would like to thank my family, Evan, Barbara, Tara, Greg, and Lauren, for their love and support throughout the PhD (and before). Finally, I would like to thank Lauren for her support throughout the PhD, and for making the last few years so amazing.

Funding statement

This work was funded by a National Institute for Health Research Doctoral Research Fellowship (DRF-2016-09-005). This thesis presents independent research funded by the National Institute for Health Research (NIHR). The views expressed are those of the authors and not necessarily those of the NHS, the NIHR or the Department of Health.

Abstract

Re-randomisation trials allow patients to be re-enrolled and re-randomised for each new treatment episode they experience. For example, in a trial evaluating treatments for acute sickle cell pain crises, patients could be re-randomised each time they have a new pain crisis. However, uptake of this design has been slow, likely because of uncertainty around its validity. The purpose of this thesis is to evaluate the methodological properties of the re-randomisation design.

Chapter 2 defines a set of treatment estimands that can be used for re-randomisation trials, and chapters 3 and 4 evaluate the use of independence estimators and mixed-effects models for these estimands. I find that independence estimators are generally unbiased, though can be biased for certain estimands in specific situations. Mixed-effects models are generally biased, except under very strong assumptions. In Chapter 5 I compare re-randomisation with cluster, crossover, and parallel group designs. I find that re-randomisation compares favourably with the other designs, though depending on the specific research question (i.e. estimand of interest), other designs may be more appropriate in certain settings. In chapter 6 I evaluate a set of trials of granulocyte colony-stimulating factors for patients with febrile neutropenia which include both parallel group and re-randomisation designs. I found that using re-randomisation led to an increase in recruitment and provided similar results to parallel group trials. In conclusion, the re-randomisation design is a valid design option, and should be used more often.

Contents

List of Figures

List of Tables

1 Introduction

In many clinical settings, patients may require treatment on multiple occasions. For example, patients who experience acute sickle cell pain crises will require treatment to manage pain each time they experience a new crisis; patients who experience severe asthma exacerbations will require treatment to manage symptoms for each new exacerbation; and patients who develop febrile neutropenia as a result of chemotherapy will require treatment for each new round of chemotherapy leading to a neutropenic episode [1, 2, 3]. In these settings, patients are typically given the same treatment for each new episode; for instance, morphine is typically used to treat all hospital admissions for sickle cell pain crises [4]. I refer to these as 'multi-episode' settings, to indicate that some patients may experience multiple treatment episodes over a given period of time.

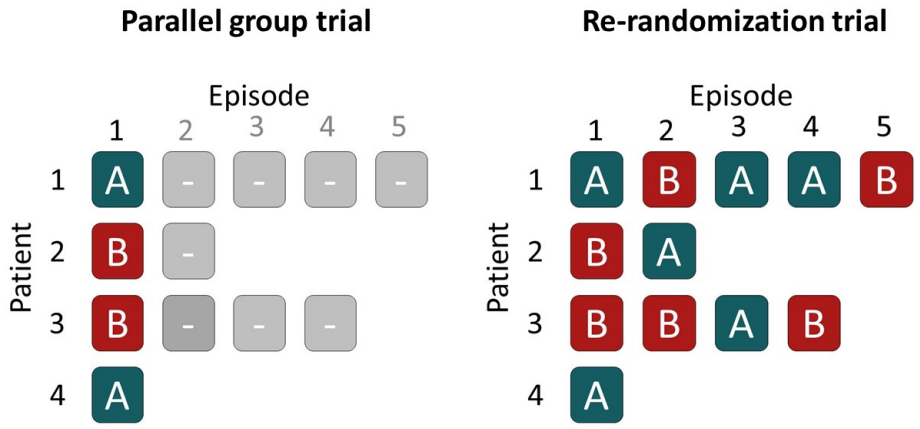
1.1 Randomised trials in multi-episode settings

Randomised controlled trials (RCTs) are considered the gold standard for assessing the effectiveness of healthcare interventions, as randomisation ensures that, on average, treatment groups are balanced for both known and unknown factors, thereby allowing unbiased estimation of the treatment effect [5]. The most common trial design is that of a parallel group design, where patients are enrolled for a single episode only; subsequent treatment episodes are excluded (figure 1.1).

However, allowing only a single episode per patient can have limitations in multi-episode settings. These are described in section 1.1.1 below. Because of these potential limitations, it may be useful to consider alternate trial designs which allow enrolment of multiple treatment episodes. The purpose of this thesis is to evaluate clinical trial designs for use in multi-episode settings. The main design I focus on is the re-randomisation design; this is described in section 1.3 below. However, I discuss alternative design options in chapter 5.

In section 1.1 I describe the potential limitations of allowing trials to enrol only one episode per patient in multi-episode settings, and I provide a motivating example for this work in section 1.2. I give a brief overview of the re-randomisation design in section 1.3, and then list some examples of trials that have used this design in section 1.4. In section 1.5 I provide an overview of the notation that will be used in this thesis, then in section 1.6 I summarise the methodological features of re-randomisation trials,

Figure 1.1: Re-randomisation vs. parallel group trials. This figure depicts the treatment episodes occurring during the trial recruitment period that are eligible for enrollment under a parallel group and re-randomisation design. Grey episodes denote the patient was not eligible, A = allocated to treatment A, B = allocated to treatment B. Reproduced from *Kahan BC, Morris TP, Harris E, Pearse R, Hooper R, Eldridge E. Re-randomization increased recruitment and provided similar treatment estimates as parallel designs in trials of febrile neutropenia. Journal of Clinical Epidemiology 97 (2018) 14-19. DOI: https://doi.org/10.1016/j.jclinepi.2018.02.002* with permission.



and detail what is still unknown. In section 1.7 I give an overview of the research topics that I cover in chapters 2-6 of this thesis.

1.1.1 Limitations of allowing trials to enrol only one episode per patient in multi-episode settings

1.1.1.1 Recruitment

Despite the widespread use of RCTs, recruitment of patients remains a major challenge. Two reviews of trials funded by the National Institute for Health Research (NIHR) Health Technology Assessment (HTA) and the UK Medical Research Council (MRC) found that many trials fail to reach their target sample size [6, 7]. The first review assessed 122 trials from 1994-2002, and found that 69% did not reach their target sample size [6]. The second review assessed 73 trials from 2002-2008, and found that 45% did not meet their target sample size [7]. A review of trials posted on clinicaltrials.gov found that 48,027 patients had taken part in 481 trials that were at risk of being unable to address their primary research question due to poor recruitment [8].

Poor recruitment can have adverse impacts on patient care. It can delay trial results, leading to delays of successful interventions being adopted as part of routine care, or of unsuccessful or harmful treatments being discontinued [9]. Longer recruitment periods or early discontinuation of trials can lead to higher costs and wasted research funds,

resulting in fewer trials being funded overall [9]. Finally, difficulties in recruitment can make conducting trials in rare diseases impractical [9]. For example, a large number of randomised trials in sickle cell disease have been stopped early due to poor recruitment [4, 10]. As a result, guidelines for these diseases may be based on poor evidence from underpowered trials or non-randomised studies, leading to potentially suboptimal care for patients. Poor recruitment is a major barrier to conducting effective RCTs, and has been identified as the primary research priority for leads of UK Clinical Trials Units [11].

Allowing trials to only enrol one episode per patient can be inefficient in multi-episode settings, as it discards a potentially large portion of the available episodes [4, 9]. For instance, there are not very many people who experience sickle cell pain crises, but those who do often experience a large number of episodes [4, 9]. Limiting enrolment to one pain crisis per patient could substantially reduce the pool of eligible episodes, which may make recruitment a much larger challenge than it would be otherwise.

One of the main advantages of trial designs which allow multiple episodes per patient to be enrolled is that they can increase recruitment compared to parallel group designs [12]. This gain in efficiency from re-enrolling patients for subsequent episodes could have large benefits to healthcare systems, by (a) allowing trials to be conducted more quickly, leading to successful treatments being introduced into systems sooner; (b) leading to lower trial costs, allowing more trials to be funded; and (c) allowing trials in some rare diseases to be more easily conducted, resulting in an improved evidence base for the care provided to these patients.

1.1.1.2 Generalisability

A second issue with allowing trials to enrol only one episode per patient is that the results from these trials would likely be used to inform practice for all treatment episodes that occur, even the ones that are excluded from the trial by design. For instance, imagine that a parallel group trial finds that a new intervention is effective for treating pain in acute sickle cell pain crises. When rolled out into clinical practice, this intervention would not be limited to use in a single pain crisis per patient; it would instead be used to treat patients each time they came in with a new pain episode.

The parallel group design estimates the effectiveness and safety of the intervention in a single episode from each patient, and then extrapolates this effect to all subsequent episodes. This approach assumes the effect found in the parallel group design will generalise to the excluded episodes, i.e. that because it was found to be safe and effective when studied in a single episode per patient, it will be equally safe and effective if used for all episodes from each patient. However, this is a very strong assumption; treatments may become less effective or more toxic the more often they are used, or the effectiveness of the treatment may be different in patients who require treatment more often. If the effect differs in subsequent episodes, the extrapolation

from a parallel group design could be misleading, and could result in inappropriate use of the intervention in practice.

Designs which allow multiple episodes per patient to be enrolled do not require extrapolation, as they are able to directly assess whether interventions are safe and effective to use across all treatment episodes. For instance, investigators can evaluate whether there is an increased risk of toxicity due to repeated exposures to the intervention, or whether the intervention becomes less effective on re-use. Because these designs do not require extrapolation it is reasonable to expect they may provide a more realistic picture of how useful the treatments would be in practice.

1.2 Motivating example: the SWIM trial

In the early 2010s I worked on the SWIM (Sickle With Ibuprofen & Morphine) trial (ISRCTN97241637) while I was employed at the MRC Clinical Trials Unit [4] . SWIM was a trial in patients admitted to hospital with an acute sickle cell pain crisis who were treated with patient-controlled analgesia (either morphine or diamorphine). It evaluated whether ibuprofen could reduce the total amount of opioids consumed over a four day period compared to placebo. The rationale was that opioids have negative side effects, and if ibuprofen could reduce the amount of opioids required for effective pain control, this would be of benefit to patients.

The interventions (ibuprofen or placebo) were given for a maximum of four days. The follow-up period for each patient was four weeks from hospital discharge, and it was expected that most patients would be discharged within four days of admission. Therefore, it was anticipated that the follow-up period would be under 5 weeks for most patients enrolled in the trial. SWIM was designed as a parallel group trial, so patients could only be enrolled for a single pain episode. The sample size target was 316 patients.

Recruitment to SWIM was much slower than expected, and so partway through the recruitment period the trial management group convened to discuss options to increase recruitment. One member of the group did an audit of all admissions with a sickle cell pain crisis to one of the SWIM recruiting centres to see whether they could identify reasons for the poor recruitment. One major finding from this audit was that many of the admissions to hospital for pain crises were from the same group of patients. Over a one year period, there were 121 pain crises from 46 patients [9]. The median number of pain crises per patient was 2 (interquartile range 1 to 3; range 1 to 11), and 30/46 patients (65%) were admitted for multiple pain crises [9].

Based on these findings, the trial management group decided to modify the design of the SWIM trial to allow patients to be re-enrolled for each new pain crisis they experienced, provided they had completed the follow-up period from their previous enrolment. It was anticipated that this change could potentially increase the recruitment

rate enough to allow SWIM to reach its recruitment target. However, due to the poor recruitment, the funder decided to terminate the trial early, before the changes to the trial design were implemented.

1.3 Re-randomisation trials

Re-randomisation trials have been proposed as an alternative to the parallel group design to evaluate interventions in multi-episode settings (figure 1.1) [9, 13, 14]. An overview of the re-randomisation design is given in table 1.1.

The re-randomisation design involves re-enrolling and re-randomising patients for each new treatment episode they experience (presuming they continue to meet all eligibility criteria for the trial). In the SWIM example described above, patients could be re-enrolled and re-randomised for each new pain crisis they experience. Importantly, the number of times each patient is enrolled in the trial is not specified in advance, but instead depends on how many treatment episodes they experience during the trial [9]. For instance, in the SWIM example above, some patients may be enrolled for a single pain crisis; others may be enrolled for two or more episodes.

Re-randomisation trials can be used to evaluate interventions that are intended to be used to treat each new episode that occurs (i.e. interventions that are fixed across episodes). The duration of the intervention and duration of the follow-up period for each patient must also be less than the length of the trial recruitment period (otherwise the trial recruitment period would end before the patient completes their first enrolment) [9, 12].

These criteria are met in the SWIM trial described above: (a) some patients may experience more than one pain crisis during the course of the trial; (b) if found to be effective, ibuprofen would likely be used to treat patients each time they were admitted to hospital with a pain crisis, and its use would be fixed across episodes; (c) ibuprofen is a short term intervention (maximum use of four days in the SWIM trial); and (d) the follow-up period in the SWIM trial was also short compared to the overall duration of the trial.

Conversely, re-randomisation trials cannot be used to evaluate interventions which might be modified or changed between episodes (i.e. adaptive interventions), or long-term interventions whose duration is longer than that of the trial recruitment period. It can also not be used if the intended follow-up period for patients is longer than the trial recruitment period. For example, re-randomisation could not be used for: (i) a treatment strategy where the dose of ibuprofen is modified for each new episode based on the patient's response to treatment in their previous episode; (ii) a long-term intervention, such as taking a daily pill on an indefinite basis to reduce the severity of pain crises; or (iii) if investigators wished to evaluate quality of life scores at 5 years post-randomisation, and the duration of recruitment to the trial is expected to take

only 3 years. The re-randomisation design would not be appropriate for use in the above settings, and alternative trial designs would be required.

There are two design requirements for re-randomisation trials (table 1.1) [9, 12]. The first is that patients can only be enrolled and re-randomised after they have completed the follow-up period from their previous enrolment. For example, in a trial where outcomes and other data are only collected up to one year from randomisation, and the patient's involvement in the trial is considered to be finished after the one-year mark, patients can only be re-enrolled after this point. This is to ensure there are no overlapping treatment periods between different randomisations.

The second design requirement is that randomisations for the same patient are performed independently, i.e. the patient's treatment allocation in episode two is not affected by their allocation in episode one. This means that 'patient' is not used in the randomisation process, e.g. patient cannot be used as a stratification factor to force balance in treatment allocations within patients.

1.4 Examples of re-randomisation trials in practice

Several trials which have used the re-randomisation design are shown in table 1.2. I summarise these trials below.

1.4.1 L-arginine in children with sickle cell disease admitted to hospital with vaso-occlusive pain episodes

This trial (NCT01796678) evaluated L-arginine vs. placebo in children with sickle cell disease admitted to hospital with vaso-occlusive pain episodes [1]. This trial allowed patients to be re-enrolled and re-randomised for each new admission to hospital for a vaso-occlusive pain episode. They enrolled 56 episodes from 38 individual children (a 47% increase in the number of episodes enrolled compared to if only one episode per child was allowed). They found that L-arginine led to reductions in the total amount of opioid use (L-arginine 1.9 mg/kg [standard deviation (SD) 2.0] vs. placebo 4.1 [4.1], p=0.02) and lower pain scores at discharge (1.9 [2.4] vs. 3.9 [2.9], p=0.01), though there was little evidence of a difference in length of hospital stay (4.1 days [1.8] vs. 4.8 [2.5], p=0.34). The authors concluded that L-arginine may be a potentially useful adjunct to standard care in this setting and that a larger trial is warranted to confirm or refute these findings.

1.4.2 Azithromycin in children for treatment of acute episodes of asthma-like symptoms

This trial (NCT01233297) evaluated azithromycin vs. placebo to treat children aged 1-3 years who experienced asthma-like symptoms [2]. Children could be re-randomised for each new acute episode of asthma-like symptoms they experienced. They enrolled 158 episodes from 72 children (a 119% increase in the number of episodes enrolled). They found that azithromycin shortened the duration of episodes by an estimated

Table 1.1: Overview of re-randomisation trials. Reproduced from *Kahan BC. Using re-randomization to increase the recruitment rate in clinical trials – an assessment of three clinical areas. Trials. 2016;17:595. doi:10.1186/s13063-016-1736-z* with permission.

Settings requirements for re-randomisation trials	1. Some patients may require treatment on multiple occasions 2. The intervention(s) would be used for each new treatment episode 3. The intervention duration and length of the follow-up period for each treatment episode are less than the overall length of the trial recruitment period
Design requirements for re-randomisation trials	1. Patients are only re-enrolled and re-randomised when they have completed the follow-up period from their previous randomisation 2. Randomisations for the same patient are performed independently
Implementation of re-randomisation trials	1. Patients are enrolled as usual, randomised to a treatment group, and followed-up until all outcomes have been collected 2. If patients experience new treatment episodes and require further treatment, they can be re-enrolled and re-randomised, provided they have completed the follow-up period from their previous randomisation 3. This process is repeated until the target sample size is met

Table 1.2: Examples of trials using re-randomisation design

Trial	**Population**	**Interventions**	**Treatment episode which is re-randomised**
L-arginine in children with sickle cell disease admitted to hospital with vaso-occlusive pain episodes (NCT01796678) [1]	Children with sickle cell disease admitted to hospital for vaso-occlusive pain episodes	L-arginine vs. placebo	New hospitalisation for vaso-occlusive pain episode
Azithromycin in children for treatment of acute episodes of asthma-like symptoms (NCT01233297) [2]	Children aged 1-3 who experience recurrent asthma-like symptoms	Azithromycin vs. placebo	New acute episodes of asthma-like symptoms
Granulocyte colony-stimulating factor in paediatric patient for treatment of febrile neutropenia [3]	Paediatric patients undergoing chemotherapy who experience fever and severe neutropenia	Granulocyte colony-stimulating factor vs. placebo	New episode of fever and severe neutropenia
Pathogen-inactivated platelets in patients with haematologic malignancies and thrombocytopenia (NCT02783313) [15]	Patients with haematologic malignancies and thrombocytopenia	Pathogen-inactivated platelets vs. untreated platelets	New hospital admission requiring transfusion
High-dose influenza vaccine in adults 65 years or older (NCT01427309) [16]	Adults 65 years or older	High-dose influenza vaccine vs. standard dose influenza vaccine	New influenza season
Ambient light for completion of biophysical profiles (NCT02453230) [17]	Women with single gestations scheduled to undergo biophysical profile testing	Ambient light vs. darkened room	New biophysical profile test
Albumin to treat patients admitted to hospital with complications from cirrhosis (ISRCTN14174793) [18]	Patients admitted to hospital with complications of cirrhosis	Albumin vs. standard care	New hospitalisation for complication of cirrhosis

63.3% (95% confidence interval (CI) 56.0 to 69.3; p<0.0001), with no increase in adverse events in the azithromycin group (azithromycin 23% vs. placebo 30%; p=0.30). The authors concluded that azithromycin could be useful in the acute management of exacerbations.

1.4.3 Granulocyte colony-stimulating factor in paediatric patients for treatment of febrile neutropenia

This trial compared granulocyte colony-stimulating factor vs. placebo in paediatric patients undergoing chemotherapy with fever and severe neutropenia [3]. Patients could be re-randomised for each new episode of febrile neutropenia (up to a maximum of 4 episodes). They enrolled 186 episodes from 112 patients (a 66% increase in recruitment compared to limiting each patient to one episode). They found moderate evidence that granulocyte colony-stimulating factor led to a shorter duration of hospital stay (granulocyte colony-stimulating factor median 5 days [range 4 to 8] vs. placebo 7 [range 4 to 10]; p=0.04) and fewer days of antibiotic use (median 5 days [range 3 to 7] vs. placebo 6 [range 4 to 9]; p=0.02), though there was little evidence of a shorter duration of fever (median 2 days [range 1 to 5] vs. 3 days [range 1 to 5]; p=0.30). The authors concluded that use of granulocyte colony-stimulating factor has some benefit in treatment of febrile neutropenia.

1.4.4 Pathogen-inactivated platelets in patients with haematologic malignancies and thrombocytopenia

This trial (NCT02783313) compared the use of pathogen-inactivated platelets vs. untreated platelets for transfusion in patients with haematologic malignancies and thrombocytopenia [15]. They evaluated whether pathogen-inactivated platelets were non-inferior to untreated platelets for the primary outcome of grade 2 bleeding. They used a non-inferiority margin of 12.5 percentage points for the upper limit of the 95% CI for the absolute difference (pathogen-inactivated vs. untreated). Patients could be re-randomised for each new hospital admission requiring transfusion. They enrolled 567 episodes from 469 patients (a 21% increase in recruitment compared to limiting enrolment to one episode per patient). They found in an intention to treat analysis that there was evidence that pathogen-inactivated platelets were non-inferior (pathogen-inactivated 54% vs. untreated 51%, absolute difference 3 percentage points, 95% CI -6 to 11; p for non-inferiority 0.012). However, a per-protocol analysis did not show non-inferiority (52% vs. 44%, absolute difference 8 percentage points, 95% CI -2 to 18; p for non-inferiority 0.19). As such, the authors did not make any strong conclusions around the non-inferiority of pathogen-inactivated platelets.

1.4.5 High-dose influenza vaccine in adults 65 years or older

This trial (NCT01427309) compared high-dose vs. standard-dose influenza vaccine in adults 65 years or older [16]. Patients were allowed to be re-randomised at the start

of each new influenza season. They enrolled 31,989 episodes from 24,344 patients (a 31% increase in recruitment). They found that the high-dose vaccine was effective in reducing the incidence of laboratory-confirmed influenza (high-dose vaccine 1.4% vs. standard-dose vaccine 1.9%, relative efficacy 24.2%, 95% CI 9.7 to 36.5), with a slight decrease in the number of serious adverse events (8.3% vs. 9.0%, relative risk 0.92, 95% CI 0.85 to 0.99). The authors concluded that high-dose vaccine offered better protection against influenza compared to the standard-dose vaccine.

1.4.6 Ambient light for completion of biophysical profiles

This trial (NCT02453230) assessed whether the time to complete a biophysical profile (a pre-natal ultrasound evaluation of the foetus) was shorter using ambient light compared to a darkened room [17]. Patients could be re-randomised each time they returned on a subsequent visit for a biophysical profile. They enrolled 357 biophysical profiles from 224 patients (a 59% increase in recruitment). They found no difference in the time to complete the biophysical profile (ambient light median 6.6 minutes [IQR 2.6 to 13.4] vs. darkened room 6.1 minutes [2.4 to 12.7]; p=0.73). They concluded that ambient light was not useful in decreasing the time needed to complete a biophysical profile.

1.4.7 Albumin to treat patients admitted to hospital with complications from cirrhosis

ATTIRE (Albumin To prevenT Infection in chronic liveR failurE) is a trial (ISRCTN14174793) comparing albumin vs. standard care in patients admitted to hospital with complications from cirrhosis [18]. Patients can be re-randomised for each new admission to hospital provided it has been more than 30 days from the hospital discharge from their previous enrolment. At the time of writing this thesis, ATTIRE has not yet finished recruitment. Its primary outcome is a composite of new infection, renal dysfunction or mortality within the trial treatment period (up to 14 days or hospital discharge if sooner). The trial will aim to recruit 866 episodes.

1.5 Notation

In this section I summarise some of the key notation that will be used in this thesis. A summary of notation is provided in table 1.3. Some of the notation introduced in this section will be expanded on in later chapters, and so there are some pieces of notation in table 1.3 that are explained in later chapters rather than in this section.

Table 1.3: Summary of notation

Notation	**Definition**
Indices	
i	Indexes patient
j	Indexes episode within patient

Continuation of Table 1.3	
Notation	**Definition**
Notation related to number of episodes/patients	
M_i	The number of episodes for which patient i is enrolled in the trial
W_i	A weight for patient i (in this thesis W_i is used to weight patients by the inverse of their number of episodes, $W_i = \frac{1}{M_i}$) (see section 3.1.2)
$M_{T(j)}$	The total number of patients for whom $M_i = j$
M_T	The total number of episodes enrolled in the trial
N_j	The number of patients who are enrolled in the trial for at least j episodes
p	The proportion of patients in the trial who enrolled for two episodes (i.e. $p = \frac{M_{T(2)}}{N_T}$)
$N_{j,\widetilde{Z}=\widetilde{z}}$	The number of patients enrolled at episode j with treatment history $\widetilde{Z} = \widetilde{z}$ (i.e. $N_{2,\widetilde{Z}=(0)}$ denotes the number of patients enrolled at episode 2 who were allocated to control in their first episode, and $N_{2,\widetilde{Z}=(1)}$ is the number allocated intervention in their first episode) (see section 4.3.2)
N_T	The total number of patients enrolled in the trial
Variables	
Y_{ij}	Outcome for patient i during episode j
Z_{ij}	Treatment allocation for patient i during episode j (where 0=control, 1=intervention)
$Z_{i,j-1}$	Treatment allocation in the patient's previous episode (defined as 0 for $j = 1$)
$\widetilde{Z}_{ij}$	A vector of previous treatment allocations for patient i (for example, $\widetilde{Z}_{13}$ would be the vector (Z_{11}, Z_{12})); this is referred to as the 'treatment history'
$X_{ep_{ij}}$	Indicator for episode 2 (i.e. $X_{ep_{ij}} = 1$ for episode 2, and 0 otherwise) (see section 3.1.3)
X_{M_i}	Indicator for patients with $M_i = 2$ (i.e. $X_{M_i} = 1$ if $M_i = 2$, and 0 otherwise) (see section 3.4.2.1)
X_{PL_i}	An unobserved binary patient-level variable (i.e. it is constant across episodes) (see section 3.3.10)

Continuation of Table 1.3	
Notation	**Definition**
$X_{EL_{ij}}$	An unobserved binary episode-level variable (i.e. it can vary across episodes) (see section 3.3.11)
R_{ij}	Denotes whether patient i is enrolled in the trial at episode j (where $R_{ij} = 1$ means the patient was enrolled, $R_{ij} = 0$ means they were not enrolled) (see section 2.4.2.2)
ε_{ij}	A random error term for episode j from patient i
μ_i	Random-intercept for patient i
Parameters	
σ^2_ε	Variance of ε_{ij}
σ^2_μ	Variance of μ_i
σ^2	Total variance of outcome Y_{ij} conditional on treatment allocation Z_{ij}, i.e. $V\left(Y_{ij}\middle\vert Z_{ij}\right) = \sigma^2$
β	Denotes a treatment effect (either the effect of treatment in a model, an estimand, or a potential treatment effect)
α	Represents an intercept in a model
γ	Represents the carry forward effect in a model (i.e. the effect associated with $Z_{i,j-1}$)
δ	Represents the interaction between Z_{ij} and $Z_{i,j-1}$ in a model
ϑ	The intraclass correlation coefficient, which represents the degree of correlation between outcomes from episodes from the same patient
ϑ^*	The expected value of ϑ
Notation related to sampling of episodes/patients (used for estimands)	
I^E	A random variable, where $I^E = i$ with probability $\frac{M_i}{M_T}$
J^E	A random variable, which, conditional on I^E has a uniform distribution on $(1, \ldots, M_{I^E})$
$(IJ)^E$	Represents $I^E J^E$, which represents a randomly selected episode from within the trial, where each episode has an equal probability of being selected ($\frac{1}{M_T}$)
$Y_{(IJ)^E}$	The outcome for a randomly selected episode in the trial
$Z_{(IJ)^E}$	The treatment allocation for a randomly selected episode in the trial
$\widetilde{Z}_{(IJ)^E}$	The treatment history for a randomly selected episode in the trial

Continuation of Table 1.3	
Notation	**Definition**
I^P	A random variable which has a uniform distribution on $(1, \ldots, N_T)$
J^P	A random variable, which, conditional on I^P has a uniform distribution on $(1, \ldots, M_{I^P})$
$(IJ)^P$	Represents $I^P J^P$, which represents a randomly selected episode from a randomly selected patient from within the trial, where each patient has a probability $\frac{1}{N_T}$ of being selected, and each episode from the chosen patient has a probability $\frac{1}{M_{I^P}}$ of being selected
$Y_{(IJ)^P}$	The outcome for a randomly selected episode from a randomly selected patient
$Z_{(IJ)^P}$	The treatment allocation for a randomly selected episode from a randomly selected patient
$\widetilde{Z}_{(IJ)^P}$	The treatment history for a randomly selected episode from a randomly selected patient
End of Table	

I will illustrate this notation using the simple fictitious example shown in table 1.4. Let i index patient, and j index the episode number within the trial. M_i represents the number of episodes for which patient i is enrolled in the trial. In this fictitious example, $M_1 = 1$ because patient 1 was only enrolled for one episode, and $M_5 = 4$ because patient 5 was enrolled for four episodes.

M_T is the total number of episodes enrolled in the trial, and $M_{T(j)}$ represents the total number of patients for whom $M_i = j$ (i.e. the number of patients enrolled for j episodes). In this example, $M_T = 10$ because there are 10 episodes enrolled, $M_{T(1)} = 2$ because two patients are enrolled for only one episode, $M_{T(2)} = 2$ because two patients are enrolled for two episodes, $M_{T(3)} = 0$ because no patients are enrolled for three episodes, and $M_{T(4)} = 1$ because one patient is enrolled for four episodes.

N_T denotes the total number of patients enrolled in the trial, and N_j is the number of patients who are enrolled for at least j episodes. In this example, $N_T = 5$ because there are five patients enrolled in the trial, $N_1 = 5$ because all five patients were enrolled for at least one episode (N_1 is always the same as N_T), $N_2 = 3$ because three patients were enrolled for at least two episodes, and $N_4 = 1$ because only one patient was enrolled for at least four episodes (N_3 is also equal to 1).

Let Y_{ij} denote the outcome for patient i during episode j, and Z_{ij} denote the treatment allocation for patient i during episode j (where 0=control, 1=intervention). $\widetilde{Z}_{ij}$ represents a vector of treatment allocations in the patient's previous episodes

Table 1.4: Simple fictitious re-randomisation trial

Patient ID (i)	**Episode** (j)	**Treatment allocation** (Z_{ij})	**Treatment history** ($\widetilde{Z}_{ij}$)
1	1	0	-
2	1	1	-
3	1	0	-
3	2	1	(0)
4	1	1	-
4	2	0	(1)
5	1	0	-
5	2	1	(0)
5	3	1	(0, 1)
5	4	0	(0, 1, 1)

(referred to as the 'treatment history'). In this fictitious example, $\widetilde{Z}_{54}$ is the vector $(Z_{51}, Z_{52}, Z_{53}) = (0,\ 1,\ 1)$, and $\widetilde{Z}_{32} = (Z_{31}) = (0)$.

Next, I define some notation that will be used for chapter 2 (this notation will become clearer in chapter 2 when I expand on it further). Let I be a random variable, where $I = i$ with a probability that is specified by design (i.e. chosen by the investigators); I represents a randomly selected patient from the trial. Then, let J be a random variable, where $J = j$ with a probability that is specified by design; J represents a randomly selected episode from patient $I = i$. Then, Y_{IJ} represents the outcome for a randomly selected episode from the trial, where each episode has a certain probability of being selected, depending on the specified probabilities for I and J; I discuss different specifications of these probabilities in chapter 2. Similarly, Z_{IJ} and $\widetilde{Z}_{IJ}$ represent the treatment assignment and treatment history for a randomly selected episode.

In this thesis I will use the potential outcomes framework [19, 20]. I will discuss this further in chapter 2, but I briefly define the notation here. Consider a single-episode setting, where patients experience a maximum of one episode. Let Y_i denote the outcome for patient i, and Z_i denote their treatment allocation (0=control, 1=intervention). Then, $Y_i^{(Z_i=1)}$ is the outcome that would occur for patient i if they were allocated to the intervention, and $Y_i^{(Z_i=0)}$ is the outcome that would occur if they were allocated to control. In the multi-episode setting, we need to include the patient's treatment history ($\widetilde{Z}_{ij}$) in the potential outcome definition as this may impact their outcome Y_{ij}. So, $Y_{ij}^{\left(Z_{ij}=0, \widetilde{Z}_{ij}=\widetilde{z}_{ij}\right)}$ would represent patient i's potential outcome at episode j under $Z = 0$ and treatment history $\widetilde{Z}_{ij} = \widetilde{z}_{ij}$, and $Y_{ij}^{\left(Z_{ij}=1, \widetilde{Z}_{ij}=\widetilde{z}_{ij}\right)}$ would represent their potential outcome under $Z = 1$. For clarity, I will drop the subscripts inside the brackets, as these are the same as subscripts on the outside of the brackets; for example, $Y_{ij}^{\left(Z=1, \widetilde{Z}=\widetilde{z}\right)}$ is the same as $Y_{ij}^{\left(Z_{ij}=1, \widetilde{Z}_{ij}=\widetilde{z}_{ij}\right)}$, and $Y_{IJ}^{\left(Z=1,\ \widetilde{Z}=\widetilde{1}\right)}$ is the same

as $Y_{IJ}^{\left(Z_{IJ}=1,\ \widetilde{Z}_{IJ}=\widetilde{1}\right)}$.

Finally, in this thesis I will describe both analysis models and data generating models. Data generating models describe the true underlying process that gives rise to the data, and includes true parameter values; analysis models are the statistical models we use to estimate parameters from the data. I will differentiate the two types of models by including hats on the parameters in the analysis models (to indicate they are estimates from the data), but not on the parameters in the data generating models (as these are the true values). For example, the following is a data generating model, where the outcome is based only on a covariate X_{ij} and a random error term ε_{ij}:

$$Y_{ij} = \alpha + \tau X_{ij} + \varepsilon_{ij}$$

whilst the following is an analysis model which is used to estimate the effect of treatment Z_{ij} (we can see from the data generating model above that the treatment effect is null):

$$Y_{ij} = \hat{\alpha} + \hat{\beta} Z_{ij} + \varepsilon_{ij}$$

1.6 Previous research into methodological properties of re-randomisation trials

In this section I summarise some of the previous research into the properties of the re-randomisation design. I begin by summarising the research around the analysis of re-randomisation trials in sections 1.6.1 and 1.6.2, then summarise several articles which report results of re-analyses of completed re-randomisation trials in section 1.6.3. I then describe research around the sample size calculation in section 1.6.4, then describe previous research into the potential increase in recruitment from re-randomisation in section 1.6.5. In section 1.6.6 I summarise the research which has compared re-randomisation with other trial designs

1.6.1 Analysis of re-randomisation trials

1.6.1.1 Overview of analysis models

The data from a re-randomisation trial will be clustered, with episodes nested within patients. It is therefore natural to assume that episodes from the same patient might be correlated (that is, outcomes from two episodes from the same patient will be more similar to each other than they are to outcomes from episodes from different patients).

There are two broad analysis approaches which have been evaluated for re-randomisation trials: (a) independence estimators; and (b) mixed-effects models with a random intercept for patient (hereafter referred to as 'mixed-effects models'). These

two analysis methods differ primarily in how they deal with the possible correlation between episodes from the same patient.

Independence estimators use a working independence correlation structure to estimate treatment effects [21, 22, 23, 24, 25]; that is, they estimate the treatment effect under the assumption that all episodes are independent, even those from the same patient. They can be used in conjunction with robust standard errors which allow for clustering [21, 26], though the use of robust standard errors for re-randomisation trials has not been formally evaluated in the literature.

The use of working independence correlation structures has been previously suggested for analysing correlated data, under the premise that they can still provide consistent estimates of regression parameters, even if the correlation structure has been misspecified (i.e. if episodes from the same patient are in fact correlated) [21, 22, 23, 24, 25]. However, they are less efficient than methods which correctly specify the correlation structure [21].

For a continuous outcome Y_{ij} the analysis model for an independence estimator can be written as:

$$Y_{ij} = \hat{\alpha} + \hat{\beta} Z_{ij} + \varepsilon_{ij} \tag{1.1}$$

where Y_{ij} is a continuous outcome for episode j from patient i, Z_{ij} is the treatment allocation (0=control, 1=intervention), $\hat{\beta}$ is the estimated treatment effect, and $\varepsilon_{ij} \sim N\left(0, \hat{\sigma}_{\varepsilon}^2\right)$ is a random error term for episode j in patient i (i.e. the off-diagonal elements in the variance-covariance matrix are all set to 0).

In contrast, mixed-effects models directly model the clustering structure by including a random intercept for patients, and allow for correlation between episodes from the same patient. Mixed-effects models with a random-intercept assume an exchangeable correlation structure, where the correlation between any two episodes from the same patient is the same (i.e. all episodes within a patient are equally correlated).

For a continuous outcome Y_{ij} the mixed-effects model can be written as:

$$Y_{ij} = \hat{\alpha} + \hat{\beta} Z_{ij} + \mu_i + \varepsilon_{ij} \tag{1.2}$$

where $\mu_i \sim N\left(0, \hat{\sigma}_{\mu}^2\right)$ is a random-intercept for patient i, and $\varepsilon_{ij} \sim N\left(0, \hat{\sigma}_{\varepsilon}^2\right)$ is a random error term for episode j in patient i (and μ_i and ε_{ij} are assumed to be independent).

1.6.2 Evaluation of analysis models

1.6.2.1 Independence estimators (continuous outcomes)

Before I began work on this thesis, I was involved in a research paper which evaluated independence estimators [9]. We primarily evaluated the performance of the independence estimator when the true data generating model was:

$$Y_{ij} = \alpha + \beta Z_{ij} + \mu_i + \varepsilon_{ij} \tag{1.3}$$

where $\mu_i \sim N\left(0, \sigma_\mu^2\right)$ is a random-intercept for patient i, and $\varepsilon_{ij} \sim N\left(0, \sigma_\varepsilon^2\right)$ is a random error term for episode j in patient i (and μ_i and ε_{ij} are independent). Model 1.3 implies that episodes from the same patient are correlated, with an intraclass correlation coefficient (ICC) of $\frac{\sigma_\mu^2}{\sigma_\mu^2+\sigma_\varepsilon^2}$. It also implies that the treatment effect (β) is constant across all patients and episodes. Note that this model differs to model 1.2 in the use of hats for the parameters (i.e. β vs. $\hat{\beta}$, σ_μ^2 vs. $\hat{\sigma}_\mu^2$), because model 1.2 represents an analysis model, whereas model 1.3 represents a true data generating model.

We initially evaluated independence estimators under the true data generating mechanism in model 1.3, and under the assumption that there is no non-enrolment of episodes; that is, that patients re-enrol for all episodes they experience (i.e. all patients who experience two episodes enrol in the trial for both; there are no patients who enrol for their first episode, but then do not re-enrol for the second episode). Under these assumptions, we showed analytically and using simulation [27] that independence estimators provide unbiased estimates of treatment effect, and the standard error will also be unbiased, implying that confidence intervals and p-values from independence estimators will be valid.

We also evaluated independence estimators under different data generating models (i.e. under models that differ to model 1.3 above). These models were:

- Patients who experience more episodes have different outcomes than patients who experience fewer episodes.
- Patients experience different outcomes in different episodes.
- Outcomes depend upon previous treatment allocation (e.g. the effect of the intervention carries forward into the subsequent episode).

We also evaluated some scenarios with non-enrolment (for example, where some patients who experience two episodes do not re-enrol in the trial for their 2nd episode). These scenarios were:

- The probability of being re-enrolled is the same for all patients;

- The probability of being re-enrolled depends on treatment allocation in the previous episode;
- The probability of being re-enrolled depends on the outcome in the previous episode.

We found that in all scenarios described above the independence estimators provided unbiased estimates of treatment effect and valid standard errors.

1.6.2.2 Independence estimators (binary outcomes)

Dunning and Reeves [13] evaluated an independence estimator for a difference in proportions. They evaluated bias in the estimated standard error and the type I error rate, using both simulation and mathematical derivations. In their simulations, they assumed a true data generating model similar to model 1.3 (except adapted for a binary outcome). In their mathematical derivations they did not assume any particular data generating model. Their simulation study showed that independence estimators provided valid type I error rates. Their mathematical derivations found that under the null hypothesis (i.e. when there is no treatment effect) estimated standard errors were unbiased. However, they showed that under the alternative hypothesis (i.e. when there is a treatment effect) estimated standard errors from an independence estimator may be biased in some instances. The degree of bias depends on the correlation between outcomes, and the event rates in each treatment group. They found that for positive correlation where episodes within a patient are more similar to each other than to episodes from other patients (as would be expected in re-randomisation trials), the bias will be downwards, leading to confidence intervals that are too narrow and p-values that are too small. However, they concluded that in most instances, the bias in the standard errors would be small. For instance, they used the vaccine trial discussed in section 1.4.5 as an example, and found that the bias in the estimated standard errors for this trial would be $\leq 0.2\%$. They did not evaluate bias in the estimated treatment effects.

In our earlier work [9], we also evaluated independence estimators for an odds ratio for binary outcomes. We used simulation to evaluate the type I error rate under a similar data generating model to model 1.3 (adapted for a binary outcome). We found that independence estimators provided valid type I error rates. We did not evaluate bias in estimated treatment effects.

Takada *et al* [28] also evaluated independence estimators for an odds ratio for binary outcomes. They examined this in the context of trials where only patients who do not experience an event (i.e. $Y_{ij} = 0$) were re-randomised. Their motivating example is of a fertility trial, where patients who experience an event (become pregnant, $Y_{ij} = 1$) do not need to undergo further treatment to induce pregnancy. They used a simulation study to evaluate bias and type I error rate under four different data

generating models. All were based on the data generating mechanism in model 1.3 (adapted for a binary outcome), both with and without the random intercept for patient μ_i (the model without the random intercept denotes there is no correlation between episodes from the same patient), and with and without an episode effect (i.e. patients experience different outcomes in different episodes). They found the type I error rate was valid in all scenarios. However, they found that estimated treatment effects were biased in all scenarios (except under the null hypothesis, i.e. no treatment effect). However, it is not immediately clear whether this is true bias, or is because they did not compare the estimated treatment effects to the right parameter. For example, in several scenarios they used a conditional odds ratios (conditional on the random intercept μ_i, or the episode effect) in the data generating model, but estimated a marginal odds ratio. These two quantities will differ due to the non-collapsibility of the odds ratio [29, 30, 31]; however this is not bias, but simply that marginal and conditional odds ratios are estimating different things.

1.6.2.3 Mixed-effects models (continuous outcomes)

In our earlier work, we also evaluated mixed-effects models for continuous outcomes [9]. We found that this method of analysis provided unbiased estimates of treatment effect, valid type I error rates, and higher power than independence estimators for most of the data generating models discussed in section 1.6.2.1. The one exception was when the outcome depended on the treatment allocation from a previous episode. An example of this scenario can be seen in the following model:

$$Y_{ij} = \alpha + \beta Z_{ij} + \gamma Z_{i,j-1} + \mu_i + \varepsilon_{ij}$$

where $Z_{i,j-1}$ is the treatment allocation from the patient's previous episode (and is set to 0 for episode 1). This data generating model represents a situation where the intervention effect carries forward into the subsequent episode. In this scenario, the treatment effect estimate from a mixed-effects model was biased. However, this bias can be negated by correctly modelling the carryover effect in the analysis (for instance, by including the term $Z_{i,j-1}$ as a covariate in the analysis model).

1.6.2.4 Mixed-effects models (binary outcomes)

In our earlier work, we also evaluated mixed-effects logistic models for binary outcomes [9]. We used simulation to evaluate the type I error rate under a similar data generating model to model 1.3 (adapted for a binary outcome). We found that mixed-effects logistic models provided valid type I error rates and had higher power than independence estimators. We did not evaluate bias in the estimated treatment effects.

Nason and Follmann [14] evaluated a mixture model based on a beta distribution in the context of trials where only patients who do not experience an event (i.e. $Y_{ij} = 0$) were re-randomised (i.e. when an event, $Y_{ij} = 1$, precludes future episodes). They used

an example of a trial of anti-HIV microbicide gels where once a patient experiences an event (contracts HIV) it no longer makes sense to use a microbicide gel to prevent HIV. They found using simulation that mixture models provided correct type I error rates in most settings, except when the number of patients was small and the correlation between episodes was large. They did not evaluate bias in the estimated treatment effects.

Takada *et al* [28] used a simulation study to evaluate a mixed-effects logistic regression model with a random-intercept (using the same data generating models as discussed in section 1.6.2.2); they found the type I error rate was valid in all settings. However, they also found that estimated treatment effects were biased in all scenarios (except under the null hypothesis, i.e. no treatment effect). This is partly explained by the non-collapsibility of the odds ratio (i.e. in some scenarios the marginal odds ratio was compared against an odds ratio conditional on an episode effect). However, this analysis method was biased even when the analysis model matched the data generating model, which cannot be explained by the difference between marginal vs. conditional odds ratios. It is therefore not immediately apparent why this analysis method would be biased in this scenario; it may be due to a similar issue of using the incorrect 'true' parameter to evaluate bias (i.e. comparing estimates against the wrong parameter, or 'estimand'; this is described further in chapter 2).

1.6.3 Re-analyses of completed re-randomisation trials

1.6.3.1 High-dose influenza vaccine in adults 65 years or older (NCT01427309)

This trial was discussed earlier in section 1.4.5 [16]. Briefly, this trial compared high-dose vs. standard-dose influenza vaccine in adults 65 years or older. Patients were enrolled over two influenza seasons; participants who were enrolled in the first influenza season could be re-enrolled and re-randomised for the second season.

The aim of their re-analysis [32] was to evaluate whether treatment allocation in the first episode (season one) affected either the outcome or the treatment effect in the second episode (season two); that is they wanted to know whether receipt of the high-dose vaccine in season one either made people less likely to have influenza in season two, or made the high-dose vaccine more (or less) effective in season two. This re-analysis only included participants who were enrolled for both seasons.

They evaluated four different definitions of influenza (laboratory confirmed vs. culture confirmed, any strain vs. vaccine-similar strain). They found no statistical evidence that season one treatment allocation affected either the outcomes or the treatment effect in season two. However, the statistical power for these analyses was low, and consequently the confidence intervals were very wide, and so these results do not rule out the possibility of season one treatment allocation influencing what happens in season two.

1.6.3.2 Web- vs. telephone-based strategies for smoking cessation (NCT01124110)

This trial compared a web-based strategy vs. a telephone-based strategy to reduce smoking amongst heavy equipment operators [33, 34]. The main outcomes were smoking cessation rates at 30 days and at 6 months. This trial was initially designed as a parallel group trial; however, partway through they modified the design to allow patients who did not manage to quit smoking to be re-enrolled and to try again. They initially reported episode one results alone (i.e. the results for the first attempt to quit smoking for each patient) [33]. Their purpose in using re-randomisation was to evaluate whether any patients successfully quite smoking in their second attempt [34]; if so, this may indicate that some smoking cessation interventions could be used again for those who were unsuccessful in previous attempts. Furthermore, the use of a re-randomisation trial design would allow investigators to evaluate the effectiveness of such interventions in subsequent episodes.

Of the 145 patients enrolled in the trial, 128 did not successfully quit smoking in their first attempt; of these 41 (32%) decided to try again, and were re-enrolled and re-randomised. Overall smoking cessation rates in the second episode were 4/41 (9.8%) at 30 days and 5/31 (12.2%) at 6 months. The authors concluded that because many smokers make multiple attempts at quitting, subsequent exposure to interventions after unsuccessful attempts may increase the number of participants who achieve smoking cessation.

Because the numbers were so small, comparison between groups in the second episode was challenging. The quit rate was higher in the web-based group at 30 days (web group 4/23 [17.4%] vs. phone group 0/18 [0%]), though this difference was not statistically significant (p=0.118). Quit rates were similar between groups at 6 months (web group 3/23 [13.0%] vs. phone group 2/18 [11.1%], p=1.00).

1.6.4 Sample size calculations for re-randomisation trials

There have been no formal sample size formulas published for re-randomisation trials. However, in our earlier work [9] we showed that for a continuous outcome, the standard error from an independence estimator in a re-randomisation trial was the same as the standard error from a parallel group design when model 1.3 (constant treatment effect) was the true data generating model. This means that an independence estimator from a re-randomisation trial would have equivalent power to a parallel group trial with the same number of observations. For example, a re-randomisation trial with 200 episodes from 50 patients would have equivalent power to a parallel group trial with 200 individual patients. Using simulation, we found a similar result for binary outcomes (i.e. independence estimators also had the same power to a parallel group design with an equivalent number of observations for binary outcomes).

These findings imply that, under the assumption that model 1.3 is the true underlying data generating model, the same sample size calculation from a parallel

group trial could be used, but instead of recruiting the specified number of individual patients, a re-randomisation trial could recruit the specified number of episodes [9]. For instance, if the sample size calculation for a parallel group design specified 500 patients, then the re-randomisation trial could recruit 500 episodes to have equivalent power.

We used simulation to evaluate what effect departures from model 1.3 as the true data generating model would have on power from an independence estimator with a continuous outcome. We evaluated this for the same scenarios described in section 1.6.2.1. We found that in these scenarios the independence estimator had lower power than a parallel group trial with an equivalent number of observations, though the difference was generally quite small (e.g. less than a 3% difference in most cases).

We also found that mixed-effects models usually had much higher power than a parallel group trial with an equivalent number of observations, for both continuous and binary outcomes. This implies that re-randomisation could enrol a smaller number of observations than a parallel group trial and still have equivalent power; however, sample size calculations for mixed-effects models in re-randomisation trials have not been developed.

1.6.5 Potential increase to recruitment from re-randomisation

Based on the results in section 1.6.4 above, if we assume that the true data generating model follows model 1.3, then for a re-randomisation trial we can use the sample size calculation from a parallel group trial and recruit the specified number of episodes. Because re-randomisation trials allow patients to be re-enrolled this should allow the target sample size to be reached more quickly than under a parallel group design. This gain in recruitment will depend on a number of factors, including how often patients experience new episodes, the length of the follow-up period, and the length of the trial.

We can see from some of the trials described in section 1.4 what effect re-randomisation had on recruitment. For example, the trial of high dose vaccine based their sample size calculation on that of a parallel group trial. Based on this calculation, they required 30,000 episodes. They ended up enrolling 31,989 episodes from 24,344 patients over two influenza seasons. Using a parallel group design, they would have had to recruit for at least one additional season to reach their target sample size, and so using re-randomisation allowed them to cut their recruitment duration from three years to two years (a 33% reduction).

Similarly, the trial of ambient light for biophysical profiles based their sample size calculation on that of a parallel group trial. The calculation required 346 evaluable episodes; they recruited 357 evaluable episodes from 224 patients over approximately 17 months. Assuming a constant rate of recruitment for individual patients of 224/17=13.2 patients/month, they would have required 27 months to recruit 357 patients under a parallel group design. Therefore, re-randomisation allowed them to cut their recruitment

duration from 27 to 17 months (a 37% reduction).

As a third example, the trial of pathogen-inactivated platelets in patients with haematologic malignancies and thrombocytopenia also based their sample size calculation on that of a parallel group trial. They enrolled 567 episodes from 469 patients over approximately a 66 month period. Assuming a constant recruitment rate for individual patients of 469/66=7.1 patients/month, they would have required 79.9 months to recruit 567 patients under a parallel group design. Re-randomisation therefore allowed them to complete recruitment 13.9 months earlier than they would have under a parallel group design (a 17% reduction).

Before I began work on this thesis, I did a modelling study to evaluate how much re-randomisation might increase recruitment in practice [12]. I looked at three clinical areas: (i) acute sickle cell pain crises; (ii) severe asthma exacerbations; and (iii) in vitro fertilisation. Based on different assumptions around the target sample size, length of recruitment for a parallel group design and rate of new episodes, I calculated how much more quickly recruitment could be completed by using re-randomisation. Across all the scenarios considered, I estimated that re-randomisation could reduce the recruitment time by between 19% and 45% (absolute reductions of between 4 and 22 months).

Although this modelling study is entirely reliant on the assumptions that went into it, the results are consistent with the increases in recruitment seen in the re-randomisation trials described in section 1.4. It does appear that using re-randomisation can help trials achieve their target sample size much more quickly than a parallel group design in certain settings.

1.6.6 Comparison between re-randomisation and other designs

Nason and Follmann [14] and Takada *et al* [28] both compared re-randomisation to other trial designs in the context of a trial where experiencing an outcome event precludes future episodes (and so only patients who do not experience an event can be re-enrolled). As discussed in section 1.6.2.4, Nason and Follmann [14] used the example of a trial of anti-HIV microbicide gels , where patients who contract HIV no longer require the use of the gel; and Takada *et al* [28] used the example of a fertility trial, where those who become pregnant no longer need to undergo further treatment to induce pregnancy.

Nason and Follmann [14] compared re-randomisation to both a parallel group design and a crossover design. In the crossover design, patients are switched to the alternate treatment for their second episode (i.e. if they were allocated to the intervention for their first episode, they would be switched to the control for their second episode, and vice versa). Using simulation, they evaluated type I error, precision (i.e. which design had lower variance), and power; they did not evaluate bias. They evaluated the different designs under different data generating mechanisms, which included different degrees of correlation between episodes from the same patients, different

sample sizes, whether there was a trend across episodes (i.e. whether outcomes differed across episodes), and whether treatment allocation in the previous episode affected outcomes in the current episode. They found that both crossover and re-randomisation trials had higher precision (lower variance) and higher power than parallel group trials in most settings. Crossover designs generally had higher power than re-randomisation trials. Type I error rate was generally adequate for all designs in the settings they considered.

Takada *et al* [28] compared re-randomisation with a crossover design and a cluster design (which they call a two-period two-treatment comparison design). In the cluster design, patients acted as the cluster and received the same treatment for all episodes (i.e. if they are allocated to intervention in the first episode they also receive intervention in the second episode, and similarly for control). Using simulation, they evaluated type I error, power, and bias. They evaluated the different designs under different data generating mechanisms, which included different degrees of correlation between episodes from the same patients, different sample sizes, whether there was a trend across episodes (i.e. whether outcomes differed across episodes). They found that all designs generally had valid type I error rates, and that crossover designs usually had the highest power. They found that all designs gave biased estimates of treatment effect across all settings (except when there was no treatment effect). It is not clear why this occurred; as discussed earlier in section 1.6.2.2 in relation to re-randomisation trials, it may be partially explained by the non-collapsibility of the odds ratio. However, they identified bias even when the analysis method was identical to the data generating model (i.e. included all the same parameters), and so non-collapsibility cannot fully explain this bias. As discussed earlier, it may be due to a similar issue of using the incorrect 'true' parameter to evaluate bias (i.e. comparing estimates against the wrong parameter, or 'estimand'; this is described further in chapter 2).

1.7 Thesis focus

The main focus of this thesis is to evaluate the methodological properties of the re-randomisation design. My primary focus is on whether re-randomisation is able to provide unbiased estimates of treatment effect.

In chapter 2 I will define a set of estimands that can be used in multi-episode settings. An estimand is a precise definition of the target we want to estimate (i.e. what is the treatment effect we wish to estimate from our trial?) [35, 36, 37, 38] Defining the estimand of interest is essential, as this allows us to choose a trial design and analysis approach which can provide valid results for the chosen estimand. To my knowledge, estimands have not been previously defined for multi-episode settings, and so it is not clear what exactly is being estimated from a re-randomisation trial.

In chapters 3 and 4 I will evaluate whether analysing re-randomisation trials using either independence estimators or mixed-effects models can provide unbiased estimates

of the estimands defined in chapter 2. Although independence estimators and mixed-effects models have been investigated previously [9, 13, 14, 28], these investigations have mainly assumed a true data generating model where the treatment effect is constant across episodes and patients, which may not always be realistic. They have also not been evaluated against the estimands defined in chapter 2. It is therefore of interest to know whether a re-randomisation design in conjunction with either an independence estimator or a mixed-effects model can provide unbiased estimates for relevant treatment effects.

In chapter 5 I will compare re-randomisation with other designs that could be used in multi-episode settings (parallel group, cluster, or crossover designs). As above, although the different designs have been compared previously [14, 28], this has been primarily done under the assumption of a constant treatment effect, and they have not been compared against the estimands defined in chapter 2.

In chapter 6 I will review a set of trials which evaluated the use of granulocyte colony-stimulating factor in patients with febrile neutropenia, which contained a mixture of both re-randomisation and parallel group trials. I evaluate whether treatment effect estimates were different between re-randomisation and parallel group trials, as well as some design characteristics for the re-randomisation trials (e.g. method of analysis, etc). This chapter has been published as a journal article in the *Journal of Clinical Epidemiology* [39].

I finish with a summary of results in chapter 7, and discuss further work.

In this thesis I will put some restrictions on the setting for simplicity. I consider a setting where:

- Patients experience a maximum of two episodes during the trial;
- The main outcome of interest is continuous;
- The treatment allocation in a given episode does not affect whether future episodes occur (i.e. patients would experience the same number of episodes during the trial period regardless of which treatments they receive during those episodes).

These restrictions are intended to make the scope of this thesis manageable. They are also not uncommon in practice. For instance, the trial of high-dose influenza vaccine and some trials of granulocyte colony-stimulating factor have enrolled a maximum of two episodes per patient [39]. Similarly, most trials described in section 1.4 have at least one continuous outcome. Finally, the trial of ambient light for biophysical profiles and the SWIM trial described in section 1.2 are examples of trials where treatment allocation would not affect occurrence of future episodes. For instance, in the SWIM trial, the interventions under study (ibuprofen, placebo) would not affect

the occurrence of future episodes (e.g. ibuprofen would not prevent future pain crises from occurring), and similarly for the trial of ambient light. To clarify, although I do not consider the setting where treatment allocation precludes the occurrence of future episodes, I do consider the setting where treatment allocation affects whether patients re-enrol in the trial for subsequent episodes (the difference between the two settings being that in the first, future episodes do not exist, whereas in the second they do exist but the patient may not enrol in the trial for them).

2 Estimands in multi-episode settings

In this chapter I will define a set of treatment estimands that can be used in multi-episode settings [35, 36, 37, 38]. An estimand is *"the target of estimation to address the scientific question of interest posed by the trial objective"* [35]; i.e. it is a precise definition of the treatment effect that we wish to estimate in our study. It is important to define the estimand of interest, so that a trial design and analysis method can be chosen which can satisfactorily address the estimand; otherwise there is a risk the study is unable to address its objective.

In section 2.1 I will provide an example which motivates the need for estimands in the multi-episode setting. In section 2.2 I provide some further description of potential outcomes, and in section 2.3 I provide some background on informative cluster sizes (both concepts are used in the estimands defined in this chapter). Then in section 2.4 I provide definitions for a set of estimands that can be used in multi-episode settings.

2.1 Motivating example

If we assume the true data generating mechanism is model 1.3:

$$Y_{ij} = \alpha + \beta Z_{ij} + \mu_i + \varepsilon_{ij}$$

where the treatment effect is constant across all patients and all episodes, then we can see that the overall treatment effect (estimand) should be β. However, it is less clear what the estimand should be under more complicated data generating mechanisms. For example, imagine if in the SWIM trial patients with more severe forms of sickle cell disease were pre-disposed to experience a larger number of pain crises, but were less likely to respond to intervention than patients with less severe forms. Here, the true underlying data generating mechanism would be:

$$Y_{ij} = \begin{cases} \alpha + \beta_1 Z_{ij} + \mu_i + \varepsilon_{ij} \text{ if } M_i = 1 \\ \alpha + \beta_2 Z_{ij} + \mu_i + \varepsilon_{ij} \text{ if } M_i = 2 \end{cases} \tag{2.1}$$

Under this data generating mechanism, the treatment effect is β_1 for patients who experience one episode, and β_2 for patients who experience two episodes. In this scenario, it is not immediately clear what the overall treatment effect should be. For

example, if 50% of patients experience one episode and 50% experience two episodes, then the average treatment effect across patients is $0.5\beta_1 + 0.5\beta_2$, whereas the average treatment effect across episodes is $0.33\beta_1 + 0.67\beta_2$ (this difference comes from weighting the treatment effects by the percentage of patients vs. by the percentage of episodes that they correspond to).

This highlights the need to define treatment estimands for multi-episode settings, so that we can define our precise question of interest (e.g. how effective is the intervention for the typical patient vs. how effective is the intervention for the typical treatment episode), and identify estimators which can be used to answer these questions.

2.2 Potential outcomes

I will use the potential outcomes framework to define estimands in this thesis [19, 20]. Consider a single-episode setting, where patients experience a maximum of one episode. Let Y_i denote the outcome for patient i, and Z_i denote their treatment allocation (0=control, 1=intervention). Then, $Y_i^{(Z_i=1)}$ is the outcome that would occur for patient i if they were allocated to the intervention, and $Y_i^{(Z_i=0)}$ is the outcome that would occur if they were allocated to control. These are called potential outcomes. For ease, I will abbreviate these as $Y_i^{(Z=0)}$ and $Y_i^{(Z=1)}$ (i.e. I am omitting the subscripts inside of the brackets, as these will always match the subscripts outside of the brackets). We can only ever observe one of these potential outcomes, either $Y_i^{(Z=0)}$ or $Y_i^{(Z=1)}$, depending on whether the patient is allocated to control or intervention.

In the multi-episode setting, there will be potential outcomes for each patient i at each different episode j. Furthermore, the potential outcomes at episode j might depend not only on the treatment at episode j, but also on the treatments in previous episodes (i.e. the treatment history $\widetilde{Z}_{ij}$). For example, consider the following data generating model:

$$Y_{ij} = \alpha + \beta Z_{ij} + \gamma Z_{i,j-1} + \mu_i + \varepsilon_{ij} \tag{2.2}$$

where $Z_{i,j-1}$ represents the treatment in a patient's previous episode (and is defined as 0 for the patient's first episode). Under this data generating model, the potential outcome at episode 2, $Y_{i2}^{(Z=z)}$, is not well defined, as its value will differ depending on whether $Z_{i1} = 0$ or 1. We therefore need to incorporate the treatment history $\widetilde{Z}_{ij}$ into the potential outcome definitions, i.e. $Y_{ij}^{\left(Z=z,\widetilde{Z}=\widetilde{z}\right)}$ (note I am dropping the subscripts inside the brackets, as these match the subscripts on the outside of the brackets, i.e. $Y_{ij}^{\left(Z=1,\widetilde{Z}=\widetilde{z}\right)}$ is the same as $Y_{ij}^{\left(Z_{ij}=1,\widetilde{Z}_{ij}=\widetilde{z}_{ij}\right)}$.

In this thesis, I will assume that potential outcomes are deterministic, rather than stochastic. This is most easily explained for a binary outcome, $Y_{ij} = 0$ or $Y_{ij} = 1$. A deterministic potential outcome means that under a certain treatment (and treatment

history), the patient will either experience the event or not experience the event with 100% probability, e.g. $Y_{13}^{\left(Z=1,\widetilde{Z}=(0,1)\right)} = 1$ means that if patient 1 receives the intervention at episode 3 (having also received control and intervention in their first two episodes respectively), they are certain to experience an event. In contrast, if potential outcomes are stochastic then the patient would have a certain probability of experiencing an event (e.g. 90%). The reason I have chosen to use deterministic potential outcomes is for simplicity, as these are easier to explain and it makes little difference to the estimand definitions.

There are two key assumptions I will make about the potential outcomes throughout this thesis. The first assumption is that there will be no interference. This is also sometimes referred to as the stable-unit-treatment value assumption (SUTVA for short). No interference means that a patient's potential outcome does not depend on any other patient's treatment allocation. This assumption could be violated in a trial evaluating a group intervention (for instance, where patients receive the intervention alongside other patients) where a patient's outcome might be affected by their interaction with other patients. Formally, this assumption is that the value $Y_i^{(Z=z)}$ is not affected by $Z_{i'}$ for $i \neq i'$ [19, 20].

The second assumption is consistency. This means that if the potential outcomes for the two treatments $Z = 0$ and $Z = 1$ are $Y_i^{(Z=0)}$ and $Y_i^{(Z=1)}$, then if a patient is allocated to control ($Z = 0$) their observed outcome Y_i is equal to $Y_i^{(Z=0)}$, and similarly for if they were allocated to intervention. Formally, this means that $Y_i = Y_i^{(Z=z)}$ for all values of $Z_i = z_i$. One example where this might not be the case is if there are different versions of the treatment, and outcomes may depend on which version is received. For example, in a trial comparing surgery vs. no surgery, patient outcomes may depend on which surgeons performs their surgery, and so the Y_i might be different for different versions of $Z_i = 1$.

2.3 Informative cluster sizes

The multi-episode setting shares some similarities with the informative cluster size setting [22, 23, 24, 25, 40, 41, 42, 43, 44, 45]. Informative cluster sizes occur in clustered data when the relationship between covariates and outcomes depends on the size of the cluster [23]; that is, if $E(Y|X, M) \neq E(Y|X)$, where Y is the outcome, X the covariate of interest, and M the cluster size [23].

There are two main estimands of interest that have been defined in the informative cluster size literature [23, 24, 25, 41, 44, 45]; the first relates to the effect in a typical unit (irrespective of cluster); the second, to a typical unit from a typical cluster. The difference between these two estimands comes from how each unit is weighted; in the first, each unit is given equal weight; in the second, each cluster is given equal weight (and therefore, each unit from larger clusters is given less weight than each unit from smaller clusters). These estimands are typically defined based on a

sampling framework. For the first estimand, all units are sampled with equal probability (irrespective of which cluster they belong to). The second estimand uses a two-stage sampling procedure; in the first stage, clusters are sampled with equal probability, and in the second stage, the units within the selected cluster are sampled with equal probability.

In a multi-episode setting, the cluster size, M_i, is defined by the number of episodes for which a patient is enrolled in the study. In this setting, cluster size is informative if the association between Z_{ij} and Y_{ij} is different across different values of M_i [23]. For example, under data generating model 1.3, the cluster size will not be informative because the treatment effect is β in all patients, regardless of the number of episodes they experience. Conversely, under data generating model 2.1, the cluster size is informative, as the treatment effect is different in those who experience one episode vs. those who experience two episodes.

One key difference in the multi-episode setting compared to other informative cluster size scenarios is that in the multi-episode setting episodes occur sequentially in time, and outcomes or treatment effects in a patient's current episode may depend on the treatments they received in previous episodes. For example, consider data generating model 2.2:

$$Y_{ij} = \alpha + \beta Z_{ij} + \gamma Z_{i,j-1} + \mu_i + \varepsilon_{ij}$$

Under this data generating model, the effect of the intervention carries forward into the next episode by the amount γ. Conditional on the same treatment history, the treatment effect in episode two is β. However, we may wish to assess the effect of a treatment policy, where patients receive the intervention for all episodes vs. the control for all episodes (i.e. at episode two it would compare treatment sequences $Z = (1, 1)$ vs. $Z = (0, 0)$). In this case, the treatment effect at episode two would be $\beta + \gamma$, which differs from the treatment effect conditional on treatment history. We can therefore define different estimands based on different ways of incorporating a patient's treatment history.

2.4 Treatment effect estimands for multi-episode settings

In this section, I outline a framework that can be used to define treatment effect estimands in the multi-episode setting. This framework is based on the sampling scheme approach used in the informative cluster size setting. The estimands are based on two main components: (a) the sampling scheme, which denotes the probability with which each patient and episode is selected; and (b) the type of treatment history ($\widetilde{Z}_{ij}$) that we are interested in.

My aim is to provide a set of estimands that researchers can choose from to answer their primary question of interest; depending on the specific aims of the trial, different

estimands may be more or less relevant. Researchers could also use this framework to define their own estimands if required.

I discuss the two components in this framework (sampling scheme and treatment history) in sections 2.4.1 and 2.4.2 respectively. In section 2.4.3 I combine these two components to define a set of estimands for the multi-episode setting, which are summarised in table 2.1. Section 2.4.4 introduces episode-specific estimands. In section 2.5 I discuss some of the differences between the estimands, and illustrate these differences in a simple fictitious example.

2.4.1 Sampling scheme: per-episode vs. per-patient estimands

In this section I discuss two different approaches to defining the sampling scheme, which lead to two different types of estimands: the per-episode estimand, and the per-patient estimand. For the moment, I will assume that a patient's treatment history $\widetilde{Z}_{ij}$ does not affect either the potential outcomes or treatment effect in their current episode, i.e. $Y_{ij}^{\left(Z=0,\widetilde{Z}=\widetilde{z}\right)}$ is the same for all values of $\widetilde{Z}_{ij}$, and similarly for $Y_{ij}^{\left(Z=1,\widetilde{Z}=\widetilde{z}\right)} - Y_{ij}^{\left(Z=0,\widetilde{Z}=\widetilde{z}\right)}$. Therefore, in this section I will omit $\widetilde{Z}_{ij}$ from the estimand definitions; however I will relax this assumption in section 2.4.2.

2.4.1.1 Per-episode estimands

The per-episode estimand (denoted by β_E) gives the average effect across all episodes (i.e. "What is the effect of intervention in a typical episode?"). For instance, consider data generating model 2.1:

$$Y_{ij} = \begin{cases} \alpha + \beta_1 Z_{ij} + \mu_i + \varepsilon_{ij} \text{ if } M_i = 1 \\ \alpha + \beta_2 Z_{ij} + \mu_i + \varepsilon_{ij} \text{ if } M_i = 2 \end{cases}$$

In this scenario, the per-episode estimand would weight the treatment effects β_1 and β_2 according to the number of episodes they correspond to.

I define the per-episode estimand as the expected difference in potential outcomes between intervention vs. control for a randomly selected episode. It is defined based on a sampling scheme where each episode in the trial has an equal probability of being selected; because there are M_T episodes, the probability of selection for each episode is $P\left(I = i,\ J = j\right) = \frac{1}{M_T}$. The estimand is defined as:

$$\beta_E = \ E\left(Y_{(IJ)^E}^{(Z=1)} - Y_{(IJ)^E}^{(Z=0)}\right)$$

where $(IJ)^E$ denotes a randomly selected episode (selected with probability $\frac{1}{M_T}$). I use E to denote that the sampling scheme is based on a per-episode estimand, and $Y_{(IJ)^E}^{(Z=1)} = Y_{(IJ)^E}^{\left(Z_{(IJ)^E}=1\right)}$ (as I am omitting subscripts within the brackets). Formally,

$(IJ)^E$ represents $\left(I^E, J^E\right)$, where I^E and J^E are random variables; $I^E = i$ with probability $\frac{M_i}{M_T}$, and, conditional on I^E, J^E is uniformly distributed on $(1, \ldots, M_{I^E})$ (where M_{I^E} represents the total number of episodes, M_i, for which patient I^E was enrolled in the trial).

2.4.1.2 Per-patient estimands

The per-patient estimand (denoted by β_P) gives the average effect across patients (i.e. "What is the effect of intervention in a typical patient?"). Consider again data generating model 2.1; the per-patient estimand would weight the treatment effects β_1 and β_2 according to the number of patients they correspond to (rather than according to the number of episodes, as in the per-episode estimand).

I define the per-patient estimand as the expected difference in potential outcomes between intervention vs. control for a randomly selected episode from a randomly selected patient. It is based on a two-stage sampling scheme, where a patient is randomly selected in the first stage (each with equal probability), then an episode from within that patient is randomly selected in the second stage (each with equal probability). Because there are N_T patients in the population of interest, the probability of selection for each patient is $\frac{1}{N_T}$, and because each patient has M_i episodes, the probability of selection for each episode, given that patient i has been selected, is $\frac{1}{M_i}$. Therefore, the overall probability of selection for each episode is $P\left(I = i,\ J = j\right) = \left(\frac{1}{N_T}\frac{1}{M_i}\right)$. The estimand is defined as:

$$\beta_P = \ E\left(Y_{(IJ)^P}^{(Z=1)} - Y_{(IJ)^P}^{(Z=0)}\right)$$

where $(IJ)^P$ denotes randomly selected episode from a randomly selected patient (I use P to denote that the sampling scheme is based on a per-patient estimand). Formally, $(IJ)^P$ represents $I^P J^P$, where I^P and J^P are random variables; I^P has a uniform distribution on $(1, \ldots, N_T)$, and, conditional on I^P, J^P is uniformly distributed on $(1, \ldots, M_{I^P})$ (where M_{I^P} represents the value of M_i for the randomly selected patient I^P).

2.4.2 Treatment history: added-benefit vs. policy-benefit estimands

In this section, I allow for the fact that a patient's treatment history may influence the potential outcomes or treatment effect in their current episode (for instance, that the value of $Y_{ij}^{\left(Z=0,\widetilde{Z}=\widetilde{z}\right)}$ may be different for different values of $\widetilde{Z}_{ij}$). I discuss two approaches for incorporating treatment history into the estimand definition: the policy-benefit estimand, and the added-benefit estimand. For the moment, I do not distinguish between per-episode and per-patient sampling schemes, and instead use Y_{IJ} as a placeholder for one of these two sampling schemes.

2.4.2.1 Policy-benefit estimands

The policy-benefit estimand (denoted by β^{PB}) gives the effect of a treatment policy where patients either receive intervention for all episodes, or control for all episodes (i.e. "how much benefit is there to always treating vs. never treating?"). For instance, consider data generating model 2.2:

$$Y_{ij} = \alpha + \beta Z_{ij} + \gamma Z_{i,j-1} + \mu_i + \varepsilon_{ij}$$

In this scenario, under a treatment policy (intervention for all episodes vs. control for all episodes) the difference in potential outcomes at episode 2 involves the term γ (as this term represents benefit carried forward from receiving the intervention in episode 1), and the policy-benefit estimand therefore would include this term.

I define the policy-benefit estimand as the expected difference in potential outcomes for a randomly selected episode (based either on a per-episode or per-patient basis), which has been allocated intervention for the current and all previous episodes vs. control for the current and all previous episodes. The estimand can be written as:

$$\beta^{PB} = E\left(Y_{IJ}^{\left(Z=1,\widetilde{Z}=\widetilde{1}\right)} - Y_{IJ}^{\left(Z=0,\widetilde{Z}=\widetilde{0}\right)}\right)$$

where $\widetilde{Z} = \widetilde{1}$ means the patient has received the intervention for all previous episodes in the trial, and $\widetilde{Z} = \widetilde{0}$ means they received the control for all previous episodes.

2.4.2.2 Added-benefit estimands

The added-benefit estimand (denoted by β^{AB}) gives the additional effect of receiving the intervention in the current episode (i.e. "what is the benefit to taking the intervention in this episode, over and above the benefit from previous episodes?"). Consider again data generating model 2.2; the added-benefit estimand would omit the term γ, as this represents carried over benefit from previous episodes, rather than any new benefit from the intervention in the current episode.

One problem in defining this estimand is that the treatment effect may depend on treatment history, and so the additional benefit of receiving intervention might be different for different histories. For instance, if the intervention became more or less effective the more often it is used, then its benefit in a particular episode will depend on the number of times a patient has received it previously. Let $\beta_{ij}^{AB\left(\widetilde{Z}=\widetilde{z}\right)} = Y_{ij}^{\left(Z=1,\widetilde{Z}=\widetilde{z}\right)} - Y_{ij}^{\left(Z=0,\widetilde{Z}=\widetilde{z}\right)}$ represent the difference in potential outcomes (intervention vs. control) for patient i at episode j under treatment history $\widetilde{Z} = \widetilde{z}$; this can be thought of as the potential treatment effect for patient i at episode j under a specific treatment history (because I view each of the potential outcomes as deterministic,

this implies that $\beta_{ij}^{AB\left(\widetilde{Z}=\widetilde{z}\right)}$ is also deterministic for a specific value of $\widetilde{Z}=\widetilde{z}$). There will be 2^{j-1} different potential treatment histories at episode j, and therefore this many possible values of $\beta_{ij}^{AB\left(\widetilde{Z}=\widetilde{z}\right)}$ (for patients with $M_i \geq j$). Therefore, in order to define an added-benefit estimand, we must first define the distribution of $\widetilde{Z}$ that we are interested in.

A simple approach might be to define a single treatment history. For instance, we could use $\widetilde{Z}_{ij}=\widetilde{0}$, which indicates the patient received control in all previous episodes. We could then define the estimand as $E\left(Y_{IJ}^{\left(Z=1,\widetilde{Z}=\widetilde{0}\right)}-Y_{IJ}^{\left(Z=0,\widetilde{Z}=\widetilde{0}\right)}\right)$, which represents the added benefit of receiving intervention in a patient's current episode, given they received control in all previous episodes. Alternatively, we could use $\widetilde{Z}_{ij}=\widetilde{1}$, which would lead to an estimand that represents the added benefit of receiving intervention in an episode, given the patient had received intervention in all previous episodes. One drawback to this approach is that the treatment effect based on the chosen treatment history may not be representative of other treatment histories. This is not a problem if we are only interested in the single treatment history specified, but may be a problem otherwise.

My solution to this issue is to define the added-benefit estimand as a weighted average across all possible treatment histories, with each value of $\beta_{ij}^{AB\left(\widetilde{Z}=\widetilde{z}\right)}$ weighted according to the probability of treatment history $\widetilde{Z}=\widetilde{z}$ being observed at episode j for patient i. This probability, denoted as $P\left(\widetilde{Z}_{ij}=\widetilde{z}_{ij}\right)$, depends on two factors; the allocation probabilities used in the study (e.g. 1:1 allocation ratio vs. 2:1 allocation ratio), and whether the patient would be enrolled in the trial at episode j under treatment history $\widetilde{Z}_{ij}=\widetilde{z}_{ij}$ (for instance, patients who experience two episodes may decide to enrol in the trial for their second episode under one treatment history, but not under a different history). Note that this probability is not known in practice, as although we know the allocation ratio, we will not know whether patients would be enrolled in the trial under different treatment histories.

Let R_{ij} denote whether patient i is enrolled in the trial at episode j (where $R_{ij}=1$ means the patient was enrolled, and $R_{ij}=0$ means they were not enrolled; note that $R_{ij}=0$ if $M_i<j$), and let $R_{ij}^{\left(\widetilde{Z}=\widetilde{z}\right)}$ denote the patient's potential enrolment status at episode j under treatment history $\widetilde{Z}_{ij}=\widetilde{z}_{ij}$. Then, $P\left(\widetilde{Z}_{ij}=\widetilde{z}_{ij}\right)=0$ if $R_{ij}^{\left(\widetilde{Z}=\widetilde{z}\right)}=0$; that is, there is no chance that $\widetilde{Z}_{ij}=\widetilde{z}_{ij}$ for patient i at episode j if that patient would not be enrolled in the trial under that treatment history. Under a 1:1 allocation ratio, all treatment histories $\widetilde{Z}_{ij}=\widetilde{z}_{ij}$ for which $R_{ij}^{\left(\widetilde{Z}=\widetilde{z}\right)}=1$ are equally like to be observed; hence, a simple average over the $\beta_{ij}^{AB\left(\widetilde{Z}=\widetilde{z}\right)}$ for which $R_{ij}^{\left(\widetilde{Z}=\widetilde{z}\right)}=1$ could be taken. Note that $\sum_{\widetilde{Z}_{ij}} P\left(\widetilde{Z}_{ij}=\widetilde{z}_{ij}\right)=1$.

Then, I define the added-benefit estimand as:

$$\beta^{AB} = E\left(Y_{IJ}^{\left(Z=1,\widetilde{Z}\right)} - Y_{IJ}^{\left(Z=0,\widetilde{Z}\right)}\right)$$

where the expectation is over both the IJ and the distribution of $\widetilde{Z}$ (which was defined above, i.e. the expectation is taken over the different treatment histories according to their probability of being observed for each patient, $P\left(\widetilde{Z}_{ij} = \widetilde{z}_{ij}\right)$). Further detail on how this estimand can be related to the potential treatment effects is given in section 2.4.3.1.

There are several important implications that follow on from this definition. The first is that the estimand depends on the distribution of $\widetilde{Z}$, and so changes to this distribution may lead to different values of the estimand. For example, in a re-randomisation trial, changing the allocation ratio from 1:1 to 2:1 will change the distribution of the treatment history. This implication is inherent to any definition of the estimand which averages over different treatment histories. Of note, if the treatment effect is not affected by treatment history (i.e. $Y_{ij}^{\left(Z=1,\widetilde{Z}=\widetilde{z}\right)} - Y_{ij}^{\left(Z=0,\widetilde{Z}=\widetilde{z}\right)}$ is the same for all values of $\widetilde{Z}_{ij} = \widetilde{z}_{ij}$), then the added-benefit estimand does not depend on the distribution of $\widetilde{Z}$ (this is the case even if $Y_{ij}^{\left(Z=0,\widetilde{Z}=\widetilde{z}\right)}$ is affected by $\widetilde{Z}_{ij}$, e.g. if the treatment effect carries forward as in data generating mechanism 2.2). Then, the manner in which the treatment histories are weighted does not matter, as all definitions will be equivalent.

Another implication of this definition is that the estimand excludes values of $\beta_{ij}^{AB\left(\widetilde{Z}=\widetilde{z}\right)}$ for which $R_{ij}^{\left(\widetilde{Z}=\widetilde{z}\right)} = 0$. That is, it excludes potential treatment effects for episodes which correspond to treatment histories that would never be observed in the trial for that particular patient (for instance because under that treatment history they would not have re-enrolled in the trial at episode j). One benefit of defining the estimand to exclude episodes that would not be enrolled in the trial is that this definition is compatible with the setting where treatment allocation may influence the occurrence of subsequent episodes (for instance, in a trial evaluating an intervention to induce pregnancy in couples with difficulty conceiving), as this definition excludes episodes which would not have occurred under specific treatment histories. Although I do not consider this setting in this thesis, I feel it is useful for definitions to apply to more complicated settings when possible.

2.4.3 Main estimands of interest

In this section, I combine the two components (sampling scheme, treatment history) discussed in the previous sections to define a set of estimands which can be used in multi-episode settings. A summary of these estimands is provided in table 2.1.

2.4.3.1 Per-episode added-benefit estimand

The per-episode added-benefit estimand (β_E^{AB}) is defined as:

$$\beta_E^{AB} = E\left(Y_{(IJ)^E}^{\left(Z=1,\widetilde{Z}\right)} - Y_{(IJ)^E}^{\left(Z=0,\widetilde{Z}\right)}\right)$$

where $Y_{(IJ)^E}^{(Z=1,\widetilde{Z})}$ represents $Y_{(IJ)^E}^{(Z_{(IJ)^E}=1,\widetilde{Z}_{(IJ)^E})}$ (and similarly for $Y_{(IJ)^E}^{(Z=0,\widetilde{Z})}$). This expectation is over both $(IJ)^E$ and $\widetilde{Z}_{(IJ)^E}$. As described in section 2.4.2.2, $\widetilde{Z}_{(IJ)^E}$ is a random variable which takes on the different treatment histories $\widetilde{Z}_{ij}$ according to the probability of that history being observed in the trial for patient i at episode j (denoted by $P\left(\widetilde{Z}_{ij} = \widetilde{z}_{ij}\right)$).

The estimand can be related to the potential treatment effects for each i, j (the $\beta_{ij}^{AB\left(\widetilde{Z}=\widetilde{z}\right)}$) through the expression:

$$\beta_E^{AB} = \frac{1}{M_T}\sum_{ij}\bar{\beta}_{ij}^{AB}$$

where $\bar{\beta}_{ij}^{AB}$ is calculated as a weighted average of the $\beta_{ij}^{AB\left(\widetilde{Z}=\widetilde{z}\right)}$, with weights equal to the probability of the treatment history being observed:

$$\bar{\beta}_{ij}^{AB} = \sum_{\widetilde{Z}_{ij}} \beta_{ij}^{AB\left(\widetilde{Z}=\widetilde{z}\right)} P\left(\widetilde{Z}_{ij} = \widetilde{z}_{ij}\right)$$

where the summation $\sum_{\widetilde{Z}_{ij}}(\bullet)$ is taken across all possible values of $\widetilde{Z}_{ij}$ at episode j.

2.4.3.2 Per-episode policy-benefit estimand

The per-episode policy-benefit estimand (β_E^{PB}) is defined as:

$$\beta_E^{PB} = E\left(Y_{(IJ)^E}^{\left(Z=1,\ \widetilde{Z}=\widetilde{1}\right)} - Y_{(IJ)^E}^{\left(Z=0,\ \widetilde{Z}=\widetilde{0}\right)}\right)$$

Let $\beta_{ij}^{PB} = Y_{ij}^{\left(Z=1,\widetilde{Z}=\widetilde{1}\right)} - Y_{ij}^{\left(Z=0,\widetilde{Z}=0\right)}$ represent the difference in potential outcomes for patient i at episode j under a policy of all intervention vs. all control. Because I view potential outcomes as deterministic, this implies that β_{ij}^{PB} is also deterministic for a particular i, j. Then, the estimand can be related to the potential treatment effects for each i, j through the expression:

$$\beta_E^{PB} = \frac{1}{M_T}\sum_{ij}\beta_{ij}^{PB}$$

2.4.3.3 Per-patient added-benefit estimand

The per-patient added-benefit estimand (β_P^{AB}) is defined as:

$$\beta_P^{AB} = E\left(Y_{(IJ)^P}^{\left(Z=1,\widetilde{Z}\right)} - Y_{(IJ)^P}^{\left(Z=0,\widetilde{Z}\right)}\right)$$

This estimand can be related to the potential treatment effects for each i, j through the expression:

$$\beta_P^{AB} = \frac{1}{N_T}\sum_{ij}\frac{1}{M_i}\bar{\beta}_{ij}^{AB}$$

where $\bar{\beta}_{ij}^{AB}$ was previously defined in section 2.4.3.1.

2.4.3.4 Per-patient policy-benefit estimand

The per-patient policy-benefit estimand (β_P^{PB}) is defined as:

$$\beta_P^{PB} = E\left(Y_{(IJ)^P}^{\left(Z=1,\ \widetilde{Z}=\widetilde{1}\right)} - Y_{(IJ)^P}^{\left(Z=0,\ \widetilde{Z}=\widetilde{0}\right)}\right)$$

This estimand can be related to the potential treatment effects for each i, j through the expression:

$$\beta_P^{PB} = \frac{1}{N_T}\sum_{ij}\frac{1}{M_i}\beta_{ij}^{PB}$$

2.4.4 Episode-specific estimands

I now define episode-specific estimands (denoted by β_j for episode j). These estimands represent the treatment effect at episode j (note that β_j only includes patients for whom $M_i \geq j$). For example, β_2 represents the treatment effect at episode 2, in patients with $M_i \geq 2$. I note that the per-episode vs. per-patient distinction does not apply here, as patient's can be enrolled only once at each episode. These estimands could be of interest if we wanted to ascertain whether the treatment effect changes across episodes, or whether the treatment becomes less effective the more often it is used.

Under the added-benefit framework, the episode-specific estimand is:

$$\beta_j^{AB} = E\left(Y_{I^{ES}{}_j}^{\left(Z=1,\widetilde{Z}\right)} - Y_{I^{ES}{}_j}^{\left(Z=0,\widetilde{Z}\right)}\right)$$

Table 2.1: Summary of estimands for multi-episode settings

Estimand	**Definition**	**Description**
Per-episode added-benefit	$\beta_E^{AB} = E\left(Y_{(IJ)^E}^{\left(Z=1,\widetilde{Z}\right)} - Y_{(IJ)^E}^{\left(Z=0,\widetilde{Z}\right)}\right)$	• Provides the additional effect of receiving the intervention in the current episode, over and above the benefit of receiving the intervention in previous episodes • Provides an average effect across episodes
Per-episode policy-benefit	$\beta_E^{PB} = E\left(Y_{(IJ)^E}^{\left(Z=1,\ \widetilde{Z}=\widetilde{1}\right)} - Y_{(IJ)^E}^{\left(Z=0,\ \widetilde{Z}=\widetilde{0}\right)}\right)$	• Provides the effect of a treatment policy where patients either receive intervention vs. control for all episodes • Provides an average effect across episodes
Per-patient added-benefit	$\beta_P^{AB} = E\left(Y_{(IJ)^P}^{\left(Z=1,\widetilde{Z}\right)} - Y_{(IJ)^P}^{\left(Z=0,\widetilde{Z}\right)}\right)$	• Provides the additional effect of receiving the intervention in the current episode, over and above the benefit of receiving the intervention in previous episodes • Provides an average effect across patients
Per-patient policy-benefit	$\beta_P^{PB} = E\left(Y_{(IJ)^P}^{\left(Z=1,\ \widetilde{Z}=\widetilde{1}\right)} - Y_{(IJ)^P}^{\left(Z=0,\ \widetilde{Z}=\widetilde{0}\right)}\right)$	• Provides the effect of a treatment policy where patients either receive intervention vs. control for all episodes • Provides an average effect across patients

where I^{ES} is a random variable which represents a randomly selected patient (with equal probability) at episode j (from the subset of patients with $M_i \geq j$). I use the ES superscript to denote episode-specific.

This estimand can be related to the potential treatment effects by:

$$\beta_j^{AB} = \frac{1}{N_j} \sum_{i:\ M_i \geq j} \bar{\beta}_{ij}^{AB}$$

Under the policy-benefit framework, the estimand can be defined as:

$$\beta_j^{PB} = E\left(Y_{I^{ES}j}^{\left(Z=1,\widetilde{Z}=\widetilde{1}\right)} - Y_{I^{ES}j}^{\left(Z=0,\widetilde{Z}=\widetilde{0}\right)}\right)$$

and can be related to the potential treatment effects by:

$$\beta_j^{PB} = \frac{1}{N_j} \sum_{i:\ M_i \geq j} \beta_{ij}^{PB}$$

I note that the estimands above relate to the subset of patients with $M_i \geq j$, and so may not be useful for directly comparing the effect of the intervention across different episodes. For example, if we want to know whether the intervention is more effective the first time it is used vs. the second time, a comparison between β_1 vs. β_2 does not necessarily tell us this, because β_1 applies to patients for whom $M_i = 1$, whereas β_2 does not. If the treatment effect is different in those for whom $M_i = 1$ compared to $M_i > 1$, then β_1 and β_2 may differ, even if the effect is the same both the first and second time the intervention is used.

An alternate way to define episode-specific estimands is to restrict the subset of patients for each β_j to those where $M_i \geq c$, where c is the number of episodes we are interested in. For example, if we want to know whether the treatment effect is the same the first three times the intervention is used, then we could restrict to the subset of patients for whom $M_i \geq 3$ and compare $\beta_{1(M_i \geq 3)}$ vs. $\beta_{2(M_i \geq 3)}$ vs. $\beta_{3(M_i \geq 3)}$.

Although I have included the episode-specific estimands here for completeness, I do not consider them further in this thesis, as my primary focus is on estimands which provide an average treatment effect across all patients and episodes, rather than a subset of patients or episodes.

2.5 Comparison between estimands

I now discuss some of the differences between the per-episode vs. per-patient, and the added- vs. policy-benefit estimands, and then explore these differences in a simple fictitious example.

2.5.1 *Comparison between per-episode vs. per-patient estimands*

Following on from the literature on informative cluster sizes [23, 24, 25, 44, 45], the per-episode and per-patient estimands should coincide unless cluster size is informative. A simple way to evaluate whether the cluster size is informative under a particular data generating model is to take the mean of the potential treatment effects (either $\bar{\beta}_{ij}^{AB}$ or β_{ij}^{PB}) across episodes for each patient; then if the mean potential treatment effect for each patient differs according to M_i, the cluster size is informative (as this indicates the association between Z_{ij} and Y_{ij} is different across different values of M_i).

For instance, consider data generating model 1.3, where the treatment effect is constant across episodes and patients. Then, the mean potential treatment effect is β for patients with both $M_i = 1$ and $M_i = 2$; therefore, $\beta_P = \beta_E$.

However, consider data generating model 2.1, where the treatment effect is different between patients where $M_i = 1$ vs. $M_i = 2$. Here, the mean potential treatment effect is β_1 for patients where $M_i = 1$, and β_2 for patients where $M_i = 2$; therefore, $\beta_P \neq \beta_E$.

Some further examples of scenarios where the estimands differ are shown in section 2.5.3.

2.5.2 *Comparison between added-benefit vs. policy-benefit estimands*

The added-benefit and policy-benefit estimands will coincide if $\bar{\beta}_{ij}^{AB} = \beta_{ij}^{PB}$. This will occur when the treatment history $\widetilde{Z}_{ij}$ does not affect either $Y_{ij}^{\left(Z=0,\widetilde{Z}=\widetilde{z}\right)}$ or $Y_{ij}^{\left(Z=1,\widetilde{Z}=\widetilde{z}\right)} - Y_{ij}^{\left(Z=0,\widetilde{Z}=\widetilde{z}\right)}$ (i.e. when $Y_{ij}^{\left(Z=0,\widetilde{Z}=\widetilde{z}\right)}$ takes the same value for all values of $\widetilde{Z}_{ij}$, and similarly for $Y_{ij}^{\left(Z=1,\widetilde{Z}=\widetilde{z}\right)} - Y_{ij}^{\left(Z=0,\widetilde{Z}=\widetilde{z}\right)}$). When this is not the case then the added-benefit and policy-benefit estimands will usually differ.

2.5.3 *Differences between estimands in a fictitious example*

I now illustrate the similarities and differences between the four estimands of interest using a simple fictitious example. Consider a trial where patients experience a maximum of two episodes, with a 1:1 allocation ratio. Further, assume there is no non-enrolment (i.e. patients who experience two episodes enrol in the trial for both episodes, that is, $R_{i2} = 1$ for all patients with $M_i = 2$). Let p denote the proportion of patients in the trial who enrolled for two episodes (i.e. $p = \frac{M_{T(2)}}{N_T}$).

I will evaluate the estimands under a variety of different generating mechanism using the formulas described in section 2.4.3. As a reminder, I repeat these formulas here.

For the per-episode added-benefit estimand, I use:

$$\beta_E^{AB} = \frac{1}{M_T} \sum_{ij} \bar{\beta}_{ij}^{AB}$$

Where:

$$\bar{\beta}_{ij}^{AB} = \sum_{\widetilde{Z}_{ij}} \beta_{ij}^{AB\left(\widetilde{Z}=\widetilde{z}\right)} P\left(\widetilde{Z}_{ij} = \widetilde{z}_{ij}\right)$$

For the per-episode policy-benefit estimand, I use:

$$\beta_E^{PB} = \frac{1}{M_T} \sum_{ij} \beta_{ij}^{PB}$$

For the per-patient added-benefit estimand, I use:

$$\beta_P^{AB} = \frac{1}{N_T} \sum_{ij} \frac{1}{M_i} \bar{\beta}_{ij}^{AB}$$

And for the per-patient policy-benefit estimand, I use:

$$\beta_P^{PB} = \frac{1}{N_T} \sum_{ij} \frac{1}{M_i} \beta_{ij}^{PB}$$

In these derivations, I will replace $M_T = N_T(1+p)$, as this simplifies some calculations.

I evaluate the estimands under five different data generating models (which I refer to as 'treatment effect scenarios'): (a) constant treatment effect (the treatment effect is the same across all patients and episodes); (b) the treatment effect varies across episodes (the treatment effect is different in episode 1 vs. episode 2); (c) the treatment effect varies across patients depending on whether $M_i = 1$ vs. $M_i = 2$ (the treatment effect is different in patients who require treatment less often vs. more often); (d) the treatment effect carries forward (patients who received the intervention in their 1st episode have different outcomes in their 2nd episode compared to those who received control in their 1st episode); and (e) the treatment becomes less effective on re-use (patients who received the intervention in their 1st episode have different treatment effects in their 2nd episode compared to those who received control in their 1st episode). The exact data generating mechanism is shown in each example below.

2.5.3.1 Treatment effect scenario 1: constant treatment effect

Consider data generating model 1.3:

$$Y_{ij} = \alpha + \beta Z_{ij} + \mu_i + \varepsilon_{ij}$$

This implies that:

$$\bar{\beta}_{ij}^{AB} = \beta_{ij}^{PB} = \beta \; for \; all \; i, j$$

Therefore, all estimands are the same:

$$\beta_E^{AB} = \beta_E^{PB} = \beta_P^{AB} = \beta_P^{PB} = \beta$$

2.5.3.2 Treatment effect scenario 2: treatment effect varies across episode

Consider the following data generating model:

$$Y_{ij} = \begin{cases} \alpha + \beta_1 Z_{ij} + \mu_i + \varepsilon_{ij} \text{ for } j = 1 \\ \alpha + \beta_2 Z_{ij} + \mu_i + \varepsilon_{ij} \text{ for } j = 2 \end{cases} \tag{2.3}$$

This implies that:

$$\bar{\beta}_{ij}^{AB} = \beta_{ij}^{PB} = \begin{cases} \beta_1 \text{ if } j = 1 \\ \beta_2 \text{ if } j = 2 \end{cases}$$

For the derivation of β_E^{AB} and β_E^{PB} we need to split the numerator into two components: the first component is for episodes for which $\bar{\beta}_{ij}^{AB} = \beta_{ij}^{PB} = \beta_1$, and the second is for episodes for which $\bar{\beta}_{ij}^{AB} = \beta_{ij}^{PB} = \beta_2$. We can see that there are N_T episodes where $\bar{\beta}_{ij}^{AB} = \beta_{ij}^{PB} = \beta_1$ (because there are N_T episodes where $j = 1$), and there are $N_T p$ episodes where $\bar{\beta}_{ij}^{AB} = \beta_{ij}^{PB} = \beta_2$ (because there are $N_T p$ episodes where $j = 2$). Therefore, the first component is $N_T \beta_1$, and the second component is $N_T p \beta_2$. Therefore:

$$\beta_E^{AB} = \beta_E^{PB} = \frac{N_T \beta_1 + N_T p \beta_2}{N_T (1 + p)} = \frac{\beta_1 + p\beta_2}{1 + p}$$

For the derivation of β_P^{AB} and β_P^{PB}, we need to split the numerator into three components: the first component is for $j = 1$ for patients with $M_i = 1$, the second is for $j = 1$ for patients with $M_i = 2$, and the third is for $j = 2$ for patients with $M_i = 2$.

The first component is equal to $N_T (1 - p) \left(\frac{1}{1}\right) \beta_1$, where $N_T (1 - p)$ is the number of episodes in this component, $\left(\frac{1}{1}\right)$ is the weight $(\frac{1}{M_i})$, and β_1 is the treatment effect. Similarly, we can see that the second component is equal to $N_T p \left(\frac{1}{2}\right) \beta_1$, and the third equal to $N_T p \left(\frac{1}{2}\right) \beta_2$. Therefore, the overall estimands are:

$$\beta_P^{AB} = \beta_P^{PB} = \frac{1}{N_T}\left(N_T\left(1-p\right)\left(\frac{1}{1}\right)\beta_1 + N_T p\left(\frac{1}{2}\right)\beta_1 + N_T p\left(\frac{1}{2}\right)\beta_2\right)$$

$$= \left(1-p\right)\beta_1 + \frac{p}{2}\beta_1 + \frac{p}{2}\beta_2 = \left(1-\frac{p}{2}\right)\beta_1 + \frac{p}{2}\beta_2$$

2.5.3.3 Treatment effect scenario 3: Treatment effect varies across patients with different values of M_i

Consider data generating model 2.1:

$$Y_{ij} = \begin{cases} \alpha + \beta_1 Z_{ij} + \mu_i + \varepsilon_{ij} \text{ if } M_i = 1 \\ \alpha + \beta_2 Z_{ij} + \mu_i + \varepsilon_{ij} \text{ if } M_i = 2 \end{cases}$$

This implies that:

$$\bar{\beta}_{ij}^{AB} = \beta_{ij}^{PB} = \begin{cases} \beta_1 \text{ for } M_i = 1 \\ \beta_2 \text{ for } M_i = 2 \end{cases}$$

As before, we can split the numerator of β_E^{AB} and β_E^{PB} into the two components, $N_T\left(1-p\right)\beta_1$ (as there are $N_T\left(1-p\right)$ episodes where $M_i = 1$) and $2N_T p\beta_2$ (because there are $2N_T p$ episodes where $M_i = 2$). Therefore:

$$\beta_E^{AB} = \beta_E^{PB} = \frac{1}{N_T\left(1+p\right)}\left(N_T\left(1-p\right)\beta_1 + 2N_T p\beta_2\right)$$

$$= \frac{\left(1-p\right)\beta_1 + 2p\beta_2}{\left(1+p\right)}$$

For the derivation of β_P^{AB} and β_P^{PB} we need to split the numerator into three components: $N_T\left(1-p\right)\left(\frac{1}{1}\right)\beta_1$, $N_T p\left(\frac{1}{2}\right)\beta_2$, and $N_T p\left(\frac{1}{2}\right)\beta_2$. Because components two and three are the same, these can be combined:

$$\beta_P^{AB} = \beta_P^{PB} = \frac{1}{N_T}\left(N_T\left(1-p\right)\left(\frac{1}{1}\right)\beta_1 + 2N_T p\left(\frac{1}{2}\right)\beta_2\right) = \left(1-p\right)\beta_1 + p\beta_2$$

2.5.3.4 Treatment effect scenario 4: Treatment effect carries forward into 2nd episode

Consider data generating model 2.2:

$$Y_{ij} = \alpha + \beta Z_{ij} + \gamma Z_{i,j-1} + \mu_i + \varepsilon_{ij}$$

For the added-benefit estimands, $\beta_{i2}^{AB\left(\widetilde{Z}=\widetilde{z}\right)} = \beta$ for each value of $Z_{i,j-1}$, and $\beta_{i1}^{AB\left(\widetilde{Z}=\widetilde{z}\right)} = \beta$ as well. Therefore, $\bar{\beta}_{ij}^{AB} = \beta$ for all i, j in this scenario, meaning that $\beta_E^{AB} = \beta_P^{AB} = \beta$.

For the policy-benefit estimands:

$$\beta_{ij}^{PB} = \begin{cases} \beta \text{ for } j = 1 \\ \beta + \gamma \text{ for } j = 2 \end{cases}$$

Therefore:

$$\beta_E^{PB} = \frac{1}{N_T\left(1+p\right)}\left(N_T\beta + N_T p\left(\beta+\gamma\right)\right) = \beta + \frac{p\gamma}{\left(1+p\right)}$$

And:

$$\beta_P^{PB} = \frac{1}{N_T}\left(N_T\left(1-p\right)\left(\frac{1}{1}\right)\beta + N_T p\left(\frac{1}{2}\right)\beta + N_T p\left(\frac{1}{2}\right)\left(\beta+\gamma\right)\right) = \beta + \frac{p\gamma}{2}$$

2.5.3.5 Treatment effect scenario 5: Treatment becomes less effective on re-use

Consider the data generating model:

$$Y_{ij} = \begin{cases} \alpha + \beta Z_{ij} + \mu_i + \varepsilon_{ij} \text{ for } Z_{i,j-1} = 0 \\ \alpha + (\beta+\delta) Z_{ij} + \mu_i + \varepsilon_{ij} \text{ for } Z_{i,j-1} = 1 \end{cases} \tag{2.4}$$

i.e. the treatment effect is β the first time an intervention is used, and $\beta + \delta$ the second time it is used (for simplicity, I have coded δ as positive, but we can imagine it as a negative quantity so that the treatment effect is smaller the second time the intervention is used).

For the added-benefit estimands, $\beta_{i2}^{AB\left(\widetilde{Z}=\widetilde{z}\right)} = \beta$ for $Z_{i,j-1} = 0$, and $\beta_{i2}^{AB\left(\widetilde{Z}=\widetilde{z}\right)} = \beta + \delta$ for $Z_{i,j-1} = 1$ (and $\beta_{i1}^{AB\left(\widetilde{Z}=\widetilde{z}\right)} = \beta$). Therefore, $\bar{\beta}_{i1}^{AB} = \beta$, and because treatment history $Z_{i,j-1} = 0$ and $Z_{i,j-1} = 1$ both have a 50% probability of occurring at episode 2 (because there is a 1:1 allocation ratio and no non-enrolment), $\bar{\beta}_{i2}^{AB}$ is $\frac{1}{2}\left(\beta + (\beta+\delta)\right) = \beta + \frac{\delta}{2}$. We can split the numerator for the per-episode estimand into two components, the first for episodes where $j = 1$, and the second where $j = 2$; there are N_T episodes for the first component, $N_T p$ for the second components. The estimand therefore is:

$$\beta_E^{AB} = \frac{1}{N_T\left(1+p\right)}\left(N_T\beta + N_T p(\beta + \frac{\delta}{2})\right) = \beta + \frac{p\delta}{2\left(1+p\right)}$$

The numerator of the per-patient estimand can be split into three components. The first is $N_T\,(1-p)\left(\frac{1}{1}\right)\beta$ (corresponding to $j=1$ for patients with $M_i=1$), the second is $N_T p\left(\frac{1}{2}\right)\beta$ (corresponding to $j=1$ for patients with $M_i=2$), and the third component is $N_T p\left(\frac{1}{2}\right)\left(\beta+\frac{\delta}{2}\right)$ (corresponding to $j=2$ for patients with $M_i=2$). Putting this together, we have:

$$\beta_P^{AB} = \frac{1}{N_T}\left(N_T(1-p)(\frac{1}{1})\beta + N_T p(\frac{1}{2})\beta + N_T p(\frac{1}{2})(\beta+\frac{\delta}{2})\right) = \beta + \frac{p\delta}{4}$$

For the policy-benefit estimands:

$$\beta_{ij}^{PB} = \begin{cases} \beta \text{ for } j=1 \\ \beta+\delta \text{ for } j=2 \end{cases}$$

Then:

$$\beta_E^{PB} = \frac{1}{N_T\,(1+p)}\left(N_T\beta + N_T p\,(\beta+\delta)\right) = \beta + \frac{p\delta}{(1+p)}$$

And:

$$\beta_P^{PB} = \frac{1}{N_T}\left(N_T(1-p)(\frac{1}{1})\beta + N_T p(\frac{1}{2})\beta + N_T p(\frac{1}{2})(\beta+\delta)\right) = \beta + \frac{p\delta}{2}$$

2.5.3.6 Summary of differences between estimands

Values for the different estimands under different treatment effect scenarios is shown in table 2.2.

All four estimands coincide under a constant treatment effect mechanism; this is because the cluster size is not informative (i.e. the potential treatment effects $\bar{\beta}_{ij}^{AB}$ and β_{ij}^{PB} do not differ according to M_i), and the potential outcomes and potential treatment effects ($Y_{ij}^{\left(Z=0,\widetilde{Z}=\widetilde{z}\right)}$ and $Y_{ij}^{\left(Z=1,\widetilde{Z}=\widetilde{z}\right)} - Y_{ij}^{\left(Z=0,\widetilde{Z}=\widetilde{z}\right)}$) are not affected by $\widetilde{Z_{ij}}$.

When the treatment effect varies across episodes or across patients with different values of M_i, the per-episode and per-patient estimands differ (because the cluster size is informative), though the added-benefit and policy-benefit estimands do not (because treatment history does not affect either the potential outcomes or potential treatment effects).

When the treatment effect carries forward, the per-episode added-benefit and per-patient added-benefit estimands coincide, because the $\bar{\beta}_{ij}^{AB}$ treatment effect does

not differ according to M_i. However, the per-episode and per-patient policy-benefit estimands differ; this is because the β_{ij}^{PB} treatment effects do differ according to M_i.

Finally, when the intervention becomes less effective on re-use, all four estimands differ. This is because both sets of potential treatment effects ($\bar{\beta}_{ij}^{AB}$ and β_{ij}^{PB}) differ according to M_i, and the potential treatment effects ($Y_{ij}^{\left(Z=1,\widetilde{Z}=\widetilde{z}\right)} - Y_{ij}^{\left(Z=0,\widetilde{Z}=\widetilde{z}\right)}$) are affected by $\widetilde{Z_{ij}}$.

2.6 Discussion

In this chapter I have proposed a set of treatment estimands that can be used in multi-episode settings. Each estimand corresponds to a different question around the usefulness of the intervention. These estimands can be used by researchers to help decide on their primary question of interest, and then identify a trial design and method of analysis that matches the chosen estimand. Furthermore, researchers can use this proposed framework to define alternate estimands if required.

In the next two chapters I evaluate different methods of estimating these estimands in re-randomisation trials (chapters 3 and 4), and then compare re-randomisation to other trial designs (parallel group, cluster, crossover) in chapter 5 to see which designs are best suited to which estimands.

Table 2.2: Treatment estimands under different treatment effect mechanisms in a fictitious trial with a maximum of two episodes. p denotes the proportion of patients in the trial who enrolled for two episodes (i.e. $p = \frac{M_{T(2)}}{N_T}$).

Scenario	**Data generating model**	**Per-episode added-benefit** (β_E^{AB})	**Per-episode policy-benefit** (β_E^{PB})	**Per-patient added-benefit** (β_P^{AB})	**Per-patient policy-benefit** (β_P^{PB})
Constant treatment effect	$Y_{ij} = \alpha + \beta Z_{ij} + \mu_i + \varepsilon_{ij}$	β	β	β	β
Treatment effect varies across episodes	$Y_{ij} = \begin{cases} \alpha + \beta_1 Z_{ij} + \mu_i + \varepsilon_{ij} \text{ for } j = 1 \\ \alpha + \beta_2 Z_{ij} + \mu_i + \varepsilon_{ij} \text{ for } j = 2 \end{cases}$	$\frac{\beta_1 + p\beta_2}{1+p}$	$\frac{\beta_1 + p\beta_2}{1+p}$	$\left(1 - \frac{p}{2}\right)\beta_1 + \frac{p}{2}\beta_2$	$\left(1 - \frac{p}{2}\right)\beta_1 + \frac{p}{2}\beta_2$
Treatment effect varies across patients with different values of M_i	$Y_{ij} = \begin{cases} \alpha + \beta_1 Z_{ij} + \mu_i + \varepsilon_{ij} \text{ if } M_i = 1 \\ \alpha + \beta_2 Z_{ij} + \mu_i + \varepsilon_{ij} \text{ if } M_i = 2 \end{cases}$	$\frac{(1-p)\beta_1 + 2p\beta_2}{(1+p)}$	$\frac{(1-p)\beta_1 + 2p\beta_2}{(1+p)}$	$(1-p)\beta_1 + p\beta_2$	$(1-p)\beta_1 + p\beta_2$
Treatment effect carries forward	$Y_{ij} = \alpha + \beta Z_{ij} + \gamma Z_{i,j-1} + \mu_i + \varepsilon_{ij}$	β	$\beta + \frac{p\gamma}{(1+p)}$	β	$\beta + \frac{p\gamma}{2}$
Treatment becomes less effective on re-use	$Y_{ij} = \begin{cases} \alpha + \beta Z_{ij} + \mu_i + \varepsilon_{ij} \text{ for } Z_{i,j-1} = 0 \\ \alpha + (\beta + \delta) Z_{ij} + \mu_i + \varepsilon_{ij} \text{ for } Z_{i,j-1} = 1 \end{cases}$	$\beta + \frac{p\delta}{2(1+p)}$	$\beta + \frac{p\delta}{(1+p)}$	$\beta + \frac{p\delta}{4}$	$\beta + \frac{p\delta}{2}$

3 *Independence estimators for re-randomisation trials*

In this chapter I will describe a set of independence estimators that can be used to analyse re-randomisation trials. As described in chapter 1, independence estimators have been previously evaluated for re-randomisation trials, in the form of analysis model 1.1:

$$Y_{ij} = \hat{\alpha} + \hat{\beta} Z_{ij} + \varepsilon_{ij}$$

Briefly, independence estimators use a working independence correlation structure, that is, estimates of treatment effect are based on a working assumption that episodes from the same patient are uncorrelated. Previous articles have evaluated analysis model 1.1 under a constant treatment effect mechanism (e.g. data generating model 1.3), but have not explicitly evaluated this estimator against any of the estimands defined in chapter 2.

In this chapter I will propose a set of independence estimators for re-randomisation trials, one for each estimand in chapter 2 (per-episode added-benefit, per-patient added-benefit, per-episode policy-benefit, and per-patient policy-benefit). I will then evaluate the bias of these estimators against their corresponding estimand under a range of data generating mechanisms and non-enrolment scenarios.

3.1 Independence estimators

3.1.1 Per-episode added-benefit estimator

In order to correspond to the per-episode added-benefit estimand, our estimator must fulfil two conditions: (1) it must give equal weight to each episode; and (2) it must compare outcomes between intervention and control at each episode for patients with the same treatment history.

The simplest way to do this is to take a difference in means between the intervention and control episodes:

$$\hat{\beta}_E^{AB} = \frac{\sum_{ij} Y_{ij} Z_{ij}}{\sum_{ij} Z_{ij}} - \frac{\sum_{ij} Y_{ij} (1 - Z_{ij})}{\sum_{ij} (1 - Z_{ij})} \tag{3.1}$$

This estimator gives equal weight to each episode, and implicitly compares outcomes at each episode for patients with the same treatment history (i.e. the above estimator could be re-written as a weighted average across the different treatment histories at each episode, with weights based on how often each history had occurred in the trial).

This estimator is identical to analysis model 1.1 (i.e. equation 3.1 is the least-squares estimate from analysis model 1.1). This implies that previous research evaluating independence estimators in the form of analysis model 1.1 have been evaluating an analysis approach which targets the per-episode added-benefit estimand.

3.1.2 Per-patient added-benefit estimator

For the per-patient added-benefit estimator we need to re-write the estimator in equation 3.1 so that each patient has equal weight, rather than each episode. This can be done by weighting each patient by the inverse of their number of episodes (i.e. $W_i = \frac{1}{M_i}$), which leads to the estimator:

$$\hat{\beta}_P^{AB} = \frac{\sum_{ij} W_i Y_{ij} Z_{ij}}{\sum_{ij} W_i Z_{ij}} - \frac{\sum_{ij} W_i Y_{ij} (1 - Z_{ij})}{\sum_{ij} W_i (1 - Z_{ij})} \tag{3.2}$$

This estimator is the least squares estimate of analysis model 1.1 when patients are weighted by $W_i = \frac{1}{M_i}$, which requires minimizing the function:

$$\sum_{ij} W_i (Y_{ij} - \hat{\alpha} - \hat{\beta} Z_{ij})^2$$

3.1.3 Per-episode policy-benefit estimator

To estimate the per-episode policy-benefit treatment effect, we must first estimate the β_{ij}^{PB}'s (as defined in section 2.4.3.2) and then use these to calculate the overall treatment effect. We can do this by specifying a causal model for the effect of treatment history ($\widetilde{Z}_{ij}$) on the potential outcomes. For example, in a trial where patients experience a maximum of two episodes, we might assume the following analysis model:

$$Y_{ij} = \hat{\alpha} + \hat{\beta} Z_{ij} + \hat{\gamma} Z_{i,j-1} + \hat{\delta} Z_{ij} Z_{i,j-1} + \hat{\beta}_{ep} X_{ep_{ij}} + \varepsilon_{ij} \tag{3.3}$$

where $Z_{i,j-1}$ is the treatment allocation in the previous episode (and is set to 0 for $j = 1$), and $X_{ep_{ij}}$ is an indicator for episode 2 (i.e. $X_{ep_{ij}} = 1$ for episode 2, and 0 otherwise). This model allows the effect of the intervention in episode 1 to carry forward into episode 2 (the term $\hat{\gamma}$), and for the intervention to get more (or less) effective the 2nd time it is used (the term $\hat{\delta}$).

After obtaining estimates for $\hat{\beta}$, $\hat{\gamma}$, and $\hat{\delta}$, we can use these to estimate the β_{ij}^{PB}'s; for example, following on from section 2.4.3.2 and model 3.3 above, the $\hat{\beta}_{ij}^{PB}$ for all first episodes is $\hat{\beta}$, and for all second episodes is $\hat{\beta} + \hat{\gamma} + \hat{\delta}$.

We can then use these estimates to get an overall estimate of β_E^{PB} using results from section 2.4.3.2 as follows:

$$\hat{\beta}_E^{PB} = \frac{N_1}{M_T}\left(\hat{\beta}\right) + \frac{N_2}{M_T}\left(\hat{\gamma} + \hat{\beta} + \hat{\delta}\right)$$

Although the term $\hat{\beta}_{ep}$ is not directly used to estimate the treatment effect, it is necessary to include $X_{ep_{ij}}$ in model 3.3, as estimates may be biased otherwise. This is because $X_{ep_{ij}}$ is associated with $Z_{i,j-1}$, and so may act as a confounder if omitted from the model, resulting in biased estimates of γ. For trials with $j > 2$, separate indicator variables for each episode should be included in the model.

3.1.4 Per-patient policy-benefit estimator

For the per-patient policy-benefit estimator we need to obtain estimates from model 3.3 using weighted least-squares, where each patient is weighted by the inverse of their number of episodes, $W_i = \frac{1}{M_i}$. After obtaining estimates and calculating the $\hat{\beta}_{ij}^{PB}$'s, the overall treatment effect is calculated as:

$$\hat{\beta}_P^{PB} = \frac{1}{N_T}\sum_{ij} W_i \hat{\beta}_{ij}^{PB}$$

Using analysis model 3.3 above, this equates to:

$$\hat{\beta}_P^{PB} = \frac{M_{T(1)}}{N_T}\left(\hat{\beta}\right) + \frac{M_{T(2)}}{N_T}\left(\frac{1}{2}\hat{\beta} + (\frac{1}{2})(\hat{\beta} + \hat{\gamma} + \hat{\delta})\right)$$

where $M_{T(j)}$ represents the total number of patients for whom $M_i = j$. In this equation, $\frac{M_{T(1)}}{N_T}\left(\hat{\beta}\right)$ is the component for patients where $M_i = 1$, and $\frac{M_{T(2)}}{N_T}\left(\frac{1}{2}\hat{\beta} + \left(\frac{1}{2}\right)\left(\hat{\beta} + \hat{\gamma} + \hat{\delta}\right)\right)$ is the component for patients where $M_i = 2$ (with $\frac{1}{2}\hat{\beta}$ being the 1st episode component, and $\left(\frac{1}{2}\right)\left(\hat{\beta} + \hat{\gamma} + \hat{\delta}\right)$ being the 2nd episode component).

3.2 Mathematical derivation of bias

In this section I will evaluate the bias of the independence estimators for the per-episode and per-patient added-benefit estimands. I evaluate the policy-benefit estimators in a simulation study in section 3.4. As discussed in chapter 1, I will restrict the setting to a trial with a 1:1 allocation ratio, where patients experience a maximum of two episodes.

In section 3.2.1 I will discuss the number of patients in each different treatment sequence that could occur in a re-randomisation trial in this setting (as this will be used in the derivation of the expected value of the independence estimators). Then

in section 3.3 I will derive the expected value of the independence estimators. I will then compare the expected values against the true estimand values under a range of different data generating and non-enrolment mechanisms in section 3.3.4.

3.2.1 Number of patients at each episode for different sequences

In this section, I derive the number of episodes in the different treatment sequences that could occur in a re-randomisation trial; this information is used in the next section (section 3.3.1). For the moment, I will assume that there is no non-enrolment, i.e. patients who experience two episodes will re-enrol in the trial for their second episode. This assumption will be relaxed in sections 3.3.10 and 3.3.11.

Under these assumptions, the asymptotic number of episodes in each treatment sequence are shown in table 3.1 (where p represents the proportion of patients in the trial who experience two episodes). As a brief reminder, N_T is the total number of patients enrolled in the trial, M_T the total number of episodes, M_i is the number of episodes for which patient i is enrolled in the trial, and $M_{T(j)}$ denotes the number of patients for whom $M_i = j$. Then, $p = \frac{M_{T(2)}}{N_T}$ is the proportion of patients who are re-enrolled for a second episode.

In this setting, there are six possible treatment sequences (shown in table 3.1). Let Z denote the treatment sequence; i.e. Z is one of (0), (1), $(0,0)$, $(0,1)$, $(1,0)$, or $(1,1)$. There are $N_T\,(1-p)$ patients enrolled for a single episode; therefore, there are $\frac{N_T}{2}\,(1-p)$ patients in treatment sequences $Z=(0)$ and $Z=(1)$ respectively. There are $N_T p$ patients enrolled for two episodes; therefore, there are $\frac{N_T p}{4}$ patients in treatment sequences $Z=(0,0)$, $Z=(0,1)$, $Z=(1,0)$, and $Z=(1,1)$. The number of patients at each episode for the different treatment sequences is shown in table 3.1.

Table 3.1: Asymptotic number of patients in each treatment sequence at each episode in a re-randomisation trial with 1:1 allocation ratio and no non-enrolment

	Treatment allocation		**Number of patients**	
Sequence	**Episode 1**	**Episode 2**	**Episode 1**	**Episode 2**
$Z=(1)$	1	–	$\frac{N_T}{2}\,(1-p)$	–
$Z=(1,0)$	1	0	$\frac{N_T p}{4}$	$\frac{N_T p}{4}$
$Z=(1,1)$	1	1	$\frac{N_T p}{4}$	$\frac{N_T p}{4}$
$Z=(0)$	0	–	$\frac{N_T}{2}\,(1-p)$	–
$Z=(0,1)$	0	1	$\frac{N_T p}{4}$	$\frac{N_T p}{4}$
$Z=(0,0)$	0	0	$\frac{N_T p}{4}$	$\frac{N_T p}{4}$

3.3 Expected values of independence estimators

3.3.1 Between- and within-patient estimation components

The treatment sequences in table 3.1 can be split into into three different estimation components; two are between-patient estimation components and one is a within-

patient estimation component [9]. I derive these components below. These estimation components will be used as the basis for the mathematical derivations in this chapter, as well as in chapters 4 and 5.

Let $\hat{\beta}_{B_1}$ represent the first between-patient estimation component; it is based on treatment sequences $Z = (0)$ and $Z = (1)$. It uses between-patient information for patients who are enrolled for one episode. It is calculated as:

$$\hat{\beta}_{B_1} = \frac{\sum_{i \in Z=(1)} Y_{i1}}{\frac{N_T}{2}(1-p)} - \frac{\sum_{i \in Z=(0)} Y_{i1}}{\frac{N_T}{2}(1-p)}$$
$$= \frac{1}{\frac{N_T}{2}(1-p)} \left(\sum_{i \in Z=(1)} Y_{i1} - \sum_{i \in Z=(0)} Y_{i1} \right) \quad (3.4)$$

i.e. it is the mean of all episodes in treatment sequence $Z = (1)$ vs. the mean of all episodes in treatment sequence $Z = (0)$ (where the denominators are from table 3.1, and are derived in section 3.2.1).

Let $\hat{\beta}_{B_2}$ represent the second between-patient estimation component; it is based on treatment sequences $Z = (0,\ 0)$ and $Z = (1,\ 1)$. It uses between-patient information for patients who are enrolled for two episodes and allocated to the same treatment for each episode. It is calculated as:

$$\hat{\beta}_{B_2} = \frac{\sum_{i \in Z=(1,1)} (Y_{i1} + Y_{i2})}{\frac{N_T p}{4} + \frac{N_T p}{4}} - \frac{\sum_{i \in Z=(0,0)} (Y_{i1} + Y_{i2})}{\frac{N_T p}{4} + \frac{N_T p}{4}}$$
$$= \frac{1}{\frac{N_T p}{2}} \left(\sum_{i \in Z=(1,1)} (Y_{i1} + Y_{i2}) - \sum_{i \in Z=(0,0)} (Y_{i1} + Y_{i2}) \right) \quad (3.5)$$

i.e. it is the mean of all episodes in treatment sequence $Z = (1,\ 1)$ vs. the mean of all episodes in treatment sequences $Z = (0,\ 0)$ (where the denominators are shown in table 3.1).

Let $\hat{\beta}_W$ represents the within-patient estimation component; it is based on treatment sequences $Z = (0,\ 1)$ and $Z = (1,\ 0)$. It uses within-patient information for patients who are enrolled for two episodes and allocated to different treatments for each episode. It is calculated as:

$$\hat{\beta}_W = \frac{\sum_{i \in Z=(1,0)} (Y_{i1} - Y_{i2})}{\frac{N_T p}{4} + \frac{N_T p}{4}} - \frac{\sum_{i \in Z=(0,1)} (Y_{i1} - Y_{i2})}{\frac{N_T p}{4} + \frac{N_T p}{4}}$$
$$= \frac{1}{\frac{N_T p}{2}} \left(\sum_{i \in Z=(1,0)} (Y_{i1} - Y_{i2}) - \sum_{i \in Z=(0,1)} (Y_{i1} - Y_{i2}) \right) \quad (3.6)$$

i.e. it is the mean of intervention episodes for patients on treatment sequences $Z = (0\ 1)$ and $Z = (1\ 0)$ vs. the mean of control episodes for patients on treatment sequences $Z = (0\ 1)$ and $Z = (1\ 0)$ (where the denominators are shown in table 3.1).

3.3.2 Per-episode added-benefit estimator

This estimator can be written in terms of the components $\hat{\beta}_{B_1}$, $\hat{\beta}_{B_2}$, and $\hat{\beta}_W$, as:

$$\hat{\beta}_E^{AB} = \frac{\sum_{ij} Y_{ij} Z_{ij}}{\sum_{ij} Z_{ij}} - \frac{\sum_{ij} Y_{ij} (1 - Z_{ij})}{\sum_{ij} (1 - Z_{ij})} = \frac{\frac{N_T}{2}(1-p)\hat{\beta}_{B_1} + \frac{N_T p}{2}\hat{\beta}_{B_2} + \frac{N_T p}{2}\hat{\beta}_W}{\frac{N_T}{2}(1-p) + \frac{N_T p}{2} + \frac{N_T p}{2}}$$
$$= \frac{(1-p)\hat{\beta}_{B_1} + p\hat{\beta}_{B_2} + p\hat{\beta}_W}{1+p} \quad (3.7)$$

Taking the expectation leads to:

$$E\left(\hat{\beta}_E^{AB}\right) = \frac{(1-p)E\left(\hat{\beta}_{B_1}\right) + pE\left(\hat{\beta}_{B_2}\right) + pE\left(\hat{\beta}_W\right)}{1+p} \quad (3.8)$$

3.3.3 Per-patient added-benefit estimator

This estimator can also be written in terms of the components $\hat{\beta}_{B_1}$, $\hat{\beta}_{B_2}$, and $\hat{\beta}_W$, by using equation 3.7 above and weighting each component by $\frac{1}{M_i}$:

$$\hat{\beta}_P^{AB} = \frac{\sum_{ij} W_i Y_{ij} Z_{ij}}{\sum_{ij} W_i Z_{ij}} - \frac{\sum_{ij} W_i Y_{ij} (1 - Z_{ij})}{\sum_{ij} W_i (1 - Z_{ij})}$$
$$= \frac{\left(\frac{1}{1}\right)\left(\frac{N_T}{2}\right)(1-p)\hat{\beta}_{B_1} + \left(\frac{1}{2}\right)\left(\frac{N_T p}{2}\right)\hat{\beta}_{B_2} + \left(\frac{1}{2}\right)\left(\frac{N_T p}{2}\right)\hat{\beta}_W}{\left(\frac{1}{1}\right)\left(\frac{N_T}{2}\right)(1-p) + \left(\frac{1}{2}\right)\left(\frac{N_T p}{2}\right) + \left(\frac{1}{2}\right)\left(\frac{N_T p}{2}\right)} = (1-p)\hat{\beta}_{B_1} + \frac{p}{2}\hat{\beta}_{B_2} + \frac{p}{2}\hat{\beta}_W \quad (3.9)$$

Taking the expectation leads to:

$$E\left(\hat{\beta}_P^{AB}\right) = (1-p)E\left(\hat{\beta}_{B_1}\right) + \frac{p}{2}E\left(\hat{\beta}_{B_2}\right) + \frac{p}{2}E\left(\hat{\beta}_W\right) \quad (3.10)$$

3.3.4 Expected value of estimators under different data generating mechanisms and non-enrolment scenarios

I now evaluate the expected values of the per-episode and per-patient added-benefit estimators under a range of data generating and non-enrolment scenarios. I evaluate seven scenarios in total:

1. Constant treatment effect

2. Treatment effect varies across episode
3. Treatment effect varies across value of M_i
4. Treatment effect carries forward into second episode
5. Treatment becomes less effective on re-use
6. Constant treatment effect, differential non-enrolment based on outcome in previous episode
7. Constant treatment effect, differential non-enrolment based on expected outcome in current episode

The first five scenarios use the same treatment effect mechanisms discussed in section 2.5.3 of chapter 2, and do not involve any non-enrolment (i.e. patients who experience two episodes will enrol in the trial for both episodes). The last two scenarios use a constant treatment effect mechanism (i.e. model 1.3), but some patients who experience two episodes do not re-enrol for their second episode. In these two scenarios I examine the impact of *differential non-enrolment*. In this thesis I define differential non-enrolment to mean that different types of patients from the episode 1 intervention and control groups will re-enrol for episode 2. For example, in the episode 1 intervention group healthier patients are more likely to re-enrol than sicker patients, but in the episode 1 control group sicker patients are more likely to re-enrol. In scenario 6, non-enrolment is differential between treatment arms depending on their outcome in the first episode; patients who received the intervention in episode 1 and had a good outcome have the same probability of re-enrolling as patients who received control in episode 1 and had a bad outcome, and vice versa. In scenario 7, non-enrolment is differential between treatment arms depending on their expected outcome in the second episode (i.e. their baseline prognosis at episode 2); patients who received the intervention in episode 1 and have a good baseline prognosis at episode 2 have the same probability of re-enrolling as patients who received control in episode 1 and had a bad prognosis, and vice versa. Further details on each of these scenarios is given in the sections below.

For each scenario, I will assume that $\alpha = 0$ (this makes calculations slightly easier and has no impact on the results). Expected values of the components $\hat{\beta}_{B_1}$, $\hat{\beta}_{B_2}$, and $\hat{\beta}_W$ under the different scenarios are shown in table 3.2.

3.3.5 Scenario 1 (S1): constant treatment effect

Consider data-generating mechanism 1.3:

$$Y_{ij} = \alpha + \beta Z_{ij} + \mu_i + \varepsilon_{ij}$$

Table 3.2: Expected values of between- and within-patient estimators under different scenarios

Scenario	$E(\hat{\beta}_{B_1})$	$E(\hat{\beta}_{B_2})$	$E(\hat{\beta}_W)$
S1 – Constant treatment effect	β	β	β
S2 – Treatment effect varies across episode	β_1	$\frac{1}{2}(\beta_1 + \beta_2)$	$\frac{1}{2}(\beta_1 + \beta_2)$
S3 – Treatment effect varies across value of M_i	β_1	β_2	β_2
S4 – Treatment effect carries forward into second episode	β	$\beta + \frac{\gamma}{2}$	$\beta - \frac{\gamma}{2}$
S5 – Treatment becomes less effective on re-use	β	$\beta + \frac{\delta}{2}$	β
S6 – Constant treatment effect, differential non-enrolment based on outcome in previous episode	$\beta_{trt} + \beta_{X_{PL}} \frac{(p_{01} - p_{00})}{2(1-p)}$	$\beta_{trt} + \beta_{X_{PL}} \frac{(p_{00} - p_{01})}{2p}$	β_{trt}
S7 – Constant treatment effect, differential non-enrolment based on expected outcome in current episode	β_{trt}	$\beta_{trt} + \frac{\beta_{X_{EL}}}{4}\left(\frac{p_{00} - p_{01}}{p}\right)$	$\beta_{trt} - \frac{\beta_{X_{EL}}}{4}\left(\frac{p_{00} - p_{01}}{p}\right)$

The values of $E(Y_{ij})$ for each treatment sequence are shown in table 3.3. From this we can see that $E\left(\hat{\beta}_{B_1}\right) = E\left(\hat{\beta}_{B_2}\right) = E\left(\hat{\beta}_W\right) = \beta$.

Therefore:

$$E\left(\hat{\beta}_E^{AB}\right) = \frac{(1-p)\beta + p\beta + p\beta}{1+p} = \beta$$

And:

$$E\left(\hat{\beta}_P^{AB}\right) = (1-p)\beta + \frac{p}{2}\beta + \frac{p}{2}\beta = \beta$$

Table 3.3: Value of $E(Y_{ij})$ across each episode and treatment sequence under *S1: constant treatment effect*

Treatment allocation		$E(Y_{ij})$	
Episode 1	**Episode 2**	**Episode 1**	**Episode 2**
0	–	0	–
0	1	0	β
0	0	0	0
1	–	β	–
1	0	β	0
1	1	β	β

3.3.6 S2: Treatment effect varies across episode

Consider data-generating mechanism 2.3:

$$Y_{ij} = \begin{cases} \alpha + \beta_1 Z_{ij} + \mu_i + \varepsilon_{ij} \text{ for } j = 1 \\ \alpha + \beta_2 Z_{ij} + \mu_i + \varepsilon_{ij} \text{ for } j = 2 \end{cases}$$

The values of $E(Y_{ij})$ for each treatment sequence are shown in table 3.4. Then, plugging the values from table 3.4 into the formulas from section 3.3.1, we can see that the expected values of the components, $\hat{\beta}_{B_1}$, $\hat{\beta}_{B_2}$, and $\hat{\beta}_W$ are:

$$E\left(\hat{\beta}_{B_1}\right) = \beta_1$$

$$E\left(\hat{\beta}_{B_2}\right) = \frac{1}{2}(\beta_1 + \beta_2)$$

$$E\left(\hat{\beta}_W\right) = \frac{1}{2}(\beta_1 + \beta_2)$$

Therefore:

$$E\left(\hat{\beta}_E^{AB}\right) = \frac{(1-p)\,\beta_1 + \frac{p}{2}\,(\beta_1+\beta_2) + \frac{p}{2}\,(\beta_1+\beta_2)}{1+p} = \frac{\beta_1 + p\beta_2}{1+p}$$

And:

$$E\left(\hat{\beta}_P^{AB}\right) = (1-p)\,\beta_1 + \frac{p}{2}\frac{(\beta_1+\beta_2)}{2} + \frac{p}{2}\frac{(\beta_1+\beta_2)}{2} = \left(1-p+\frac{p}{2}\right)\beta_1 + \frac{p}{2}\beta_2$$

$$= \left(1-\frac{p}{2}\right)\beta_1 + \frac{p}{2}\beta_2$$

Table 3.4: Value of $E\left(Y_{ij}\right)$ across each episode and treatment sequence under *S2: Treatment effect varies across episode*

Treatment allocation		$E\left(Y_{ij}\right)$	
Episode 1	**Episode 2**	**Episode 1**	**Episode 2**
0	–	0	–
0	1	0	β_2
0	0	0	0
1	–	β_1	–
1	0	β_1	0
1	1	β_1	β_2

3.3.7 S3: Treatment effect varies across value of M_i

Consider data-generating mechanism 2.1:

$$Y_{ij} = \begin{cases} \alpha + \beta_1 Z_{ij} + \mu_i + \varepsilon_{ij} \text{ if } M_i = 1 \\ \alpha + \beta_2 Z_{ij} + \mu_i + \varepsilon_{ij} \text{ if } M_i = 2 \end{cases}$$

The values of $E\left(Y_{ij}\right)$ for each treatment sequence are shown in table 3.5. Then, plugging the values from table 3.5 into the formulas from section 3.3.1, we can see that the expected values of the components, $\hat{\beta}_{B_1}$, $\hat{\beta}_{B_2}$, and $\hat{\beta}_W$ are:

$$E\left(\hat{\beta}_{B_1}\right) = \beta_1$$

$$E\left(\hat{\beta}_{B_2}\right) = \beta_2$$

$$E\left(\hat{\beta}_W\right) = \beta_2$$

Therefore:

$$E\left(\hat{\beta}_E^{AB}\right) = \frac{(1-p)\,\beta_1 + p\beta_2 + p\beta_2}{1+p} = \frac{(1-p)\,\beta_1 + 2p\beta_2}{1+p}$$

And:

$$E\left(\hat{\beta}_P^{AB}\right) = (1-p)\,\beta_1 + \frac{p}{2}\beta_2 + \frac{p}{2}\beta_2 = (1-p)\,\beta_1 + p\beta_2$$

Table 3.5: Value of $E\left(Y_{ij}\right)$ across each episode and treatment sequence under *S3: Treatment effect varies across value of* M_i

Treatment allocation		$E\left(Y_{ij}\right)$	
Episode 1	**Episode 2**	**Episode 1**	**Episode 2**
0	–	0	–
0	1	0	β_2
0	0	0	0
1	–	β_1	–
1	0	β_2	0
1	1	β_2	β_2

3.3.8 S4: Treatment effect carries forward into the second episode

Consider data-generating mechanism 2.2:

$$Y_{ij} = \alpha + \beta Z_{ij} + \gamma Z_{i,j-1} + \mu_i + \varepsilon_{ij}$$

The values of $E\left(Y_{ij}\right)$ for each treatment sequence are shown in table 3.6. Then, plugging the values from table 3.6 into the formulas from section 3.3.1, we can see that the expected values of the components, $\hat{\beta}_{B_1}$, $\hat{\beta}_{B_2}$, and $\hat{\beta}_W$ are:

$$E\left(\hat{\beta}_{B_1}\right) = \beta$$

$$E\left(\hat{\beta}_{B_2}\right) = \beta + \frac{\gamma}{2}$$

$$E\left(\hat{\beta}_W\right) = \beta - \frac{\gamma}{2}$$

Therefore:

$$E\left(\hat{\beta}_E^{AB}\right) = \frac{(1-p)\,\beta + p\left(\beta + \frac{\gamma}{2}\right) + p\left(\beta - \frac{\gamma}{2}\right)}{1+p} = \beta$$

And:

$$E\left(\hat{\beta}_P^{AB}\right) = (1-p)\,\beta + \frac{p}{2}\left(\beta + \frac{\gamma}{2}\right) + \frac{p}{2}\left(\beta - \frac{\gamma}{2}\right) = \beta$$

Table 3.6: Value of $E(Y_{ij})$ across each episode and treatment sequence under *S4: Treatment effect carries forward into second episode*

Treatment allocation		$E(Y_{ij})$	
Episode 1	**Episode 2**	**Episode 1**	**Episode 2**
0	–	0	–
0	1	0	β
0	0	0	0
1	–	β	–
1	0	β	γ
1	1	β	$\beta + \gamma$

3.3.9 S5: Treatment becomes less effective on re-use

Consider data-generating mechanism 2.4:

$$Y_{ij} = \begin{cases} \alpha + \beta Z_{ij} + \mu_i + \varepsilon_{ij} \text{ for } Z_{i,j-1} = 0 \\ \alpha + (\beta + \delta) Z_{ij} + \mu_i + \varepsilon_{ij} \text{ for } Z_{i,j-1} = 1 \end{cases}$$

The values of $E(Y_{ij})$ for each treatment sequence are shown in table 3.7. Then, plugging the values from table 3.7 into the formulas from section 3.3.1, we can see that the expected values of the components, $\hat{\beta}_{B_1}$, $\hat{\beta}_{B_2}$, and $\hat{\beta}_W$ are:

$$E\left(\hat{\beta}_{B_1}\right) = \beta$$

$$E\left(\hat{\beta}_{B_2}\right) = \beta + \frac{\delta}{2}$$

$$E\left(\hat{\beta}_W\right) = \beta$$

Therefore:

$$E\left(\hat{\beta}_E^{AB}\right) = \frac{(1-p)\,\beta + p\left(\beta + \frac{\delta}{2}\right) + p\beta}{1+p} = \frac{(1+p)\,\beta + \frac{p\delta}{2}}{1+p} = \beta + \frac{p\delta}{2\,(1+p)}$$

And:

$$E\left(\hat{\beta}_P^{AB}\right) = (1-p)\,\beta + \frac{p}{2}\left(\beta + \frac{\delta}{2}\right) + \frac{p}{2}\beta = \beta + \frac{p\delta}{4}$$

Table 3.7: Value of $E(Y_{ij})$ across each episode and treatment sequence under *S5: Treatment becomes less effective on re-use*

Treatment allocation		$E(Y_{ij})$	
Episode 1	**Episode 2**	**Episode 1**	**Episode 2**
0	–	0	–
0	1	0	β
0	0	0	0
1	–	β	–
1	0	β	0
1	1	β	$\beta + \delta$

3.3.10 S6: Constant treatment effect, differential non-enrolment based on outcome in previous episode

In this section (and in section 3.3.11), I will no longer assume that there is no non-enrolment. Instead, I will assume that some patients do not re-enrol for their second episode. In this scenario I consider a situation where non-enrolment is differential between treatment arms depending on their outcome in the first episode; patients who received the intervention in episode 1 and had a good outcome have the same probability of re-enrolling as patients who received control in episode 1 and had a bad outcome, and vice versa.

Consider the following data generating mechanism:

$$Y_{ij} = \alpha + \beta_{trt} Z_{ij} + \beta_{X_{PL}} X_{PL_i} + \mu_i + \varepsilon_{ij} \tag{3.11}$$

where X_{PL_i} is an unobserved binary patient-level variable (i.e. it is constant across episodes). I use the subscript PL to denote 'patient-level'. This data generating mechanism is equivalent to model 1.3, but with the addition of X_{PL_i} (and β_{trt} in place of β), and so the treatment effect is constant across patients and episodes.

In this scenario I assume that high values of Y_{ij} are good, which means that patients for whom $X_{PL_i} = 1$ have better outcomes if $\beta_{X_{PL}}$ is positive. The purpose of X_{PL_i} in this scenario is to allow the non-enrolment to be differential based on the episode 1 outcome; this will be explained further below.

Let $R_{i2} = 1$ indicate that patient i is enrolled in the trial for episode 2, and $R_{i2} = 0$ denoting non-enrolment for the second episode. Note that $R_{i2} = 0$ if $M_i = 1$.

In this scenario, the probability of re-enrolment depends on two factors: treatment allocation in episode 1 (Z_{i1}) and the value of X_{PL_i}. I use X_{PL_i} as a marker of the

patient's outcome in episode 1; for instance, if $\beta_{X_{PL}}$ is positive and if patients with $X_{PL_i} = 1$ are more likely to re-enrol for episode 2, this indicates that patients with better outcomes in episode 1 are more likely to re-enrol in the trial for their 2nd episode. Incorporating X_{PL_i} into data generating model 3.11 and into the model for the probability of re-enrolment allows the non-enrolment to be differential between treatment arms.

Let $\pi = P(X_{PL_i} = 1)$ (i.e. π denotes the probability that $X_{PL_i} = 1$ for patient i), and let $p_{zx} = P(R_{i2} = 1|Z_{i1} = z_{i1}, X_{PL_i} = x_{PL_i})$ (i.e. p_{zx} denotes the probability of being re-enrolled for a second episode given Z_{i1} and X_{PL_i}). So, for example, $p_{00} = P(R_{i2} = 1|Z_{i1} = 0, X_{PL_i} = 0)$, and $p_{01} = P(R_{i2} = 1|Z_{i1} = 0, X_{PL_i} = 1)$.

For simplicity, I will assume that $\pi = 0.5$, and that $p_{00} = p_{11}$ and $p_{01} = p_{10}$. This has two main implications. The first is that an equal number of patients from both episode 1 treatment arms will re-enrol for a second episode (i.e. $P(R_{i2} = 1|Z_{i1} = 0) = P(R_{i2} = 1|Z_{i1} = 1)$). Second, different types of patients will re-enrol from each episode 1 treatment group; for example, sicker patients in the episode 1 control arm and healthier patients in the episode 1 intervention arm have the same probability of re-enrolling for a 2nd episode, and vice versa.

Note that in this scenario the following (asymptotic) relationship holds for p (where p represents the proportion of patients enrolled in the trial for two episodes): $p = P(R_{i2} = 1|Z_{i1} = 0) = P(R_{i2} = 1|Z_{i1} = 1) = \frac{(p_{00})(1-\pi)+p_{01}\pi}{(1-\pi)+\pi} = p_{00} - \pi p_{00} + \pi p_{01} = \frac{1}{2}(p_{00} + p_{01})$. This will be used below.

The number of episodes and expected outcomes for each combination of treatment sequence and X_{PL_i} value can be seen in table 3.8. For simplicity, I have substituted p_{10} and p_{11} with p_{01} and p_{00} respectively, as these are equal in this scenario.

Now, collapsing over X_{PL_i} in table 3.8 leads to table 3.9.

The number of observations for each treatment sequence in table 3.9 is obtained by adding together the numbers for each of $X_{PL_i} = 0$ and $X_{PL_i} = 1$. The expected values of outcomes in table 3.9 are obtained by a weighted average; this is shown in appendix A.

Then, plugging the values from table 3.9 into the formulas from section 3.3.1, we obtain:

$$E\left(\hat{\beta}_{B_1}\right) = \frac{\frac{N_T}{2}(1-p)}{\frac{N_T}{2}(1-p)}\left(\beta_{trt} + \beta_{X_{PL}}\frac{(1-p_{00})}{2(1-p)} - \beta_{X_{PL}}\frac{(1-p_{01})}{2(1-p)}\right) = \beta_{trt} + \beta_{X_{PL}}\frac{(p_{01}-p_{00})}{2(1-p)}$$

And:

Table 3.8: Number of observations and expected values in each combination of treatment sequence and value of X_{PL_i} in *S6: Constant treatment effect, differential non-enrolment based on outcome in previous episode.* E1=episode 1, E2=episode 2

X_{PL_i}	**Treatment allocation**		**Number of observations**		$E(Y_{ij})$	
	E1	**E2**	**E1**	**E2**	**E1**	**E2**
0	0	–	$\frac{N_T}{4}(1-p_{00})$	–	0	–
1	0	–	$\frac{N_T}{4}(1-p_{01})$	–	$\beta_{X_{PL}}$	–
0	0	1	$\frac{N_T}{8}p_{00}$	$\frac{N_T}{8}p_{00}$	0	β_{trt}
1	0	1	$\frac{N_T}{8}p_{01}$	$\frac{N_T}{8}p_{01}$	$\beta_{X_{PL}}$	$\beta_{trt}+\beta_{X_{PL}}$
0	0	0	$\frac{N_T}{8}p_{00}$	$\frac{N_T}{8}p_{00}$	0	0
1	0	0	$\frac{N_T}{8}p_{01}$	$\frac{N_T}{8}p_{01}$	$\beta_{X_{PL}}$	$\beta_{X_{PL}}$
0	1	–	$\frac{N_T}{4}(1-p_{00})$	–	β_{trt}	–
1	1	–	$\frac{N_T}{4}(1-p_{01})$	–	$\beta_{trt}+\beta_{X_{PL}}$	–
0	1	0	$\frac{N_T}{8}p_{01}$	$\frac{N_T}{8}p_{01}$	β_{trt}	0
1	1	0	$\frac{N_T}{8}p_{00}$	$\frac{N_T}{8}p_{00}$	$\beta_{trt}+\beta_{X_{PL}}$	$\beta_{X_{PL}}$
0	1	1	$\frac{N_T}{8}p_{01}$	$\frac{N_T}{8}p_{01}$	β_{trt}	β_{trt}
1	1	1	$\frac{N_T}{8}p_{00}$	$\frac{N_T}{8}p_{00}$	$\beta_{trt}+\beta_{X_{PL}}$	$\beta_{trt}+\beta_{X_{PL}}$

Table 3.9: Number of observations and expected values in each treatment sequence in *S6: Constant treatment effect, differential non-enrolment based on outcome in previous episode.* E1=episode 1, E2=episode 2

Treatment allocation		**Number of observations**		$E(Y_{ij})$	
E1	**E2**	**E1**	**E2**	**E1**	**E2**
0	–	$\frac{N_T}{2}(1-p)$	–	$\beta_{X_{PL}}\frac{(1-p_{01})}{2(1-p)}$	–
0	1	$\frac{N_T}{4}p$	$\frac{N_T}{4}p$	$\beta_{X_{PL}}\frac{p_{01}}{2p}$	$\beta_{trt}+\beta_{X_{PL}}\frac{p_{01}}{2p}$
0	0	$\frac{N_T}{4}p$	$\frac{N_T}{4}p$	$\beta_{X_{PL}}\frac{p_{01}}{2p}$	$\beta_{X_{PL}}\frac{p_{01}}{2p}$
1	–	$\frac{N_T}{2}(1-p)$	–	$\beta_{trt}+\beta_{X_{PL}}\frac{(1-p_{00})}{2(1-p)}$	–
1	0	$\frac{N_T}{4}p$	$\frac{N_T}{4}p$	$\beta_{trt}+\beta_{X_{PL}}\frac{p_{00}}{2p}$	$\beta_{X_{PL}}\frac{p_{00}}{2p}$
1	1	$\frac{N_T}{4}p$	$\frac{N_T}{4}p$	$\beta_{trt}+\beta_{X_{PL}}\frac{p_{00}}{2p}$	$\beta_{trt}+\beta_{X_{PL}}\frac{p_{00}}{2p}$

$$E\left(\hat{\beta}_{B_2}\right) = \frac{1}{\frac{N_T p}{2}}\left(\frac{2N_T p}{4}\left(\beta_{trt} + \beta_{X_{PL}}\frac{p_{00}}{2p}\right) - \frac{2N_T p}{4}\left(\beta_{X_{PL}}\frac{p_{01}}{2p}\right)\right) = \beta_{trt} + \beta_{X_{PL}}\frac{(p_{00} - p_{01})}{2p}$$

And:

$$E\left(\hat{\beta}_w\right) = \frac{1}{\frac{N_T p}{2}}\left(\frac{N_T p}{4}\left(\beta_{trt} + \beta_{X_{PL}}\frac{p_{00}}{2p} - \beta_{X_{PL}}\frac{p_{00}}{2p}\right) - \frac{N_T p}{4}\left(\beta_{X_{PL}}\frac{p_{01}}{2p} - \beta_{trt} - \beta_{X_{PL}}\frac{p_{01}}{2p}\right)\right) = \beta_{trt}$$

Therefore:

$$E\left(\hat{\beta}_E^{AB}\right) = \frac{(1-p)\left(\beta_{trt} + \beta_{X_{PL}}\frac{(p_{01}-p_{00})}{2(1-p)}\right) + p\left(\beta_{trt} + \beta_{X_{PL}}\frac{(p_{00}-p_{01})}{2p}\right) + p\beta_{trt}}{1+p} = \frac{\beta_{trt}\left((1-p)+p+p\right) + \beta_{X_{PL}}\left(\frac{(1-p)(p_{01}-p_{00})}{2(1-p)} + \frac{p(p_{00}-p_{01})}{2p}\right)}{1+p} = \frac{\beta_{trt}(1+p) + \frac{\beta_{X_{PL}}}{2}(p_{01} - p_{00} + p_{00} - p_{01})}{1+p} = \beta_{trt}$$

And:

$$E\left(\hat{\beta}_P^{AB}\right) = (1-p)\left(\beta_{trt} + \beta_{X_{PL}}\frac{(p_{01}-p_{00})}{2(1-p)}\right) + \frac{p}{2}\left(\beta_{trt} + \beta_{X_{PL}}\frac{(p_{00}-p_{01})}{2p}\right) + \frac{p}{2}\beta_{trt} = \beta_{trt}\left(1 - p + \frac{p}{2} + \frac{p}{2}\right) + \beta_{X_{PL}}\left((1-p)\frac{(p_{01}-p_{00})}{2(1-p)} + \frac{p}{2}\frac{(p_{00}-p_{01})}{2p}\right) = \beta_{trt} + \beta_{X_{PL}}\frac{(p_{01}-p_{00})}{4}$$

3.3.11 S7: Constant treatment effect, differential non-enrolment based on expected outcome in current episode

In this scenario I consider a situation where non-enrolment is differential between treatment arms depending on their expected outcome in the second episode (i.e. their baseline prognosis at episode 2); patients who received the intervention in episode 1 and have a good baseline prognosis at episode 2 have the same probability of re-enrolling as patients who received control in episode 1 and had a bad prognosis, and vice versa.

Consider the following data generating mechanism:

$$Y_{ij} = \alpha + \beta_{trt} Z_{ij} + \beta_{X_{EL}} X_{EL_{ij}} + \mu_i + \varepsilon_{ij} \tag{3.12}$$

where $X_{EL_{ij}}$ is an unobserved binary episode-level variable (i.e. it can vary across episodes). I use the subscript EL to denote 'episode-level'. This data generating mechanism is equivalent to model 1.3, but with the addition of $X_{EL_{ij}}$ (and β_{trt} in place of β), and so the treatment effect is constant across patients and episodes.

As in the previous scenario, high values of Y_{ij} are good, which means that patients for whom $X_{EL_{ij}} = 1$ have better outcomes in episode j if $\beta_{X_{EL}}$ is positive. The purpose of $X_{EL_{ij}}$ in this scenario is to allow the non-enrolment to be differential based on the expected outcome at episode 2 (i.e. based on the baseline prognosis at episode 2); this will be explained further below.

In this scenario, I will redefine π as $\pi = P\left(X_{EL_{ij}} = 1\right)$, and set this to $\pi = 0.5$. Similarly I will redefine p_{zx} as $p_{zx} = P\left(R_{i2} = 1 | Z_{i1} = z_{i1}, X_{EL_{i2}} = x_{EL_{i2}}\right)$ (so p_{zx} represents the probability of re-enrolment for a second episode based on the patient's episode 1 allocation and their episode 2 value of $X_{EL_{ij}}$). So, for example, $p_{00} = P\left(R_{i2} = 1 | Z_{i1} = 0, X_{EL_{i2}} = 0\right)$, and $p_{01} = P\left(R_{i2} = 1 | Z_{i1} = 0, X_{EL_{i2}} = 1\right)$.

As before, I will assume that $p_{00} = p_{11}$ and $p_{01} = p_{10}$. This, combined with setting $\pi = 0.5$ above, implies that an equal number of patients from both episode 1 treatment arms will re-enrol for a second episode (i.e. $P\left(R_{i2} = 1 | Z_{i1} = 0\right) = P\left(R_{i2} = 1 | Z_{i1} = 1\right)$), and that different types of patients will re-enrol from each episode 1 treatment group (for example, patients with a worse prognosis at episode 2 who received control in episode 1 have the same probability of re-enrolling as patients with a better prognosis at episode 2 who received intervention in episode 1, and vice versa).

As in the previous scenario, note that $p = P\left(R_{i2} = 1 | Z_{i1} = 0\right) = P\left(R_{i2} = 1 | Z_{i1} = 1\right) = \frac{(p_{00})(1-\pi)+p_{01}\pi}{(1-\pi)+\pi} = p_{00} - \pi p_{00} + \pi p_{01} = \frac{1}{2}\left(p_{00} + p_{01}\right)$.

In this scenario, I will assume for simplicity that all patients experienced two episodes (i.e. $M_i = 2$ for all patients), but that some of these patients were not re-enrolled for their 2nd episode. Therefore, even patients who were enrolled for only a single episode still have a value for $X_{EL_{i2}}$.

The number of episodes and expected outcomes for each sequence can be seen in table 3.10; note that for simplicity I have replaced p_{10} and p_{11} with p_{01} and p_{00} respectively, as these are equal in this scenario.

Now, collapsing over $X_{EL_{ij}}$ in table 3.10 leads to table 3.11 below.

The expected values of outcomes in table 3.11 are obtained by a weighted average; this is shown in appendix A.

Table 3.10: Number of observations and expected values in each combination of treatment sequence and value of X_{EL_i} in *S7: Constant treatment effect, differential non-enrolment based on expected outcome in current episode.* E1=episode 1, E2=episode 2

X_{EL_i}		**Treatment allocation**		**Number of observations**		$E(Y_{ij})$	
E1	**E2**	**E1**	**E2**	**E1**	**E2**	**E1**	**E2**
0	0	0	–	$\frac{N_T}{8}(1-p_{00})$	–	0	–
0	1	0	–	$\frac{N_T}{8}(1-p_{01})$	–	0	–
1	0	0	–	$\frac{N_T}{8}(1-p_{00})$	–	$\beta_{X_{EL}}$	–
1	1	0	–	$\frac{N_T}{8}(1-p_{01})$	–	$\beta_{X_{EL}}$	–
0	0	0	1	$\frac{N_T}{16}p_{00}$	$\frac{N_T}{16}p_{00}$	0	β_{trt}
0	1	0	1	$\frac{N_T}{16}p_{01}$	$\frac{N_T}{16}p_{01}$	0	$\beta_{trt}+\beta_{X_{EL}}$
1	0	0	1	$\frac{N_T}{16}p_{00}$	$\frac{N_T}{16}p_{00}$	$\beta_{X_{EL}}$	β_{trt}
1	1	0	1	$\frac{N_T}{16}p_{01}$	$\frac{N_T}{16}p_{01}$	$\beta_{X_{EL}}$	$\beta_{trt}+\beta_{X_{EL}}$
0	0	0	0	$\frac{N_T}{16}p_{00}$	$\frac{N_T}{16}p_{00}$	0	0
0	1	0	0	$\frac{N_T}{16}p_{01}$	$\frac{N_T}{16}p_{01}$	0	$\beta_{X_{EL}}$
1	0	0	0	$\frac{N_T}{16}p_{00}$	$\frac{N_T}{16}p_{00}$	$\beta_{X_{EL}}$	0
1	1	0	0	$\frac{N_T}{16}p_{01}$	$\frac{N_T}{16}p_{01}$	$\beta_{X_{EL}}$	$\beta_{X_{EL}}$
0	0	1	–	$\frac{N_T}{8}(1-p_{01})$	–	β_{trt}	–
0	1	1	–	$\frac{N_T}{8}(1-p_{00})$	–	β_{trt}	–
1	0	1	–	$\frac{N_T}{8}(1-p_{01})$	–	$\beta_{trt}+\beta_{X_{EL}}$	–
1	1	1	–	$\frac{N_T}{8}(1-p_{00})$	–	$\beta_{trt}+\beta_{X_{EL}}$	–
0	0	1	0	$\frac{N_T}{16}p_{01}$	$\frac{N_T}{16}p_{01}$	β_{trt}	0
0	1	1	0	$\frac{N_T}{16}p_{00}$	$\frac{N_T}{16}p_{00}$	β_{trt}	$\beta_{X_{EL}}$
1	0	1	0	$\frac{N_T}{16}p_{01}$	$\frac{N_T}{16}p_{01}$	$\beta_{trt}+\beta_{X_{EL}}$	0
1	1	1	0	$\frac{N_T}{16}p_{00}$	$\frac{N_T}{16}p_{00}$	$\beta_{trt}+\beta_{X_{EL}}$	$\beta_{X_{EL}}$
0	0	1	1	$\frac{N_T}{16}p_{01}$	$\frac{N_T}{16}p_{01}$	β_{trt}	β_{trt}
0	1	1	1	$\frac{N_T}{16}p_{00}$	$\frac{N_T}{16}p_{00}$	β_{trt}	$\beta_{trt}+\beta_{X_{EL}}$
1	0	1	1	$\frac{N_T}{16}p_{01}$	$\frac{N_T}{16}p_{01}$	$\beta_{trt}+\beta_{X_{EL}}$	β_{trt}
1	1	1	1	$\frac{N_T}{16}p_{00}$	$\frac{N_T}{16}p_{00}$	$\beta_{trt}+\beta_{X_{EL}}$	$\beta_{trt}+\beta_{X_{EL}}$

Table 3.11: Number of observations and expected values in each treatment sequence in *S7: Constant treatment effect, differential non-enrolment based on expected outcome in current episode*. E1=episode 1, E2=episode 2

Treatment allocation		**Number of observations**		$E(Y_{ij})$	
E1	**E2**	**E1**	**E2**	**E1**	**E2**
0	–	$\frac{N_T}{2}(1-p)$	–	$\frac{\beta_{X_{EL}}}{2}$	–
0	1	$\frac{N_T}{4}p$	$\frac{N_T}{4}p$	$\frac{\beta_{X_{EL}}}{2}$	$\beta_{trt} + \beta_{X_{EL}}\frac{p_{01}}{p_{00}+p_{01}}$
0	0	$\frac{N_T}{4}p$	$\frac{N_T}{4}p$	$\frac{\beta_{X_{EL}}}{2}$	$\beta_{X_{EL}}\frac{p_{01}}{p_{00}+p_{01}}$
1	–	$\frac{N_T}{2}(1-p)$	–	$\beta_{trt} + \frac{\beta_{X_{EL}}}{2}$	–
1	0	$\frac{N_T}{4}p$	$\frac{N_T}{4}p$	$\beta_{trt} + \frac{\beta_{X_{EL}}}{2}$	$\beta_{trt} + \beta_{X_{EL}}\frac{p_{00}}{p_{00}+p_{01}}$
1	1	$\frac{N_T}{4}p$	$\frac{N_T}{4}p$	$\beta_{trt} + \frac{\beta_{X_{EL}}}{2}$	$\beta_{trt} + \beta_{X_{EL}}\frac{p_{00}}{p_{00}+p_{01}}$

Then, plugging the values from table 3.11 into the formulas from section 3.3.1, we obtain:

$$E\left(\hat{\beta}_{B_1}\right) = \frac{\frac{N_T}{2}(1-p)}{\frac{N_T}{2}(1-p)}\left(\beta_{trt} + \frac{\beta_{X_{EL}}}{2} - \frac{\beta_{X_{EL}}}{2}\right) = \beta_{trt}$$

$$E\left(\hat{\beta}_{B_2}\right) = \frac{\frac{N_T p}{4}}{\frac{N_T p}{2}}\left(\beta_{trt} + \frac{\beta_{X_{EL}}}{2} + \beta_{trt} + \beta_{X_{EL}}\frac{p_{00}}{p_{00}+p_{01}} - \frac{\beta_{X_{EL}}}{2} - \beta_{X_{EL}}\frac{p_{01}}{p_{00}+p_{01}}\right)$$

$$= \frac{1}{2}\left(2\beta_{trt} + \beta_{X_{EL}}\left(\frac{p_{00}-p_{01}}{p_{00}+p_{01}}\right)\right) = \beta_{trt} + \frac{\beta_{X_{EL}}}{4}\left(\frac{p_{00}-p_{01}}{p}\right)$$

And:

$$E\left(\hat{\beta}_{w}\right) = \frac{1}{\frac{N_T p}{2}}\left(\frac{N_T p}{4}\left(\beta_{trt} + \frac{\beta_{X_{EL}}}{2} - \beta_{X_{EL}}\frac{p_{00}}{p_{00}+p_{01}} - \frac{\beta_{X_{EL}}}{2} + \beta_{trt} + \beta_{X_{EL}}\frac{p_{01}}{p_{00}+p_{01}}\right)\right)$$

$$= \frac{1}{2}\left(2\beta_{trt} - \beta_{X_{EL}}\left(\frac{p_{00}-p_{01}}{p_{00}+p_{01}}\right)\right) = \beta_{trt} - \frac{\beta_{X_{EL}}}{4}\left(\frac{p_{00}-p_{01}}{p}\right)$$

Therefore:

$$E\left(\hat{\beta}_{E}^{AB}\right) = \frac{(1-p)\beta_{trt} + p\left(\beta_{trt} + \frac{\beta_{X_{EL}}}{4}\left(\frac{p_{00}-p_{01}}{p}\right)\right) + p\left(\beta_{trt} - \frac{\beta_{X_{EL}}}{4}\left(\frac{p_{00}-p_{01}}{p}\right)\right)}{1+p}$$

$$= \frac{(1-p+p+p)\beta_{trt} + \frac{p\beta_{X_{EL}}}{4}\left(\frac{p_{00}-p_{01}-p_{00}+p_{01}}{p}\right)}{1+p} = \beta_{trt}$$

And:

$$
\begin{aligned}
&E\left(\hat{\beta}_P^{AB}\right) = \\
&(1-p)\,\beta_{trt} + \frac{p}{2}\left(\beta_{trt} + \frac{\beta_{X_{EL}}}{4}\left(\frac{p_{00}-p_{01}}{p}\right)\right) + \frac{p}{2}\left(\beta_{trt} - \frac{\beta_{X_{EL}}}{4}\left(\frac{p_{00}-p_{01}}{p}\right)\right) \\
&\qquad = \left(1-p+\frac{p}{2}+\frac{p}{2}\right)\beta_{trt} + \frac{p\beta_{X_{EL}}}{8}\left(\frac{p_{00}-p_{01}-p_{00}+p_{01}}{p}\right) = \beta_{trt}
\end{aligned}
$$

3.3.12 Summary of mathematical results

A summary of results for the per-episode added-benefit and the per-patient added-benefit estimators are available in tables 3.12 and 3.13.

From table 3.12, we can see that the per-episode added-benefit estimator is unbiased in all settings considered.

From table 3.13, we can see that the per-patient added-benefit estimator is unbiased, except when there is differential non-enrolment based on the episode 1 outcome. The reason for this bias may be that the association between Y_{ij} and M_i is different between treatment groups in this scenario. Consider the following example; imagine a trial where there is no difference between treatment groups (i.e. a null treatment effect). However, the association between Y_{ij} and M_i differs between groups as follows; good values of Y_{ij} in the control group tend to be associated with episodes where $M_i = 1$, whereas good values of Y_{ij} in the intervention group tend to be associated with episodes where $M_i = 2$. Then, when the weights $W_i = \frac{1}{M_i}$ are applied to outcomes Y_{ij}, good values of Y_{ij} in the control group are down weighted less than good values in the intervention group, which would cause a difference between treatment groups even though none exists.

This is what occurred in this scenario. Patients in the intervention group with good outcomes in episode 1 were less likely to re-enrol for episode 2. This meant that episodes with good outcomes in the intervention arm were more likely to have $M_i = 1$ than in the control group. Likewise, episodes with good outcomes in the control group were more likely to have $M_i = 2$ than in the intervention group (formally, there was a higher proportion of episodes where $M_i = 1$ and $X_{PL_i} = 1$ in the intervention arm than in the control, and a higher proportion where $M_i = 2$ and $X_{PL_i} = 0$ in the control arm). This lead to good outcomes being weighted differently between the treatment groups, which caused bias in the estimated treatment effect.

3.4 Simulation study

I conducted a simulation study to evaluate the independence estimators discussed in the previous section [27]. The main purpose of this simulation study was to evaluate the policy-benefit estimators (which were not included in the mathematical derivations in

Table 3.12: Summary of mathematical derivations for the per-episode added-benefit independence estimator

Scenario	$E\left(\hat{\beta}_E^{AB}\right)$	**Estimand (per-episode added-benefit)**
S1 – Constant treatment effect	β	β
S2 – Treatment effect varies across episode	$\frac{\beta_1+p\beta_2}{1+p}$	$\frac{\beta_1+p\beta_2}{1+p}$
S3 – Treatment effect varies across value of M_i	$\frac{(1-p)\beta_1+2p\beta_2}{(1+p)}$	$\frac{(1-p)\beta_1+2p\beta_2}{(1+p)}$
S4 – Treatment effect carries forward into second episode	β	β
S5 – Treatment becomes less effective on re-use	$\beta + \frac{p\delta}{2(1+p)}$	$\beta + \frac{p\delta}{2(1+p)}$
S6 – Constant treatment effect, differential non-enrolment based on outcome in previous episode	β_{trt}	β_{trt}
S7 – Constant treatment effect, differential non-enrolment based on expected outcome in current episode	β_{trt}	β_{trt}

the previous section), and to evaluate all estimators under more realistic data generating mechanisms than considered in the mathematical derivations in the previous section (e.g. with smaller sample sizes, and more complex data generating or non-enrolment mechanisms).

The main aim of this simulation study was to evaluate bias, though a secondary aim was to evaluate the coverage of 95% confidence intervals in settings where estimators were unbiased. I did not evaluate the precision of the different estimators, as each estimator addressed a different question and so precision is less relevant in deciding between them.

This simulation study focussed on a setting where patients were enrolled in a trial for a maximum of two episodes. For most scenarios, I chose parameter values that are larger than those we would expect to see in practice; this was to ensure that if estimators were biased in any scenarios, I would be able to identify it.

I describe the estimands, methods of analysis, and performance measures used across all simulation scenarios in section 3.4.1. In sections 3.4.2, 3.4.3 and 3.4.4 I describe the data generating models and results of the simulation studies.

All simulations were conducted using Stata v15.1.

Table 3.13: Summary of mathematical derivations for the per-patient added-benefit independence estimator

Scenario	$E\left(\hat{\beta}_P^{AB}\right)$	**Estimand (per-patient added-benefit)**
S1 – Constant treatment effect	β	β
S2 – Treatment effect varies across episode	$\left(1-\frac{p}{2}\right)\beta_1 + \frac{p}{2}\beta_2$	$\left(1-\frac{p}{2}\right)\beta_1 + \frac{p}{2}\beta_2$
S3 – Treatment effect varies across value of M_i	$(1-p)\,\beta_1 + p\beta_2$	$(1-p)\,\beta_1 + p\beta_2$
S4 – Treatment effect carries forward into second episode	β	β
S5 – Treatment becomes less effective on re-use	$\beta + \frac{p\delta}{4}$	$\beta + \frac{p\delta}{4}$
S6 – Constant treatment effect, differential non-enrolment based on outcome in previous episode	$\beta_{trt} + \beta_{X_{PL}} \frac{(p_{01}-p_{00})}{4}$	β_{trt}
S7 – Constant treatment effect, differential non-enrolment based on expected outcome in current episode	β_{trt}	β_{trt}

3.4.1 Estimands, methods of analysis, and performance measures

3.4.1.1 Estimands

For all simulation scenarios, I used the following four estimands: (a) per-episode added-benefit; (b) per-patient added-benefit; (c) per-episode policy-benefit; and (d) per-patient policy-benefit.

The values of the estimands for each simulation scenario are provided in the sections describing the data generating mechanisms below.

3.4.1.2 Methods of analysis

I implemented independence estimators corresponding to the estimands listed above. Details of how these estimators were implemented in Stata is shown in table 3.14 . I implemented policy-benefit estimators using model 3.3, which allowed the outcome and treatment effect in episode 2 to depend upon the treatment allocation in episode 1, and included an indicator for episode two. For all estimators, I used cluster-robust standard errors, with patients acting as the cluster [26].

Table 3.14: Stata code to implement independence estimators. 'y' denotes patient outcome, 'z' denotes treatment allocation, 'id' is a unique identifier for patient, 'm_i' denotes the number of episodes for which the patient is enrolled in the trial, 'z_prev' denotes the patient's treatment allocation in their previous episode (and is set to 0 if it is the patient's first episode), 'x_ep' is an indicator for episode 2, 'prop_1st_ep' and 'prop_2nd_ep' represent the proportion of episodes in the trial which are 1st and 2nd episodes respectively, and 'prop_has_1ep' and 'prop_has_2ep' denote the proportion of patients enrolled in the trial for one and two episodes respectively. In order to run the above code in Stata, 'prop_1st_ep', 'prop_2nd_ep', 'prop_has_1ep', and 'prop_has_2ep' must be saved as Stata local macros.

Estimator	**Stata code**
Added-benefit	
Per-episode	reg y z, vce(cluster id)
Per-patient	reg y z [pw=1/m_i], vce(cluster id)
Policy-benefit	
Per-episode	reg y z##z_prev x_ep, vce(cluster id) lincom /// 'prop_1st_ep'*_b[1.z] + /// 'prop_2nd_ep'*(_b[1.z]+_b[1.z_prev] + _b[1.z#1.z_prev])
Per-patient	reg y z##z_prev x_ep [pw=1/m_i], vce(cluster id) lincom /// 'prop_has_1ep'*(_b[1.z]) + /// 'prop_has_2ep'*((1/2)*(_b[1.z]) + /// (1/2)*(_b[1.z]+_b[1.z_prev] + _b[1.z#1.z_prev]))

3.4.1.3 Performance measures

My main criterion for evaluating estimators was bias. I measured bias as $E\left(\hat{\beta}\right) - \beta$, where $E\left(\hat{\beta}\right)$ represents the mean of the estimates across all simulation replications, and β represents the true value of the estimand. I compared each estimator against its corresponding estimand (i.e. $\hat{\beta}_E^{AB}$ vs. β_E^{AB}, $\hat{\beta}_P^{AB}$ vs. β_P^{AB}, etc).

I also evaluated coverage of the 95% confidence intervals. I defined coverage as the proportion of replications for which the 95% confidence interval of the estimator contained the true value of the estimand.

For each performance measure (bias, coverage) I also assessed the Monte Carlo standard error (MCSE), which provides a measure of variability for the estimated performance measure in the simulation study. I present the MCSEs as 95% confidence intervals alongside the mean bias and coverage, except in cases where this interval was too small to show up on the figure (i.e. when the width of the confidence interval was smaller than the size of the dot representing the mean bias or coverage), in which case I report the range of Monte Carlo standard errors for each performance measure

across scenarios.

I used 10,000 replications for all simulation scenarios. This was based on my previous experience conducting simulation studies (e.g. references [46], [47], and [48]) where I have found that 10,000 replications is usually more than sufficient to evaluate bias and coverage. For example, the MCSE for bias is [49]:

$$MCSE = \sqrt{\frac{V\left(\hat{\beta}\right)}{reps}}$$

where $V\left(\hat{\beta}\right)$ is the variance of the estimator and *reps* is the number of replications. The variance of the per-episode added-benefit estimator under data generating model 1.3 is:

$$V\left(\hat{\beta}_E^{AB}\right) = \frac{2\sigma^2}{M_T}$$

In the first simulation scenario (described below), $\sigma^2 = 10$ and $M_T = 450$, which means $V\left(\hat{\beta}_E^{AB}\right) = 0.044$. This implies the MCSE for bias is 0.002, which is sufficiently small for the purpose of this study. In some other scenarios the MCSE will be larger, because M_T is lower or σ^2 is larger, however the MCSE is sufficiently small in all scenarios to allow us to identify whether estimators are truly biased or not.

Similarly, the MCSE for coverage is [49]:

$$MCSE = \sqrt{\frac{Coverage(1 - Coverage)}{reps}}$$

If the true coverage is 95%, this implies the MCSE will be 0.2%, which is sufficiently small to identify whether estimators have correct coverage or not.

3.4.2 Simulation study 1: patients enrolled for all episodes they experience

3.4.2.1 Data generating methods

This simulation study is broken into three parts; simulation study 1, 2a, and 2b. In this section I describe the data generating model for simulation study 1, then describe the results in section 3.4.2.2. I describe simulation studies 2a and 2b in sections 3.4.3 and 3.4.4.

Simulation study 1 is based on a trial of 300 patients; 150 patients experience one episode during the trial period, and 150 experience two episodes (i.e. $N_T = 300$, $M_T = 450$, $M_{T(1)} = 150$, and $M_{T(2)} = 150$).

The main purpose of this simulation study is to evaluate estimators in the setting where patients are enrolled for all episodes they experience; that is, the 150 patients

Table 3.15: Description of variables used in simulation study 1 (chapter 3)

Variable	**Description**	**Method of generation**
Y_{ij}	Continuous outcome for patient i in episode j	Generated based on model 3.13
Z_{ij}	Treatment allocation (0=control, 1=intervention) for patient i in episode j	Bernoulli random variable with probability of 0.5 (implying simple randomisation)
$X_{ep_{ij}}$	Indicator for episode 2	NA
X_{M_i}	Indicator for number of episodes patient experiences (0=1 episode, 1=2 episodes); equivalent to M_i	NA
$Z_{i,j-1}$	Treatment allocation for patient i in episode $j-1$; equal to 0 for episode 1	NA
X_{PL_i}	Unobserved patient-level binary covariate, which is constant across episodes	Bernoulli random variable with probability of 0.5
$X_{EL_{ij}}$	Unobserved episode-level binary covariate, which can vary across episodes	Bernoulli random variable with probability of 0.5
μ_i	Random intercept for patient i	$\sim N(0, \sigma^2_\mu)$
ε_{ij}	Random error term for episode j in patient i	$\sim N(0, \sigma^2_\varepsilon)$

who experienced two episodes were enrolled in the trial for both episodes (i.e. there are no patients who do not re-enrol for their 2nd episode).

I consider six different data generating mechanisms (described further below); all were based on the following general model for a continuous outcome:

$$Y_{ij} = \alpha + \beta_{trt} Z_{ij} + \beta_{ep} X_{ep_{ij}} + \beta_M X_{M_i} + \beta_{TRTxEP} Z_{ij} X_{ep_{ij}} + \beta_{TRTxM} Z_{ij} X_{M_i} + \gamma Z_{i,j-1} + \delta Z_{ij} Z_{i,j-1} + \mu_i + \varepsilon_{ij} \quad (3.13)$$

where $X_{ep_{ij}}$ is an indicator variable for episode 2, and X_{M_i} is an indicator variable for patients with $M_i = 2$. A description of the variables in this model are given in table 3.15 (this table also contains some variables which are not in equation 3.13, but are used in the simulation study in section 3.4.3). Higher values of the outcome are better.

The parameter α is an intercept, β_{ep} and β_M are the effects of episode 2 and patient type (whether they experience 1 vs. 2 episodes) on outcome, and β_{trt}, β_{TRTxEP}, β_{TRTxM}, γ, and δ are components of the treatment effect (e.g. β_{TRTxEP} is the inter-

action between treatment allocation and episode number, β_{TRTxM} is the interaction between treatment and patient type, and δ is the interaction between treatment allocation in the current episode and allocation in the previous episode).

In this study, I considered six different treatment effect mechanisms. This involved varying the parameters that define the treatment effect (β_{trt}, β_{TRTxEP}, β_{TRTxM}, γ, and δ); values of these parameters for each scenario are shown in table 3.16, along with the values of the four estimands for each scenario. For each scenario, I set $\alpha = 0$, $\beta_{trt} = 3$, $\beta_{ep} = 1$, $\beta_M = 1$, $\sigma^2_\mu = 5$ and $\sigma^2_\varepsilon = 5$. I generated μ_i and ε_{ij} independently; based on the chosen variances, the intraclass correlation between episodes from the same patient is 0.50 (conditional on the other variables in the data generating model). I chose the ICC value 0.50 arbitrarily, as the ICC should not influence the level of bias, and a value of 0.50 seems plausible when patient is the cluster (e.g. a recent review of correlations between outcomes measured at different time points within the same patient found the mean correlation was 0.50, SD 0.15 [50]).

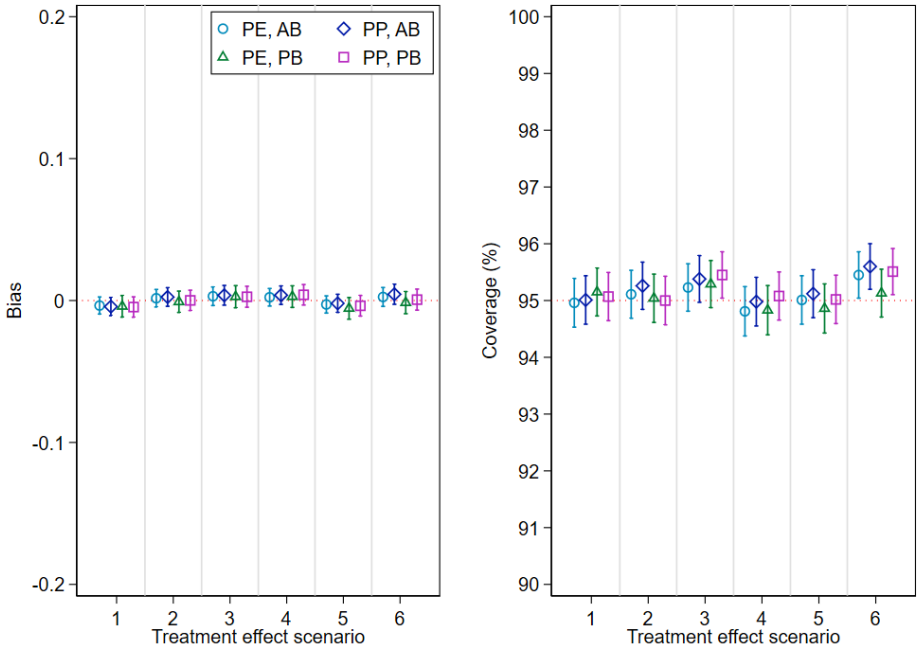
I considered the following treatment effect mechanisms:

1. Constant treatment effect: the treatment effect is the same (β_{trt}) across all episodes and patients
2. Treatment effect varies across episode: the treatment effect is different in the 1st episode (β_{trt}) vs. in the 2nd episode ($\beta_{trt} + \beta_{TRTxEP}$)
3. Treatment effect varies across patients with different values of M_i: the treatment effect is different in patients who experience 1 episode (β_{trt}) vs. those who experience 2 episodes ($\beta_{trt} + \beta_{TRTxM}$).
4. Treatment effect carries forward into the 2nd episode: patients who receive intervention in the first episode have better outcomes in their 2nd episode (by the amount γ)
5. Treatment becomes less effective on re-use: patients receiving the intervention for the 1st time have a different treatment effect (β_{trt}) than those receiving the intervention for the 2nd time ($\beta_{trt} + \delta$)
6. Treatment effect varies across episodes, across patients with different values of M_i, carries forward, and becomes less effective on re-use: the treatment effect is β_{trt} for patients who experience one episode. For patients who experience two episodes, the treatment effect is $\beta_M + \beta_{TRTxM}$ in the 1st episode, $\beta_M + \beta_{TRTxM} + \beta_{TRTxEP}$ in the 2nd episode for patients receiving the intervention for the first time (i.e. who received control in their 1st episode), and $\beta_M + \beta_{TRTxM} + \beta_{TRTxEP} + \delta$ in the 2nd episode for patients receiving the intervention for the 2nd time (i.e. received intervention in their 1st episode). Patients who receive the intervention in the first episode also have better outcomes in their 2nd episode, by the amount γ.

Table 3.16: Simulation parameters and estimands for different scenarios in simulation study 1. For all scenarios, I set $\alpha = 0$, $\beta_{trt} = 3$, $\beta_{ep} = 1$, $\beta_M = 1$, $\sigma^2_\mu = 5$ and $\sigma^2_\varepsilon = 5$.

	Parameters				Estimand values			
Scenario	β_{TRTxEP}	β_{TRTxM}	γ	δ	β_E^{AB}	β_P^{AB}	β_E^{PB}	β_P^{PB}
Scenario 1: Constant treatment effect	0	0	0	0	3	3	3	3
Scenario 2: Treatment effect varies across episode	1.5	0	0	0	3.5	3.38	3.5	3.38
Scenario 3: Treatment effect varies across patients with different values of M_i	0	3	0	0	5	4.5	5	4.5
Scenario 4: Treatment effect carries forward	0	0	1	0	3	3	3.33	3.25
Scenario 5: Treatment becomes less effective on re-use	0	0	0	-3	2.5	2.63	2	2.25
Scenario 6: Treatment effect varies across episodes, across patients with different values of M_i, carries forward, and becomes less effective on re-use	1.5	3	1	-3	5	4.5	4.83	4.38

Figure 3.1: Bias and coverage of independence estimators in simulation study 1. Error bars are 95% confidence intervals based on Monte Carlo standard errors. Scenario 1: Constant treatment effect. Scenario 2: Treatment effect varies across episode. Scenario 3: Treatment effect varies across patients with different values of M_i. Scenario 4: Treatment effect carries forward. Scenario 5: Treatment becomes less effective on re-use. Scenario 6: Treatment effect varies across episodes, across patients with different values of M_i, carries forward, and becomes less effective on re-use.



3.4.2.2 Results

Results are shown in figure 3.1. All estimators were unbiased and provided close to nominal coverage in all scenarios.

3.4.3 Simulation study 2a: some patients do not re-enrol for their 2nd episode

3.4.3.1 Data generating methods

The main purpose of this simulation study is to evaluate estimators when some of the patients who experience two episodes do not re-enrol in the trial for their second episode. For example, this may occur if patients find the trial procedures, such as number of follow-up visits, too burdensome; if they were disappointed at their treatment allocation in the first episode; or they experienced a poor outcome in their first episode. Note that this type of non-enrolment is not a form of dropout; patients only enrol in the trial for a single episode at a time, and there is no expectation that they must re-enrol for all subsequent episodes.

As before, this simulation study is based on a trial of 300 patients; 150 patients experience one episode during the trial period, and 150 experience two episodes.

All patients enrol for their first episode, but a subset of patients who experience two episodes do not re-enrol for their second episode. Therefore, $N_T = 300$ and $M_{T(1)} = 150$, however $M_{T(2)} < 150$ and $M_T < 450$; the exact values of $M_{T(2)}$ and M_T vary across simulation replications.

I simulated data by first generating outcomes for all 450 episodes (regardless of whether they were enrolled in the trial for their 2nd episode) using model 3.14 below, and then generated an indicator for each episode to denote whether it was enrolled in the trial or not using model 3.15 below. I then performed analysis only on the subset of enrolled episodes. I used six different treatment effect mechanisms (based on model 3.14 below) and five different non-enrolment mechanisms (based on model 3.15 below), leading to 6x5=30 total scenarios. The different treatment effect and non-enrolment scenarios are described below.

I generated continuous outcomes from the model:

$$Y_{ij} = \alpha + \beta_{trt} Z_{ij} + \beta_{ep} X_{ep_{ij}} + \beta_M X_{M_i} + \beta_{TRTxEP} Z_{ij} X_{ep_{ij}} + \beta_{TRTxM} Z_{ij} X_{M_i} + \gamma Z_{i,j-1} + \delta Z_{ij} Z_{i,j-1} + \beta_{X_{PL}} X_{PL_i} + \beta_{X_{EL}} X_{EL_{ij}} + \mu_i + \varepsilon_{ij} \quad (3.14)$$

This model is identical to model 3.13 from simulation study 1, except it contains two additional terms: X_{PL_i} and $X_{EL_{ij}}$. These were explained previously in sections 3.3.10 and 3.3.11; briefly, X_{PL_i} and $X_{EL_{ij}}$ are unobserved binary covariates, with X_{PL_i} being a patient-level covariate which does not vary across episodes, and $X_{EL_{ij}}$ being an episode-level covariate which can vary across episodes for the same patient; I use the subscript PL to denote 'patient-level', and EL to denote 'episode-level'. The purpose of including X_{PL_i} and $X_{EL_{ij}}$ in this model is explained below. I used positive values for $\beta_{X_{PL}}$ and $\beta_{X_{EL}}$, so that patients or episodes where $X_{PL_i} = 1$ or $X_{EL_{ij}} = 1$ have better outcomes than if X_{PL_i} or $X_{EL_{ij}}$ are 0; exact values of $\beta_{X_{PL}}$ and $\beta_{X_{EL}}$ for each scenario are shown in table 3.17.

In the subset of patients with two episodes, I generated each patient's probability of not re-enrolling for the second episode on a linear scale using the following model:

$$P(R_{i2} = 0) = \alpha^{R_2} + \gamma^{R_2} Z_{i,j-1} + \beta^{R_2}_{X_{PL}} X_{PL_i} + \beta^{R_2}_{X_{EL}} X_{EL_{i2}} + \delta^{R_2}_{Xpl} Z_{i,j-1} X_{PL_i} + \delta^{R_2}_{Xel} Z_{i,j-1} X_{EL_{i2}} \quad (3.15)$$

where R_{ij} denotes whether patient i was enrolled for their jth episode (0=not enrolled, 1=enrolled). Note that $R_{i2} = 0$ for patients who only experience one episode, and $R_{i1} = 1$ for all patients. I use the superscript R_2 for parameters to indicate that these parameters relate to the probability of not being re-enrolled for the 2nd episode. I set $\alpha^{R_2} = 0.05$ and $\gamma^{R_2} = 0.10$ for all scenarios. This implies that all patients have

a non-zero probability of not re-enrolling for their second episode, and that patients who received intervention in episode 1 are more likely to not re-enrol than those in the control group (irrespective of X_{PL_i} and $X_{EL_{i2}}$). Values for other parameters are shown in table 3.17.

In the models above, I use X_{PL_i} as a marker of the patient's outcome in episode 1, and $X_{EL_{i2}}$ as a marker for the patient's expected outcome in episode 2; that is, larger values of $\beta^{R_2}_{X_{PL}}$ denote that patients with better outcomes in episode 1 are less likely to re-enrol in the trial for their 2nd episode, and larger values of $\beta^{R_2}_{X_{EL}}$ denote that patients with better expected outcomes in episode 2 are less likely to re-enrol for that episode.

As stated above, I used six treatment effect mechanisms and five non-enrolment mechanisms. I used the same six treatment effect mechanisms as used in simulation study 1 (shown in table 3.16), apart from the addition of X_{PL_i} and $X_{EL_{ij}}$ to the model (as shown in model 3.14). All other parameter values were the same as in table 3.16, though the estimand values differed (these are described below). The five non-enrolment scenarios are shown in table 3.17.

For each scenario, I calculated estimands based on the set of episodes enrolled in the trial. I calculated each estimand by generating a single large dataset of 1,000,000 patients (1,500,000 episodes), and then excluding episodes according to model 3.15 above. I then generated both added-benefit and policy-benefit potential treatment effects for each episode, and calculated the relevant estimand based on these. These estimand values are shown in table 3.18.

3.4.3.2 Results

Results are shown in figures 3.2 and 3.3. The per-episode added-benefit estimator was unbiased across all scenarios, and had close to nominal coverage. The per-patient and policy-benefit estimators were unbiased across most scenarios, however, I identified several sources of bias which I discuss further below. Coverage of 95% confidence intervals was close to nominal for all settings in which estimators were unbiased.

The per-patient added-benefit estimator was biased for non-enrolment scenario 4, where non-enrolment was differential across treatment groups based on previous outcome. This matches the mathematical derivations from sections 3.3.5-3.3.11. I also identified a small bias in non-enrolment scenario 5 (where non-enrolment is differential across treatment groups based on prognosis at episode 2) under treatment effect scenarios 3 and 6 (when the size of the treatment effect varied across patients with different values of M_i). This bias was much smaller than that seen in non-enrolment scenario 4, but may still be large enough to cause concern.

The policy-benefit estimators (both per-patient and per-episode) were biased in non-enrolment scenarios 4 and 5. This occurred despite the fact that these estimators

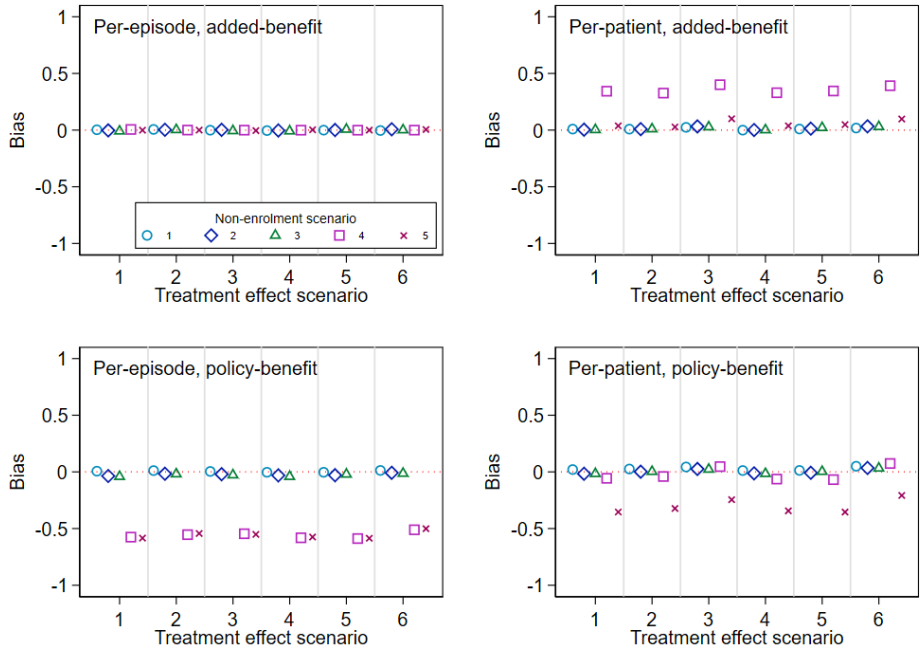
Table 3.17: Parameters for different episode 2 non-enrolment scenarios (simulation study 2a). For all scenarios, I set $\alpha^{R_2} = 0.05$ and $\gamma^{R_2} = 0.10$

Scenario	$\beta_{X_{PL}}$	$\beta_{X_{EL}}$	$\beta^{R_2}_{X_{PL}}$	$\beta^{R_2}_{X_{EL}}$	$\delta^{R_2}_{Xpl}$	$\delta^{R_2}_{Xel}$
Scenario 1 – Non-enrolment depends on previous treatment allocation	0	0	0	0	0	0
Scenario 2 – Non-enrolment depends on previous treatment allocation and previous outcome	10	0	0.25	0	0	0
Scenario 3 – Non-enrolment depends on previous treatment allocation and baseline prognosis at episode 2	0	10	0	0.25	0	0
Scenario 4 – Non-enrolment is differential between treatment groups based on previous outcome	10	0	0	0	0.5	0
Scenario 5 – Non-enrolment is differential between treatment groups based on baseline prognosis at episode 2	0	10	0	0	0	0.5

Table 3.18: Estimand values for simulation study 2a. *Treatment effect mechanism*: S1 = Constant treatment effect, S2 = Treatment effect varies across episode, S3 = Treatment effect varies across patients with different values of M_i, S4 = Treatment effect carries forward, S5 = Treatment becomes less effective on re-use, S6 = Treatment effect varies across episodes, across patients with different values of M_i, carries forward, and becomes less effective on re-use. *Non-enrolment scenarios*: S1 = Non-enrolment depends on previous treatment allocation, S2 = Non-enrolment depends on previous treatment allocation and previous outcome, S3 = Non-enrolment depends on previous treatment allocation and baseline prognosis at episode 2, S4 = Non-enrolment is differential between treatment groups based on previous outcome, S5 = Non-enrolment is differential between treatment groups based on baseline prognosis at episode 2.

Treatment effect mechanism	**Non-enrolment scenario**	β_E^{AB}	β_P^{AB}	β_E^{PB}	β_P^{PB}
S1	S1	3	3	3	3
	S2	3	3	3	3
	S3	3	3	3	3
	S4	3	3	3	3
	S5	3	3	3	3
S2	S1	3.47	3.34	3.47	3.34
	S2	3.42	3.29	3.42	3.29
	S3	3.42	3.29	3.42	3.29
	S4	3.42	3.29	3.42	3.29
	S5	3.42	3.29	3.42	3.29
S3	S1	4.97	4.5	4.97	4.5
	S2	4.92	4.5	4.92	4.5
	S3	4.92	4.5	4.92	4.5
	S4	4.92	4.5	4.92	4.5
	S5	4.92	4.5	4.92	4.5
S4	S1	3	3	3.31	3.23
	S2	3	3	3.28	3.19
	S3	3	3	3.28	3.19
	S4	3	3	3.28	3.19
	S5	3	3	3.28	3.19
S5	S1	2.56	2.68	2.07	2.33
	S2	2.61	2.73	2.16	2.42
	S3	2.61	2.73	2.16	2.42
	S4	2.67	2.77	2.16	2.42
	S5	2.67	2.77	2.16	2.42
S6	S1	4.99	4.52	4.81	4.39
	S2	4.95	4.52	4.78	4.40
	S3	4.95	4.52	4.78	4.40
	S4	5.01	4.57	4.78	4.40
	S5	5.01	4.57	4.78	4.40

Figure 3.2: Bias in estimators across different treatment effect and non-enrolment scenarios for simulation study 2a. Monte Carlo standard errors ranges: per-episode added-benefit 0.003-0.006; per-episode policy-benefit 0.004-0.008; per-patient added-benefit 0.003-0.006; per-patient policy-benefit 0.004-0.007.

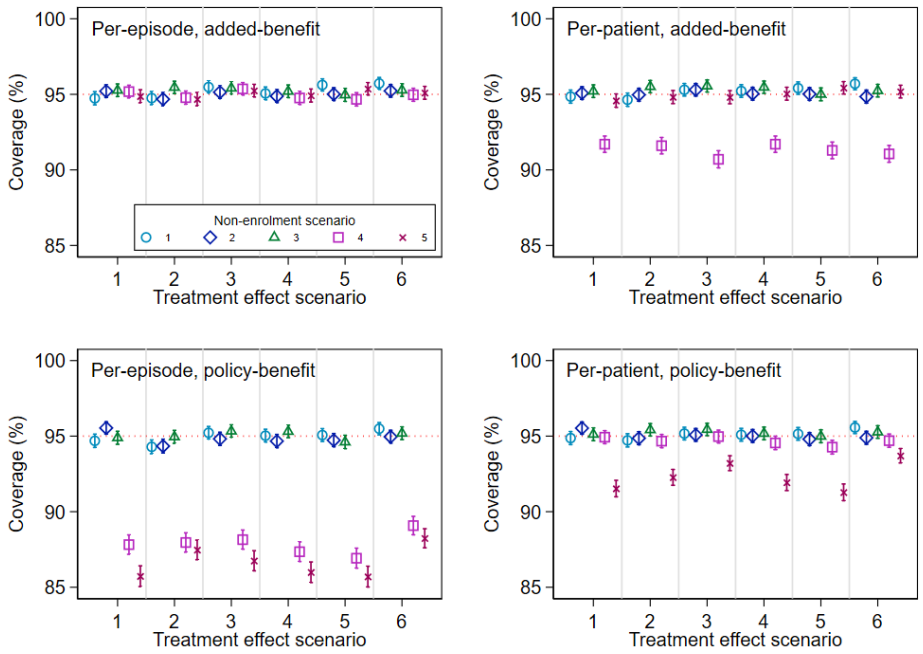


correctly modelled the causal effect of the previous treatment allocation on the outcome and treatment effect. This bias was a result of the model providing biased estimates of the parameter γ in these scenarios (which represents the effect of the previous allocation on outcome); because this parameter is used to construct policy-benefit estimates, these in turn will also be biased.

In these scenarios, episode 1 intervention patients with good outcomes were less likely to re-enrol for episode 2. At episode 2 therefore, most patients with a good outcome would have been allocated control in the previous episode. This created a false association between previous treatment allocation and outcome, which led to biased estimates of γ.

Interestingly, the per-patient policy-benefit estimator had negligible bias for non-enrolment scenario 4; this is likely because the per-patient estimator is biased upwards in this scenario, and the policy-benefit estimator is biased downwards, and the two biases cancel each other to some degree; however, under different parameter values it is likely that one of the biases would overtake the other, and the estimator would be biased. This is explored further in simulation study 2b below.

Figure 3.3: Coverage of estimators across different treatment effect and non-enrolment scenarios for simulation study 2a. Error bars are 95% confidence intervals based on Monte Carlo standard errors.



3.4.4 Simulation study 2b: further exploring bias associated with per-patient and policy-benefit estimators under non-enrolment scenarios 4 and 5

3.4.4.1 Data generating methods

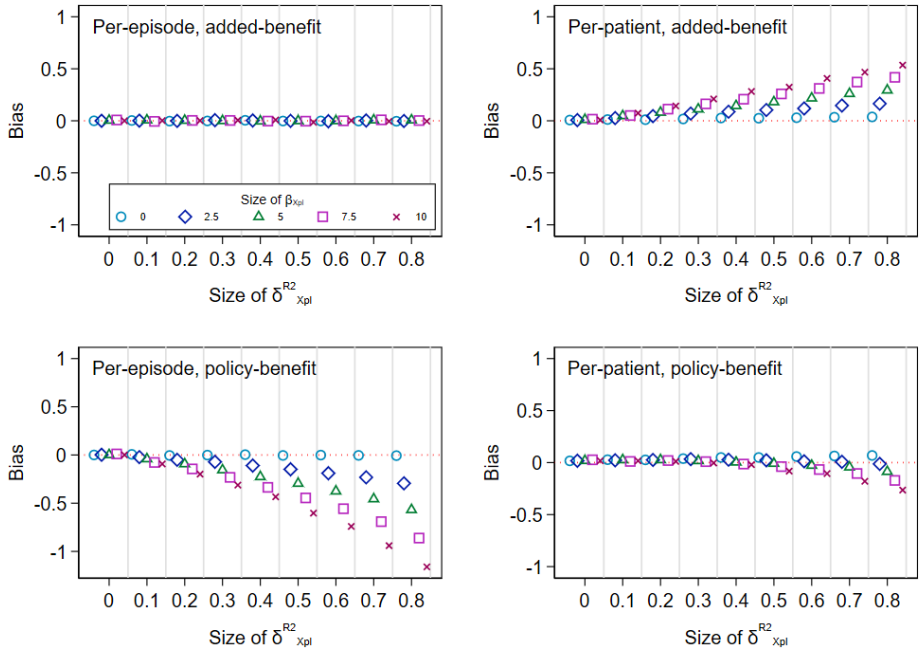
In this simulation study, I further explored some of the bias from per-patient and policy-benefit estimators associated with non-enrolment. I generated outcomes according to model 3.14, and probability of non-enrolment according to model 3.15. In the previous simulation study (2a) I used fairly large values for parameters associated with non-enrolment. In this simulation study, I used a range of values in order to assess how large the relevant parameter values needed to be in order for bias to become apparent.

Because the treatment effect mechanism did not have a large impact on bias in most scenarios in simulation study 2a, I opted to use a single treatment effect mechanism here. I used the constant treatment effect model for all scenarios (treatment effect mechanism 1); as such, the value of all estimands in these scenarios was 3.

For non-enrolment scenario 4, I varied $\beta_{X_{PL}}$ and $\delta^{R_2}_{Xpl}$ (which represents the increase in the probability of non-enrolment for patients with $X_{PL} = 1$) in a factorial manner; I varied $\beta_{X_{PL}}$ between 0, 2.5, 5, 7.5, and 10, and I varied $\delta^{R_2}_{Xpl}$ between 0, 0.1, 0.2, 0.3, 0.4, 0.5, 0.6, 0.7, and 0.8. This led to 5x9=45 scenarios. I set $\beta_{X_{EL}} = 0$ for all scenarios.

For non-enrolment scenario 5, I varied $\beta_{X_{EL}}$ and $\delta^{R_2}_{Xel}$ in a factorial manner; I

Figure 3.4: Bias in different estimators across non-enrolment scenario 4 in simulation study 2b. Monte Carlo standard errors ranges: per-episode added-benefit 0.003-0.006; per-episode policy-benefit 0.004-0.008; per-patient added-benefit 0.003-0.006; per-patient policy-benefit 0.004-0.007.



varied $\beta_{X_{EL}}$ between 0, 2.5, 5, 7.5, and 10, and I varied $\delta^{R_2}_{X_{el}}$ between 0, 0.1, 0.2, 0.3, 0.4, 0.5, 0.6, 0.7, and 0.8. This led to 5x9=45 scenarios. I set $\beta_{X_{PL}} = 0$ for all scenarios.

3.4.4.2 Results

Results are shown in figures 3.4 to 3.7. The policy-benefit estimators were biased when either X_{PL_i} or $X_{EL_{ij}}$ had strong associations with both outcome and probability of non-enrolment. When either association was small, bias was minimal, except when the other association was extremely large.

Similarly, the per-patient added-benefit estimator was biased when X_{PL_i} had a strong association with both outcome and probability of non-enrolment; when either of these associations were small, bias was negligible, except when the other association was extremely large.

Unlike in simulation study 2a, I found the per-patient policy-benefit estimator was biased in certain settings, indicating that the two competing biases will not always cancel out.

Figure 3.5: Bias in different estimators across non-enrolment scenario 5 in simulation study 2b. Monte Carlo standard errors ranges: per-episode added-benefit 0.003-0.006; per-episode policy-benefit 0.004-0.007; per-patient added-benefit 0.003-0.006; per-patient policy-benefit 0.004-0.007.

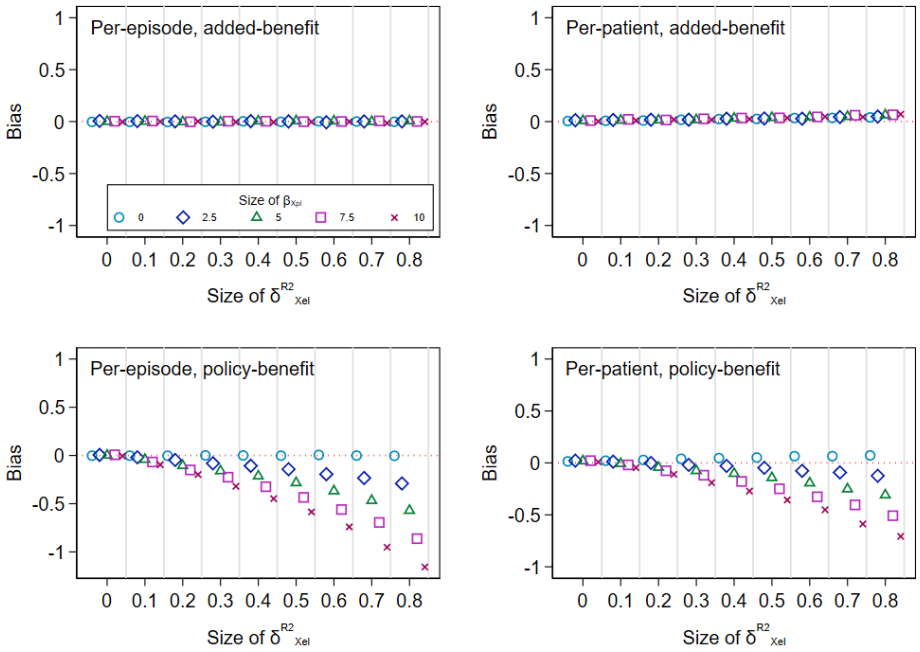


Figure 3.6: Coverage of different estimators across non-enrolment scenario 4 in simulation study 2b. Monte Carlo standard errors ranges: per-episode added-benefit 0.2-0.2; per-episode policy-benefit 0.2-0.5; per-patient added-benefit 0.2-0.3; per-patient policy-benefit 0.2-0.3.

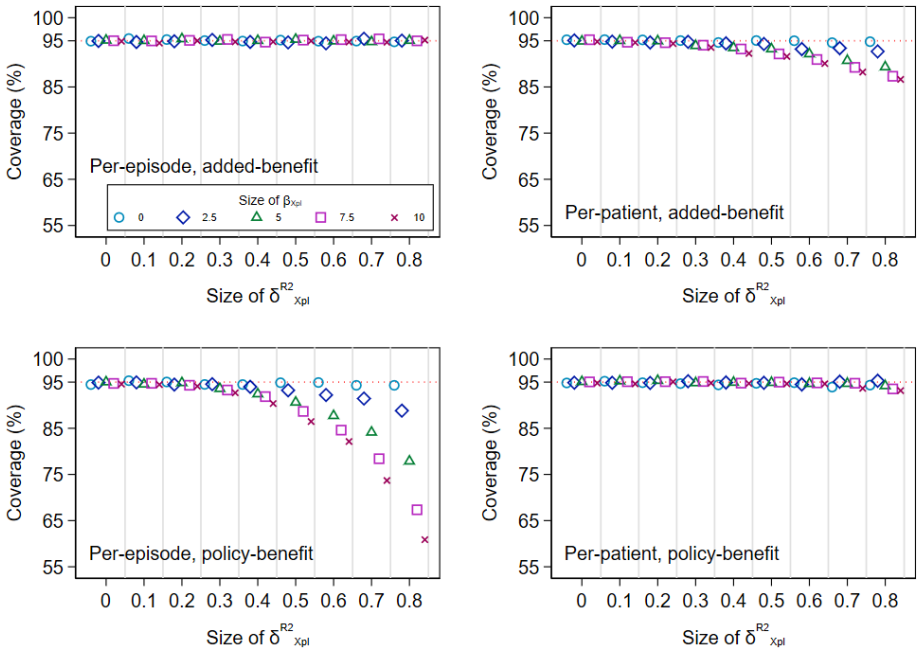
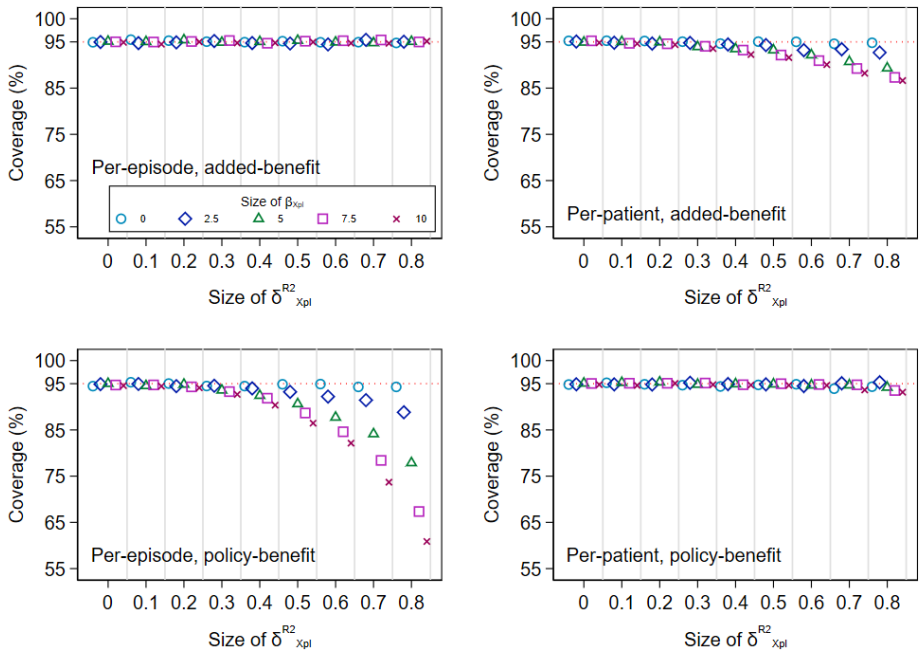


Figure 3.7: Coverage of different estimators across non-enrolment scenario 5 in simulation study 2b. Monte Carlo standard errors ranges: per-episode added-benefit 0.2-0.2; per-episode policy-benefit 0.2-0.5; per-patient added-benefit 0.2-0.2; per-patient policy-benefit 0.2-0.4.



3.5 Discussion

In this chapter I evaluated a set of independence estimators for re-randomisation trials. I evaluated these estimators under a wide range of data generating mechanisms and non-enrolment mechanisms. Not all of the scenarios I considered are likely to be common in practice. For instance, in the SWIM example ibuprofen is unlikely to become less effective each time it is used. Similarly, there is unlikely to be differential non-enrolment between treatment groups as the trial is blinded. However, it is useful to examine methods under unlikely scenarios to see if and when they will break down. Furthermore, these other scenarios may be more realistic for other trials. For example, there may be differential non-enrolment in unblinded studies, and drugs with long half-lives may be more likely to become less effective on reuse.

I found that the per-episode added-benefit estimator was unbiased in all scenarios. The per-patient estimators and policy-benefit estimators were unbiased under the assumption of no differential non-enrolment. The policy-benefit estimator also relied on the assumption that the causal model was correctly specified. If the causal model is incorrectly specified, then the policy-benefit estimators are likely to be biased. However, this depends on the type of misspecification. Including an unneeded term is unlikely to cause bias. For example, including $Z_{i,j-1}$ in the causal model will not cause bias if $Z_{i,j-1}$ does not affect the potential outcomes. However, excluding $Z_{i,j-1}$ from the

causal model will lead to bias if $Z_{i,j-1}$ does affect the potential outcomes.

The results in this chapter show that the per-episode added-benefit estimator provides valid results in all scenarios considered, and so can be recommended for use in practice without concern. The per-patient and policy-benefit estimators rely on certain assumptions about the data and therefore, if these estimators are used, it would be useful to conduct sensitivity analyses to evaluate to what extent results may be affected by violations to these assumptions.

4 Mixed-effects models

In this chapter I evaluate the use of mixed-effects models in re-randomisation trials. I discussed mixed-effects models in chapter 1 (section 1.6.1.1). I briefly summarise this method in section 4.1.1 below, then discuss its potential benefits and downsides in sections 4.1.2 and 4.1.3. In sections 4.2 and 4.3 I evaluate bias from mixed-effects models using mathematical derivations and simulation.

4.1 Mixed-effects models

4.1.1 Overview

Mixed-effects models [51] with a random intercept for patient (hereafter referred to as 'mixed-effects models') take the form of analysis model 1.2:

$$Y_{ij} = \hat{\alpha} + \hat{\beta}_{MM} Z_{ij} + \mu_i + \varepsilon_{ij}$$

where $\mu_i \sim N(0, \hat{\sigma}^2_\mu)$ and $\varepsilon_{ij} \sim N(0, \hat{\sigma}^2_\varepsilon)$, and μ_i and ε_{ij} are independent. I have used $\hat{\beta}_{MM}$ here instead of $\hat{\beta}$ to denote this is an estimate from a mixed-effects model.

This analysis model directly models the clustering structure of the data by including a random intercept for patient (the term μ_i). Therefore, this analysis model allows for correlation between episodes from the same patient. It assumes an exchangeable correlation structure, where the correlation between any two episodes from the same patient is the same (i.e. all episodes within a patient are equally correlated).

This model is typically estimated using maximum-likelihood (or restricted maximum likelihood for small sample settings); further information on maximum likelihood and restricted maximum likelihood is available elsewhere [51, 52, 53].

4.1.2 Potential benefits

The main benefit of mixed-effects models is that they are more efficient than independence estimators (that is, they have lower variance and higher power [9]). This implies that if mixed-effects models are used, the overall sample size could be reduced, allowing re-randomisation trials to complete recruitment more quickly.

4.1.3 Potential drawbacks

The main downside for mixed-effects models is that it is not clear whether they are able to provide unbiased treatment effect estimates in most settings. For instance, as discussed in chapter 1, we previously evaluated mixed-effects models under some simple data generating mechanisms [9]. We found they were unbiased when the treatment effect was constant (data generating model 1.3), but were biased when the treatment effect carried forward, as in data generating model 2.2:

$$Y_{ij} = \alpha + \beta Z_{ij} + \gamma Z_{i,j-1} + \mu_i + \varepsilon_{ij}$$

However, we found that this bias could be mitigated by including a term for $Z_{i,j-1}$ in the analysis model. One implication from this is that, for scenarios where treatment history (e.g. $Z_{i,j-1}$) matters, including treatment history in the analysis model may be sufficient to obtain unbiased estimates.

However, it is not clear whether mixed-effects models are biased for other data generating mechanisms, such as those considered in chapters 2 and 3 (e.g. when treatment effect varies by episode or by value of M_i, or becomes less effective on re-use), and if so, whether it would be possible to modify the analysis model as above to obtain unbiased estimates.

Previous research in the informative cluster size setting has found that mixed-effects models are biased for both the per-episode and per-patient treatment effects when the cluster size is informative [44], so it is plausible the same issue will arise in re-randomisation trials.

4.2 Mathematical derivation of bias

In this section I derive the expected value of the mixed-effects model estimator $\hat{\beta}_{MM}$. I then compare this against the per-episode added-benefit estimand under a range of data generating and non-enrolment mechanisms. I do not consider the per-patient or policy-benefit estimands in this section, as they require either weighting by $\frac{1}{M_i}$ or modelling the causal effect of treatment history, both of which make mathematical derivations much more complicated.

As before, I restrict the setting to a trial with a 1:1 allocation ratio, where patients experience a maximum of two episodes. I assume a large sample size, so that asymptotic results apply. I use the same set of scenarios as in chapter 3:

1. Constant treatment effect
2. Treatment effect varies across episode
3. Treatment effect varies across value of M_i

4. Treatment effect carries forward into the second episode

5. Treatment becomes less effective on re-use

6. Constant treatment effect, differential non-enrolment based on outcome in previous episode

7. Constant treatment effect, differential non-enrolment based on expected outcome in current episode

As before, the first five scenarios do not involve any non-enrolment (i.e. patients who experience two episodes will enrol in the trial for both episodes). The last two scenarios use a constant treatment effect mechanism, but some patients who experience two episodes do not re-enrol for their second episode. In scenario 6, non-enrolment is differential between treatment arms depending on their outcome in the first episode (i.e. different types of patients from each treatment arm will re-enrol). For example, patients who received the intervention in episode 1 and had a good outcome are more likely to re-enrol at episode 2 than patients who received control in episode 1 and had a poor outcome. In scenario 7, non-enrolment is differential between treatment arms depending on their expected outcome in the second episode; for example, patients who received the intervention in episode 1 and had a good prognosis at baseline for episode 2 are more likely to re-enrol than patients who received control in episode 1 and have a poor prognosis. Further details on each of these scenarios is given in the sections below, and in chapter 3. I note that in previous work we derived the expected value of $\hat{\beta}_{MM}$ for scenarios 1 and 4 [9], however I repeat these derivations here for completeness.

For each scenario, I will assume that $\alpha = 0$. Expected values of the components $\hat{\beta}_{B_1}$, $\hat{\beta}_{B_2}$, and $\hat{\beta}_W$ under the different scenarios are shown in table 3.2 in chapter 3. Let $\hat{\sigma}^2_\mu$, $\hat{\sigma}^2_\varepsilon$, and $\hat{\beta}_{MM}$ denote maximum-likelihood estimates. Then, let $V\left(Y_{ij}|Z_{ij}\right) = \hat{\sigma}^2_\mu + \hat{\sigma}^2_\varepsilon = \hat{\sigma}^2$, and let $\hat{\vartheta} = \frac{\hat{\sigma}^2_\mu}{\hat{\sigma}^2_\mu + \hat{\sigma}^2_\varepsilon}$, i.e. $\hat{\vartheta}$ is the estimated intraclass correlation coefficient (ICC). Then, let $\vartheta^* = E\left(\hat{\vartheta}\right)$, i.e. ϑ^* is the expected value of the estimated intraclass correlation coefficient. Note that the term $V\left(Y_{ij}|Z_{ij}\right)$ denotes an estimated variance; in this thesis, I use $V()$ to denote $\hat{V}()$ for simplicity (i.e. I omit the hat from the estimated variances).

In section 4.2.1 I derive the estimated variances of the components $\hat{\beta}_{B_1}$, $\hat{\beta}_{B_2}$, and $\hat{\beta}_W$ (as these are used in the derivation of the mixed-effects model estimator). In section 4.2.2 I derive the mixed-effects model estimator, and then in sections 4.2.3 to 4.2.9 I derive the expected value of this estimator under the seven scenarios listed above. I provide a summary of results in section 4.2.10.

4.2.1 Estimated variances of estimation components

I now derive the estimated variance for each of the three components $\hat{\beta}_{B_1}$, $\hat{\beta}_{B_2}$, and $\hat{\beta}_W$, based on analysis model 1.2; these variances are used in deriving the overall

estimator of treatment effect from a mixed-model. These results are shown in table 4.1.

The estimated variance of $\hat{\beta}_{B_1}$ is:

$$V\left(\hat{\beta}_{B_1}\right) = V\left(\frac{1}{\frac{N_T}{2}(1-p)}\left(\sum_{i\in Z=(1)} Y_{i1} - \sum_{i\in Z=(0)} Y_{i1}\right)\right)$$

Because outcomes from different patients are independent, the variance of the sum is equal to the sum of the variances, and so this expression becomes:

$$\frac{1}{\left(\frac{N_T}{2}(1-p)\right)^2}\left(\sum_{i\in Z=(1)} V\left(Y_{i1}\right) + \sum_{i\in Z=(0)} V\left(Y_{i1}\right)\right) = \frac{1}{\left(\frac{N_T}{2}(1-p)\right)^2} 2\left(\hat{\sigma}^2 \frac{N_T}{2}(1-p)\right) = \frac{2\hat{\sigma}^2}{\frac{N_T}{2}(1-p)}$$

Next, the estimated variance of $\hat{\beta}_{B_2}$ is:

$$V\left(\hat{\beta}_{B_2}\right) = V\left(\frac{1}{\frac{N_T p}{2}}\left(\sum_{i\in Z=(1,1)} (Y_{i1}+Y_{i2}) - \sum_{i\in Z=(0,0)} (Y_{i1}+Y_{i2})\right)\right) = \frac{1}{\left(\frac{N_T p}{2}\right)^2}\left(\sum_{i\in Z=(1,1)} V\left(Y_{i1}+Y_{i2}\right) + \sum_{i\in Z=(0,0)} V\left(Y_{i1}+Y_{i2}\right)\right) = \frac{1}{\left(\frac{N_T p}{2}\right)^2}\Big(\sum_{i\in Z=(1,1)} V\left(Y_{i1}\right) + V\left(Y_{i2}\right) + 2Cov\left(Y_{i1},Y_{i2}\right) + \sum_{i\in Z=(0,0)} V\left(Y_{i1}\right) + V\left(Y_{i2}\right) + 2Cov\left(Y_{i1},Y_{i2}\right)\Big)$$

where $Cov\left(Y_{i1},Y_{i2}\right)$ denotes the covariance between the two outcomes. For episodes from different patients, this is 0; for episodes from the same patient (as in the above expression), this becomes:

$$Cov\left(Y_{i1},Y_{i2}\right) = Corr\left(Y_{i1},Y_{i2}\right)\sqrt{V\left(Y_{i1}\right)}\sqrt{V\left(Y_{i2}\right)} = \hat{\vartheta}\hat{\sigma}^2$$

where $Corr\left(Y_{i1},Y_{i2}\right)$ represents the correlation between the two outcomes, which is $\hat{\vartheta} = \frac{\hat{\sigma}_\mu^2}{\hat{\sigma}_\mu^2+\hat{\sigma}_\varepsilon^2}$.

Then, the variance becomes:

$$\frac{1}{\left(\frac{N_T p}{2}\right)^2}\left(\sum_{i\in Z=(1,1)} 2\hat{\sigma}^2 + 2\hat{\vartheta}\hat{\sigma}^2 + \sum_{i\in Z=(0,0)} 2\hat{\sigma}^2 + 2\hat{\vartheta}\hat{\sigma}^2\right) =$$

$$\frac{1}{\left(\frac{N_T p}{2}\right)^2}\left(\frac{2N_T p}{4}\left(2\hat{\sigma}^2 + 2\hat{\vartheta}\hat{\sigma}^2\right)\right) = \frac{2\hat{\sigma}^2}{\frac{N_T p}{2}}\left(1+\hat{\vartheta}\right)$$

Finally, the estimated variance of $\hat{\beta}_W$ is:

$$V\left(\hat{\beta}_W\right) = V\left(\frac{1}{\frac{N_T p}{2}}\left(\sum_{i\in Z=(1,0)}(Y_{i1}-Y_{i2}) - \sum_{i\in Z=(0,1)}(Y_{i1}-Y_{i2})\right)\right) =$$

$$\frac{1}{\left(\frac{N_T p}{2}\right)^2}\left(\sum_{i\in Z=(1,0)} V\left(Y_{i1}-Y_{i2}\right) + \sum_{i\in Z=(0,1)} V\left(Y_{i1}-Y_{i2}\right)\right) =$$

$$\frac{1}{\left(\frac{N_T p}{2}\right)^2}\Big(\sum_{i\in Z=(1,0)} V\left(Y_{i1}\right) + V\left(Y_{i2}\right) - 2Cov\left(Y_{i1},Y_{i2}\right) +$$

$$\sum_{i\in Z=(0,1)} V\left(Y_{i1}\right) + V\left(Y_{i2}\right) - 2Cov\left(Y_{i1},Y_{i2}\right)\Big)$$

This expression becomes:

$$\frac{1}{\left(\frac{N_T p}{2}\right)^2}\left(\frac{2N_T p}{4}\left(2\hat{\sigma}^2 - 2\hat{\vartheta}\hat{\sigma}^2\right)\right) = \frac{2\hat{\sigma}^2}{\frac{N_T p}{2}}\left(1-\hat{\vartheta}\right)$$

4.2.2 Overall estimator of treatment effect from a mixed-effects model

For a re-randomisation trial with a maximum of two episodes, the estimator $\hat{\beta}_{MM}$ from analysis model 1.2 can be written as [9]:

$$\hat{\beta}_{MM} = \frac{\frac{\hat{\beta}_{B_1}}{V\left(\hat{\beta}_{B_1}\right)} + \frac{\hat{\beta}_{B_2}}{V\left(\hat{\beta}_{B_2}\right)} + \frac{\hat{\beta}_W}{V\left(\hat{\beta}_W\right)}}{\frac{1}{V\left(\hat{\beta}_{B_1}\right)} + \frac{1}{V\left(\hat{\beta}_{B_2}\right)} + \frac{1}{V\left(\hat{\beta}_W\right)}}$$

This can be re-arranged as follows:

$$\hat{\beta}_{MM} = \frac{\frac{V\left(\hat{\beta}_{B_2}\right)V\left(\hat{\beta}_W\right)\hat{\beta}_{B_1} + V\left(\hat{\beta}_{B_1}\right)V\left(\hat{\beta}_W\right)\hat{\beta}_{B_2} + V\left(\hat{\beta}_{B_1}\right)V\left(\hat{\beta}_{B_2}\right)\hat{\beta}_W}{V\left(\hat{\beta}_{B_1}\right)V\left(\hat{\beta}_{B_2}\right)V\left(\hat{\beta}_W\right)}}{\frac{V\left(\hat{\beta}_{B_2}\right)V\left(\hat{\beta}_W\right) + V\left(\hat{\beta}_{B_1}\right)V\left(\hat{\beta}_W\right) + V\left(\hat{\beta}_{B_1}\right)V\left(\hat{\beta}_{B_2}\right)}{V\left(\hat{\beta}_{B_1}\right)V\left(\hat{\beta}_{B_2}\right)V\left(\hat{\beta}_W\right)}}$$

$$= \frac{V\left(\hat{\beta}_{B_2}\right)V\left(\hat{\beta}_W\right)\hat{\beta}_{B_1} + V\left(\hat{\beta}_{B_1}\right)V\left(\hat{\beta}_W\right)\hat{\beta}_{B_2} + V\left(\hat{\beta}_{B_1}\right)V\left(\hat{\beta}_{B_2}\right)\hat{\beta}_W}{V\left(\hat{\beta}_{B_2}\right)V\left(\hat{\beta}_W\right) + V\left(\hat{\beta}_{B_1}\right)V\left(\hat{\beta}_W\right) + V\left(\hat{\beta}_{B_1}\right)V\left(\hat{\beta}_{B_2}\right)}$$

Table 4.1: Estimates and estimated variances of between- and within-patient components based on a mixed-effects model

Component	**Treatment sequences**	**Formula**	**Estimated variance**
$\hat{\beta}_{B_1}$	$Z = (1)$, $Z = (0)$	$\hat{\beta}_{B_1} = \frac{1}{\frac{N_T}{2}(1-p)} \left(\sum_{i \in Z=(1)} Y_{i1} - \sum_{i \in Z=(0)} Y_{i1} \right)$	$V\left(\hat{\beta}_{B_1}\right) = \frac{2\hat{\sigma}^2}{\frac{N_T}{2}(1-p)}$
$\hat{\beta}_{B_2}$	$Z = (0,0)$, $Z = (1,1)$	$\hat{\beta}_{B_2} = \frac{1}{\frac{N_T p}{2}} \left(\sum_{i \in Z=(1,1)} (Y_{i1} + Y_{i2}) - \sum_{i \in Z=(0,0)} (Y_{i1} + Y_{i2}) \right)$	$V\left(\hat{\beta}_{B_2}\right) = \frac{2\hat{\sigma}^2}{\frac{N_T p}{2}} \left(1 + \hat{\vartheta}\right)$
$\hat{\beta}_W$	$Z = (0,1)$, $Z = (1,0)$	$\hat{\beta}_W = \frac{1}{\frac{N_T p}{2}} \left(\sum_{i \in Z=(1,0)} (Y_{i1} - Y_{i2}) - \sum_{i \in Z=(0,1)} (Y_{i1} - Y_{i2}) \right)$	$V\left(\hat{\beta}_W\right) = \frac{2\hat{\sigma}^2}{\frac{N_T p}{2}} \left(1 - \hat{\vartheta}\right)$

Substituting in the estimated variances $V\left(\hat{\beta}_{B_1}\right)$, $V\left(\hat{\beta}_{B_2}\right)$, and $V\left(\hat{\beta}_W\right)$ from table 4.1 we get:

$$\hat{\beta}_{MM} = \frac{\left(\frac{1+\hat{\vartheta}}{p}\right)\left(\frac{1-\hat{\vartheta}}{p}\right)\hat{\beta}_{B_1} + \left(\frac{1}{1-p}\right)\left(\frac{1-\hat{\vartheta}}{p}\right)\hat{\beta}_{B_2} + \left(\frac{1}{1-p}\right)\left(\frac{1+\hat{\vartheta}}{p}\right)\hat{\beta}_W}{\left(\frac{1+\hat{\vartheta}}{p}\right)\left(\frac{1-\hat{\vartheta}}{p}\right) + \left(\frac{1}{1-p}\right)\left(\frac{1-\hat{\vartheta}}{p}\right) + \left(\frac{1}{1-p}\right)\left(\frac{1+\hat{\vartheta}}{p}\right)}$$

Note that the $\frac{2\hat{\sigma}^2}{\frac{N_T}{2}}$ terms have cancelled out. We can re-arrange this as follows:

$$\hat{\beta}_{MM} = \frac{\frac{\left(1+\hat{\vartheta}\right)\left(1-\hat{\vartheta}\right)(1-p)\hat{\beta}_{B_1} + \left(1-\hat{\vartheta}\right)p\hat{\beta}_{B_2} + \left(1+\hat{\vartheta}\right)p\hat{\beta}_W}{p^2(1-p)}}{\frac{\left(1+\hat{\vartheta}\right)\left(1-\hat{\vartheta}\right)(1-p) + \left(1-\hat{\vartheta}\right)p + \left(1+\hat{\vartheta}\right)p}{p^2(1-p)}}$$

$$= \frac{\left(1+\hat{\vartheta}\right)\left(1-\hat{\vartheta}\right)(1-p)\hat{\beta}_{B_1} + \left(1-\hat{\vartheta}\right)p\hat{\beta}_{B_2} + \left(1+\hat{\vartheta}\right)p\hat{\beta}_W}{\left(1+\hat{\vartheta}\right)\left(1-\hat{\vartheta}\right)(1-p) + \left(1-\hat{\vartheta}\right)p + \left(1+\hat{\vartheta}\right)p}$$

The numerator can be arranged as:

$$\left(1+\hat{\vartheta}\right)\left(1-\hat{\vartheta}\right)(1-p)\hat{\beta}_{B_1} + \left(1-\hat{\vartheta}\right)p\hat{\beta}_{B_2} + \left(1+\hat{\vartheta}\right)p\hat{\beta}_W$$

$$= \left([1-p]\left[1+\hat{\vartheta}\right]\hat{\beta}_{B_1} + p\hat{\beta}_{B_2}\right)\left(1-\hat{\vartheta}\right) + p\left(1+\hat{\vartheta}\right)\hat{\beta}_W$$

And the denominator can be arranged as:

$$\left(1+\hat{\vartheta}\right)\left(1-\hat{\vartheta}\right)(1-p) + \left(1-\hat{\vartheta}\right)p + \left(1+\hat{\vartheta}\right)p$$

$$= \left(1-\hat{\vartheta}^2\right)(1-p) + 2p$$

$$= 1 - \hat{\vartheta}^2 - p + \hat{\vartheta}^2 p + 2p$$

$$= 1 - \hat{\vartheta}^2 + \hat{\vartheta}^2 p + p$$

Putting the numerator and denominator together, we get:

$$\hat{\beta}_{MM} = \frac{\left((1-p)\left(1+\hat{\vartheta}\right)\hat{\beta}_{B_1} + p\hat{\beta}_{B_2}\right)\left(1-\hat{\vartheta}\right) + p\left(1+\hat{\vartheta}\right)\hat{\beta}_W}{1 - \hat{\vartheta}^2 + \hat{\vartheta}^2 p + p}$$

In the above equation, the terms $\hat{\vartheta}$, $\hat{\beta}_{B_1}$, $\hat{\beta}_{B_2}$, and $\hat{\beta}_W$ are random variables. We can take the expectation of $\hat{\beta}_{MM}$ using a first order Taylor series expansion:

$$E\left(\hat{\beta}_{MM}\right) = E\left(\frac{\left((1-p)\left(1+\hat{\vartheta}\right)\hat{\beta}_{B_1} + p\hat{\beta}_{B_2}\right)\left(1-\hat{\vartheta}\right) + p\left(1+\hat{\vartheta}\right)\hat{\beta}_W}{1-\hat{\vartheta}^2+\hat{\vartheta}^2 p + p}\right)$$
$$= \frac{\left((1-p)\left(1+\vartheta^*\right)E\left(\hat{\beta}_{B_1}\right) + pE\left(\hat{\beta}_{B_2}\right)\right)\left(1-\vartheta^*\right) + p\left(1+\vartheta^*\right)E\left(\hat{\beta}_W\right)}{1-\vartheta^{*2}+\vartheta^{*2}p+p} \quad (4.1)$$

where ϑ^* is the expected value of ϑ.

4.2.3 Scenario 1: constant treatment effect

Consider data-generating mechanism 1.3:

$$Y_{ij} = \alpha + \beta\ Z_{ij} + \mu_i + \varepsilon_{ij}$$

From chapter 3, we have $E\left(\hat{\beta}_{B_1}\right) = E\left(\hat{\beta}_{B_2}\right) = E\left(\hat{\beta}_W\right) = \beta$.

Substituting these expressions into model 4.1 above, we therefore get:

$$E\left(\hat{\beta}_{MM}\right) = \frac{\left((1-p)\left(1+\vartheta^*\right)\beta + p\beta\right)\left(1-\vartheta^*\right) + p\left(1+\vartheta^*\right)\beta}{1-\vartheta^{*2}+\vartheta^{*2}p+p} = \beta$$

4.2.4 S2: Treatment effect varies across episode

Consider data-generating mechanism 2.3:

$$Y_{ij} = \begin{cases} \alpha + \beta_1 Z_{ij} + \mu_i + \varepsilon_{ij} \text{ for } j = 1 \\ \alpha + \beta_2 Z_{ij} + \mu_i + \varepsilon_{ij} \text{ for } j = 2 \end{cases}$$

From chapter 3, we have:

$$E\left(\hat{\beta}_{B_1}\right) = \beta_1$$

$$E\left(\hat{\beta}_{B_2}\right) = \frac{1}{2}\left(\beta_1 + \beta_2\right)$$

$$E\left(\hat{\beta}_W\right) = \frac{1}{2}\left(\beta_1 + \beta_2\right)$$

Therefore:

$$E\left(\hat{\beta}_{MM}\right) = \frac{\left((1-p)\left(1+\vartheta^*\right)\beta_1 + \frac{p}{2}\left(\beta_1+\beta_2\right)\right)\left(1-\vartheta^*\right) + \frac{p}{2}\left(1+\vartheta^*\right)\left(\beta_1+\beta_2\right)}{1-\vartheta^{*2}+\vartheta^{*2}p+p}$$

The numerator can be simplified as follows:

$$\left(\left(1+\vartheta^*-p-p\vartheta^*\right)\beta_1 + \frac{p\beta_1}{2} + \frac{p\beta_2}{2}\right)\left(1-\vartheta^*\right) + \left(1+\vartheta^*\right)\left(\frac{p\beta_1}{2} + \frac{p\beta_2}{2}\right)$$

$$= \beta_1\left(1-\vartheta^{*2}+\vartheta^{*2}p\right) + \beta_2 p$$

Putting the numerator and denominator together, this expression becomes:

$$E\left(\hat{\beta}_{MM}\right) = \frac{\beta_1\left(1-\vartheta^{*2}+\vartheta^{*2}p\right)+\beta_2 p}{1-\vartheta^{*2}+\vartheta^{*2}p+p}$$

4.2.5 S3: Treatment effect varies across value of M_i

Consider data-generating mechanism 2.1:

$$Y_{ij} = \begin{cases} \alpha + \beta_1 Z_{ij} + \mu_i + \varepsilon_{ij} \text{ if } M_i = 1 \\ \alpha + \beta_2 Z_{ij} + \mu_i + \varepsilon_{ij} \text{ if } M_i = 2 \end{cases}$$

From chapter 3, we have:

$$E\left(\hat{\beta}_{B_1}\right) = \beta_1$$

$$E\left(\hat{\beta}_{B_2}\right) = \beta_2$$

$$E\left(\hat{\beta}_{W}\right) = \beta_2$$

Therefore:

$$E\left(\hat{\beta}_{MM}\right) = \frac{\left((1-p)\left(1+\vartheta^*\right)\beta_1 + p\beta_2\right)\left(1-\vartheta^*\right) + p\left(1+\vartheta^*\right)\beta_2}{1-\vartheta^{*2}+\vartheta^{*2}p+p}$$

$$= \frac{(1-p)\left(1+\vartheta^*\right)\beta_1 - \vartheta^*(1-p)\left(1+\vartheta^*\right)\beta_1 + 2p\beta_2}{1-\vartheta^{*2}+\vartheta^{*2}p+p}$$

$$= \frac{(1-p)\left(1-\vartheta^{*2}\right)\beta_1 + 2p\beta_2}{1-\vartheta^{*2}+\vartheta^{*2}p+p}$$

4.2.6 S4: Treatment effect carries forward into the second episode

Consider data-generating mechanism 2.2:

$$Y_{ij} = \alpha + \beta\ Z_{ij} + \gamma\ Z_{i,j-1} + \mu_i + \varepsilon_{ij}$$

From chapter 3, we have:

$$E\left(\hat{\beta}_{B_1}\right) = \beta$$

$$E\left(\hat{\beta}_{B_2}\right) = \beta + \frac{\gamma}{2}$$

$$E\left(\hat{\beta}_{W}\right) = \beta - \frac{\gamma}{2}$$

Therefore:

$$E\left(\hat{\beta}_{MM}\right) = \frac{\left((1-p)\left(1+\vartheta^*\right)\beta + p\left(\beta+\frac{\gamma}{2}\right)\right)\left(1-\vartheta^*\right) + p\left(1+\vartheta^*\right)\left(\beta-\frac{\gamma}{2}\right)}{1-\vartheta^{*2}+\vartheta^{*2}p+p}$$

$$= \beta - \gamma\frac{p\vartheta^*}{1-\vartheta^{*2}+\vartheta^{*2}p+p}$$

4.2.7 S5: Treatment becomes less effective on re-use

Consider data-generating mechanism 2.4:

$$Y_{ij} = \begin{cases} \alpha + \beta Z_{ij} + \mu_i + \varepsilon_{ij} \text{ for } Z_{i,j-1} = 0 \\ \alpha + (\beta+\delta) Z_{ij} + \mu_i + \varepsilon_{ij} \text{ for } Z_{i,j-1} = 1 \end{cases}$$

From chapter 3, we have:

$$E\left(\hat{\beta}_{B_1}\right) = \beta$$

$$E\left(\hat{\beta}_{B_2}\right) = \beta + \frac{\delta}{2}$$

$$E\left(\hat{\beta}_{W}\right) = \beta$$

Therefore:

$$E\left(\hat{\beta}_{MM}\right) = \frac{\left((1-p)\left(1+\vartheta^*\right)\beta + p\left(\beta+\frac{\delta}{2}\right)\right)\left(1-\vartheta^*\right) + p\beta\left(1+\vartheta^*\right)}{1-\hat{\vartheta}^2+\hat{\vartheta}^2 p+p}$$

$$= \frac{\beta\left((1-p)\left(1+\vartheta^*\right)\left(1-\vartheta^*\right) + p\left(1-\vartheta^*\right) + p\beta\left(1+\vartheta^*\right)\right) + \frac{p\delta}{2}\left(1-\vartheta^*\right)}{1-\vartheta^{*2}+\vartheta^{*2}p+p}$$

$$= \frac{\beta\left(1-\vartheta^{*2}+\vartheta^{*2}p+p\right) + \frac{p\delta}{2}\left(1-\vartheta^*\right)}{1-\vartheta^{*2}+\vartheta^{*2}p+p}$$

$$\beta + \frac{p\delta\left(1-\vartheta^*\right)}{2\left(1-\vartheta^{*2}+\vartheta^{*2}p+p\right)}$$

4.2.8 S6: Constant treatment effect, differential non-enrolment based on outcome in previous episode

This is the same scenario as in section 3.3.10 of chapter 3, where there is differential non-enrolment based on the patient's outcome in episode 1. Full details are available there. Briefly, consider the following data generating mechanism:

$$Y_{ij} = \alpha + \beta_{trt} Z_{ij} + \beta_{X_{PL}} X_{PL_i} + \mu_i + \varepsilon_{ij}$$

where X_{PL_i} is an unobserved binary patient-level variable (i.e. it is constant across episodes).

Now, $R_{i2} = 1$ denotes the patient has re-enrolled for episode 2, and $R_{i2} = 0$ denotes they have not re-enrolled. In this scenario, the probability of re-enrolment depends on two factors: treatment allocation in episode 1 (Z_{i1}) and the value of X_{PL_i} (where X_{PL_i} is a marker of the patient's outcome in episode 1).

Then, $\pi = P\left(X_{PL_i} = 1\right)$, and let $p_{zx} = P\left(R_{i2} = 1 | Z_{i1} = z_{i1}, X_{PL_i} = x_{PL_i}\right)$ (i.e. p_{zx} denotes the probability of being re-enrolled for a second episode given Z_{i1} and X_{PL_i}). So, for example, $p_{00} = P\left(R_{i2} = 1 | Z_{i1} = 0, X_{PL_i} = 0\right)$, and $p_{01} = P\left(R_{i2} = 1 | Z_{i1} = 0, X_{PL_i} = 1\right)$. For simplicity, I will assume that $\pi = 0.5$, and that $p_{00} = p_{11}$ and $p_{01} = p_{10}$.

Then, from chapter 3, we have:

$$E\left(\hat{\beta}_{B_1}\right) = \beta_{trt} + \beta_{X_{PL}} \frac{(p_{01} - p_{00})}{2(1-p)}$$

$$E\left(\hat{\beta}_{B_2}\right) = \beta_{trt} + \beta_{X_{PL}} \frac{(p_{00} - p_{01})}{2p}$$

$$E\left(\hat{\beta}_{w}\right) = \beta_{trt}$$

For the moment, I will focus on the first component of the numerator of equation 4.1:

$$(1-p)(1+\vartheta^*)E\left(\hat{\beta}_{B_1}\right)+pE\left(\hat{\beta}_{B_2}\right)=$$
$$(1-p)(1+\vartheta^*)\left(\beta_{trt}+\beta_{X_{PL}}\frac{(p_{01}-p_{00})}{2(1-p)}\right)+p\left(\beta_{trt}+\beta_{X_{PL}}\frac{(p_{00}-p_{01})}{2p}\right)$$

$$=(1-p)(1+\vartheta^*)\left(\frac{2(1-p)\beta_{trt}+\beta_{X_{PL}}(p_{01}-p_{00})}{2(1-p)}\right)+p\left(\frac{2p\beta_{trt}+\beta_{X_{PL}}(p_{00}-p_{01})}{2p}\right)$$

$$=(1+\vartheta^*)\left(\frac{2(1-p)\beta_{trt}+\beta_{X_{PL}}(p_{01}-p_{00})}{2}\right)+\left(\frac{2p\beta_{trt}+\beta_{X_{PL}}(p_{00}-p_{01})}{2}\right)$$

$$=\frac{1}{2}\Big(2(1-p)\beta_{trt}+\beta_{X_{PL}}(p_{01}-p_{00})+2\vartheta^*(1-p)\beta_{trt}+\vartheta^*\beta_{X_{PL}}(p_{01}-p_{00})+$$
$$2p\beta_{trt}+\beta_{X_{PL}}(p_{00}-p_{01})\Big)$$

$$=\frac{1}{2}\Big(\beta_{trt}(2(1-p)+2\vartheta^*(1-p)+2p)$$
$$+\beta_{X_{PL}}((p_{01}-p_{00})+\vartheta^*(p_{01}-p_{00})+(p_{00}-p_{01}))\Big)$$

$$=\frac{1}{2}(2\beta_{trt}(1+\vartheta^*-\vartheta^*p)+\vartheta^*\beta_{X_{PL}}(p_{01}-p_{00}))$$

Next, I will add in the term $(1-\vartheta^*)$ which is attached to the component above:

$$\left((1-p)(1+\vartheta^*)E\left(\hat{\beta}_{B_1}\right)+pE\left(\hat{\beta}_{B_2}\right)\right)(1-\vartheta^*)=$$
$$\frac{1}{2}(2\beta_{trt}(1+\vartheta^*-\vartheta^*p)+\vartheta^*\beta_{X_{PL}}(p_{01}-p_{00}))(1-\vartheta^*)$$

$$=\frac{1}{2}\Big(2\beta_{trt}(1+\vartheta^*-\vartheta^*p)+\vartheta^*\beta_{X_{PL}}(p_{01}-p_{00})-$$
$$2\vartheta^*\beta_{trt}(1+\vartheta^*-\vartheta^*p)-\vartheta^{*2}\beta_{X_{PL}}(p_{01}-p_{00})\Big)$$

Then, the final term in the numerator is:

$$p\left(1+\vartheta^*\right)E\left(\hat{\beta}_W\right) = p\left(1+\vartheta^*\right)\beta_{trt} = \frac{2p\beta_{trt}+2\vartheta^* p\beta_{trt}}{2}$$

Then, the overall numerator is:

$$\left((1-p)\left(1+\vartheta^*\right)E\left(\hat{\beta}_{B_1}\right)+pE\left(\hat{\beta}_{B_2}\right)\right)\left(1-\vartheta^*\right)+p\left(1+\vartheta^*\right)E\left(\hat{\beta}_W\right)$$

$$= \frac{1}{2}\Big(2\beta_{trt}\left(1+\vartheta^*-\vartheta^* p\right)+\vartheta^*\beta_{X_{PL}}\left(p_{01}-p_{00}\right)-2\vartheta^*\beta_{trt}\left(1+\vartheta^*-\vartheta^* p\right)- \vartheta^{*2}\beta_{X_{PL}}\left(p_{01}-p_{00}\right)+2p\beta_{trt}+2\vartheta^* p\beta_{trt}\Big)$$

$$= \frac{1}{2}\Big(\beta_{trt}\left(2\left(1+\vartheta^*-\vartheta^* p\right)-2\vartheta^*\left(1+\vartheta^*-\vartheta^* p\right)+2p+2\vartheta^* p\right)+ \beta_{X_{PL}}\Big(\ \vartheta^*\left(p_{01}-p_{00}\right)-\vartheta^{*2}\left(p_{01}-p_{00}\right)\Big)\Big)$$

$$= \beta_{trt}\left(1-\vartheta^{*2}+\vartheta^{*2}p+p\right)+\beta_{X_{PL}}\left(\frac{\left(\vartheta^*-\vartheta^{*2}\right)\left(p_{01}-p_{00}\right)}{2}\ \right)$$

Putting the numerator and denominator together, we have:

$$E\left(\hat{\beta}_{MM}\right) = \frac{\beta_{trt}\left(1-\vartheta^{*2}+\vartheta^{*2}p+p\right)+\beta_{X_{PL}}\left(\frac{\left(\vartheta^*-\vartheta^{*2}\right)\left(p_{01}-p_{00}\right)}{2}\ \right)}{1-\vartheta^{*2}+\vartheta^{*2}p+p}$$

And so:

$$E\left(\hat{\beta}_{MM}\right) = \beta_{trt}+\beta_{X_{PL}}\frac{\left(\vartheta^*-\vartheta^{*2}\right)\left(p_{01}-p_{00}\right)}{2\left(1-\vartheta^{*2}+\vartheta^{*2}p+p\right)}$$

4.2.9 S7: Constant treatment effect, differential non-enrolment based on expected outcome in current episode

This is the same scenario as in section 3.3.11 of chapter 3, where there is differential non-enrolment based on the patient's baseline prognosis at episode 2. Full details are available there. Briefly, consider the following data generating mechanism:

$$Y_{ij} = \alpha+\beta_{trt}Z_{ij}+\beta_{X_{EL}}X_{EL_{ij}}+\mu_i+\varepsilon_{ij}$$

where $X_{EL_{ij}}$ is an unobserved binary episode-level variable (i.e. it can vary across episodes).

In this scenario, I will redefine π as $\pi = P\left(X_{EL_{ij}} = 1\right)$, and set this to $\pi = 0.5$. Similarly I will redefine p_{zx} as $p_{zx} = P\left(R_{i2} = 1 | Z_{i1} = z_{i1}, X_{EL_{i2}} = x_{EL_{i2}}\right)$ (so p_{zx} represents the probability of re-enrolment for a second episode based on the patient's episode 1 allocation and their episode 2 value of $X_{EL_{ij}}$). So, for example, $p_{00} = P\left(R_{i2} = 1 | Z_{i1} = 0, X_{EL_{i2}} = 0\right)$, and $p_{01} = P\left(R_{i2} = 1 | Z_{i1} = 0, X_{EL_{i2}} = 1\right)$. As before, I will assume that $p_{00} = p_{11}$ and $p_{01} = p_{10}$.

In this scenario, I will assume for simplicity that all patients experienced two episodes (i.e. $M_i = 2$ for all patients), but that some of these patients were not re-enrolled for their 2nd episode. Therefore, even patients who were enrolled for only a single episode still have a value for $X_{EL_{i2}}$.

From chapter 3, we have:

$$E\left(\hat{\beta}_{B_1}\right) = \frac{\frac{N_T}{2}(1-p)}{\frac{N_T}{2}(1-p)}\left(\beta_{trt} + \frac{\beta_{X_{EL}}}{2} - \frac{\beta_{X_{EL}}}{2}\right) = \beta_{trt}$$

$$E\left(\hat{\beta}_{B_2}\right) = \frac{1}{2}\left(2\beta_{trt} + \beta_{X_{EL}}\left(\frac{p_{00} - p_{01}}{p_{00} + p_{01}}\right)\right) = \beta_{trt} + \frac{\beta_{X_{EL}}}{4}\left(\frac{p_{00} - p_{01}}{p}\right)$$

$$E\left(\hat{\beta}_{w}\right) = \frac{1}{2}\left(2\beta_{trt} - \beta_{X_{EL}}\left(\frac{p_{00} - p_{01}}{p_{00} + p_{01}}\right)\right) = \beta_{trt} - \frac{\beta_{X_{EL}}}{4}\left(\frac{p_{00} - p_{01}}{p}\right)$$

Then, the numerator of $E\left(\hat{\beta}_{MM}\right)$ is:

$$E\left(\hat{\beta}_{MM}\right) = \left((1-p)(1+\vartheta^*)\beta_{trt} + p\left(\beta_{trt} + \frac{\beta_{X_{EL}}}{4}\left(\frac{p_{00} - p_{01}}{p}\right)\right)\right)(1-\vartheta^*) + p(1+\vartheta^*)\left(\beta_{trt} - \frac{\beta_{X_{EL}}}{4}\left(\frac{p_{00} - p_{01}}{p}\right)\right)$$

For the moment, I will focus on the first component of the numerator. If we let $A = \frac{\beta_{X_{EL}}}{4}\left(\frac{p_{00} - p_{01}}{p}\right)$, then the numerator becomes:

$$\left((1-p)(1+\vartheta^*)\beta_{trt} + p(\beta_{trt} + A)\right)(1-\vartheta^*) + p(1+\vartheta^*)(\beta_{trt} - A)$$

$$= \beta_{trt}\left((1-p)(1+\vartheta^*)(1-\vartheta^*) + p(1-\vartheta^*) + p(1+\vartheta^*)\right) + p(1-\vartheta^*)A - p(1+\vartheta^*)A$$

$$= \beta_{trt}\left((1-p)\left(1-\vartheta^{*2}\right) + 2p\right) - 2p\vartheta^* A = \beta_{trt}\left(1 - \vartheta^{*2} + \vartheta^{*2}p + p\right) - 2p\vartheta^* A$$

Putting the numerator and denominator together, and substituting in for A, we have:

$$E\left(\hat{\beta}_{MM}\right) = \frac{\beta_{trt}\left(1 - \vartheta^{*2} + \vartheta^{*2}p + p\right) - 2p\vartheta^{*}\frac{\beta_{X_{EL}}}{4}\left(\frac{p_{00}-p_{01}}{p}\right)}{1 - \vartheta^{*2} + \vartheta^{*2}p + p}$$

And so:

$$E\left(\hat{\beta}_{MM}\right) = \beta_{trt} - \frac{\vartheta^{*}\beta_{X_{EL}}\left(p_{00} - p_{01}\right)}{2\left(1 - \vartheta^{*2} + \vartheta^{*2}p + p\right)}$$

4.2.10 Summary of mathematical results

In table 4.2 I present the results derived previously alongside the estimand for comparison. Mixed-effects models are biased (as defined in section 3.4.1.3) in every setting considered except when the treatment effect is constant. An exception to this is if either $\vartheta^{*} = 0$ (i.e. the expected value of the estimated intraclass correlation is 0) or $p = 0$ (i.e. all patients experience one episode). Additionally, in scenarios 6 and 7, the mixed-effects model estimate will be unbiased if $p_{01} = p_{00}$ (note in these scenarios, if $p = 0$ this implies that $p_{01} = p_{00}$). These conditions (particularly $\vartheta^{*} = 0$ or $p = 0$) do not seem plausible for most re-randomisation trials.

The reason for this bias is due to how the components $\hat{\beta}_{B_1}$, $\hat{\beta}_{B_2}$, and $\hat{\beta}_{W}$ are weighted. The mixed-effects model weights each component by its inverse variance. When the treatment effect is constant, this is useful as it leads to more precise (smaller variance) estimates that are unbiased. However, when the values of these components are different to each other, then the mixed-effects model weights them differently to how they are weighted in the estimand (where each component is weighted by its number of episodes). The mixed-effects model therefore gives too much weight to some components and too little weight to others, which leads to systematic differences to the estimand (i.e. bias). The components $\hat{\beta}_{B_1}$, $\hat{\beta}_{B_2}$, and $\hat{\beta}_{W}$ will generally be different to each other unless the treatment effect is not affected by factors such as j, M_i, or $\widetilde{Z}_{ij}$ (implying the cluster size is non-informative), and there is no differential non-enrolment between treatment groups. These conditions are violated in scenarios S2-S7, and so mixed-effects models are biased.

4.3 Simulation study

I conducted a simulation study to evaluate the performance of mixed-effects models. In particular, I wanted to look at their performance in small sample settings (as the mathematical derivations in the previous section assumed a large sample size), and to assess their performance under more realistic data generating mechanisms than considered in the mathematical derivations in the previous section. I also wanted to look at whether including a causal model for treatment history could reduce or

Table 4.2: Summary of mathematical derivations

Scenario	$E\left(\hat{\beta}_{MM}\right)$	Estimand (per-episode added-benefit)
S1 – Constant treatment effect	β	β
S2 – Treatment effect varies across episode	$\frac{\beta_1\left(1-\vartheta^{*2}+\vartheta^{*2}p\right)+\beta_2 p}{1-\vartheta^{*2}+\vartheta^{*2}p+p}$	$\frac{\beta_1+p\beta_2}{1+p}$
S3 – Treatment effect varies across value of M_i	$\frac{(1-p)\left(1-\vartheta^{*2}\right)\beta_1+2p\beta_2}{1-\vartheta^{*2}+\vartheta^{*2}p+p}$	$\frac{(1-p)\beta_1+2p\beta_2}{(1+p)}$
S4 – Treatment effect carries forward into the second episode	$\beta-\gamma\frac{p\vartheta^*}{1-\vartheta^{*2}+\vartheta^{*2}p+p}$	β
S5 – Treatment becomes less effective on re-use	$\beta+\frac{p\delta(1-\vartheta^*)}{2(1-\vartheta^{*2}+\vartheta^{*2}p+p)}$	$\beta+\frac{p\delta}{2(1+p)}$
S6 – Constant treatment effect, differential non-enrolment based on outcome in previous episode	$\beta_{trt}+\beta_{X_{PL}}\frac{\left(\vartheta^*-\vartheta^{*2}\right)(p_{01}-p_{00})}{2(1-\vartheta^{*2}+\vartheta^{*2}p+p)}$	β_{trt}
S7 – Constant treatment effect, differential non-enrolment based on expected outcome in current episode	$\beta_{trt}-\frac{\vartheta^*\beta_{X_{EL}}(p_{00}-p_{01})}{2(1-\vartheta^{*2}+\vartheta^{*2}p+p)}$	β_{trt}

eliminate bias, as this has been shown to be effective in certain scenarios (e.g. including a term for $Z_{i,j-1}$ in the model eliminates bias when the treatment effect carries forward, such as under data generating model 2.2 [9]).

In this simulation study, I focus on the per-episode added-benefit and policy-benefit estimands. I decided not to evaluate the per-patient estimands as the use of weights in mixed-effects models can be complicated (for instance, weights cannot be specified for a linear mixed-effects model in Stata if restricted maximum likelihood is used, which is the recommended method for obtaining estimates from mixed-effects models with small sample sizes [52, 53]). Further, I found the results from the simulation study on the per-episode estimators to be fairly conclusive, and in line with the mathematical derivations shown earlier, and felt it was reasonable to extrapolate these results to mixed-effects models with weights (i.e. I felt it was unlikely that adding a weight to the model would reduce the bias found from the unweighted per-episode model).

I used the same general simulation scenarios and data generating mechanisms as in chapter 3, except I omitted simulation study 2b (i.e. I only performed simulation studies 1 and 2a); this was because I felt the results from studies 1 and 2a were conclusive, and so there was no need for further evaluation in study 2b. The full methods are available in chapter 3, but I summarise them briefly below. The only difference in the data generating mechanism from chapter 3 was the value of the variance parameters (σ^2_μ and σ^2_ε); in chapter 3 I used values of 5 for both, leading to an ICC of 0.5. Mixed-effects models are more biased for larger ICCs (as shown in

sections 4.2.4-4.2.9), and in this simulation study I wanted to evaluate how poorly these models might perform under a large (but plausible) value. Therefore, I used values of $\sigma^2_\mu = 7.5$ and $\sigma^2_\varepsilon = 2.5$ to give an ICC of 0.75.

As before, I used 10,000 replications for all scenarios.

4.3.1 Estimands

As discussed above, I used the per-episode added-benefit and per-episode policy-benefit estimands. Values for each of these estimands for each simulation scenario are provided in chapter 3.

4.3.2 Methods of analysis

For the per-episode policy-benefit estimand, I used the following mixed-effects model:

$$Y_{ij} = \hat{\alpha} + \hat{\beta} Z_{ij} + \hat{\gamma} Z_{i,j-1} + \hat{\delta} Z_{ij} Z_{i,j-1} + \hat{\beta}_{ep} X_{ep_{ij}} + \mu_i + \varepsilon_{ij} \tag{4.2}$$

Following on from section 3.1.3 in chapter 3, an overall estimate of the treatment effect is then calculated as: $\frac{N_1}{M_T}\left(\hat{\beta}\right) + \frac{N_2}{M_T}\left(\hat{\gamma} + \hat{\beta} + \hat{\delta}\right)$.

For the per-episode added-benefit estimand, I used two different mixed-effects models. The first was analysis model 1.2 (this is referred to as MM in results). The second model specified the same causal model for the effect of treatment history ($\widetilde{Z}_{ij}$) on the outcome as model 4.2 for the policy-benefit estimator. The purpose of this model was to evaluate whether including a causal model for treatment history could reduce bias. This model is referred to as $MM(adj)$ in results (to denote it is adjusted for treatment history). An overall estimate of the treatment effect is then calculated as:

$$\frac{N_1}{M_T}\left(\hat{\beta}\right) + \frac{N_{2,\widetilde{Z}=(0)}}{M_T}\left(\hat{\beta}\right) + \frac{N_{2,\widetilde{Z}=(1)}}{M_T}\left(\hat{\beta} + \hat{\delta}\right)$$

where $N_{j,\widetilde{Z}=\widetilde{z}}$ denotes the number of patients enrolled at episode j with treatment history $\widetilde{Z} = \widetilde{z}$ (i.e. $N_{2,\widetilde{Z}=(0)}$ denotes the number of patients enrolled at episode 2 who were allocated to control in their first episode, and $N_{2,\widetilde{Z}=(1)}$ is the number allocated intervention in their first episode). The first term ($\frac{N_1}{M_T}\left(\hat{\beta}\right)$) is the treatment effect component for the first episode, the second term $\frac{N_{2,\widetilde{Z}=(0)}}{M_T}\left(\hat{\beta}\right)$ is the treatment effect component at the second episode for patients who received control in their first episode, and $\frac{N_{2,\widetilde{Z}=(1)}}{M_T}\left(\hat{\beta} + \hat{\delta}\right)$ is the component at the second episode for patients who received intervention in their first episode. Note that the reason the added-benefit and policy-benefit estimators based on model 4.2 calculate the 2nd episode results differently is the added-benefit effect calculates separate treatment effects depending on treatment history, whereas the policy-benefit effect does not allow for different treatment histories

Table 4.3: Stata code to implement mixed-effects models and independence estimators in simulation study. 'y' denotes patient outcome, 'z' denotes treatment allocation, 'id' is a unique identifier for patient, 'm_i' denotes the number of episodes for which the patient is enrolled in the trial, 'z_prev' denotes the patient's treatment allocation in their previous episode (and is set to 0 if it is the patient's first episode), 'x_ep' denotes the episode (0=first episode, 1=second episode), and 'prop_1st_ep' and 'prop_2nd_ep' represent the proportion of episodes in the trial which are 1st and 2nd episodes respectively, and 'prop_ep2_prevz0' and 'prop_ep2_prevz1' represent the proportion of episodes which are 2nd episodes where treatment allocation in the previous episode was 0 and 1 respectively. In order to run the above code in Stata, 'prop_1st_ep', 'prop_2nd_ep', 'prop_ep2_prevz0', and 'prop_ep2_prevz1' must be saved as Stata local macros.

Estimator	Stata code
Per-episode, added-benefit	
Mixed-effects model (MM)	mixed y z || id:, iterate(50) reml
Mixed-effects model (MM [adj])	mixed y z##z_prev x_ep || id:, iterate(50) reml lincomest 'prop_1st_ep'*_b[1.z] + 'prop_ep2_prevz0'*_b[1.z] + 'prop_ep2_prevz1'*(_b[1.z] + _b[1.z#1.z_prev])
Independence estimator	reg y z, vce(cluster id)
Per-episode, policy-benefit	
Mixed-effects model	mixed y z##z_prev x_ep || id:, iterate(50) reml lincomest 'prop_1st_ep'*_b[1.z] + 'prop_2nd_ep'*(_b[1.z]+_b[1.z_prev] + _b[1.z#1.z_prev])
Independence estimator	reg y z##z_prev x_ep, vce(cluster id) lincomest 'prop_1st_ep'*_b[1.z] + 'prop_2nd_ep'*(_b[1.z]+_b[1.z_prev] + _b[1.z#1.z_prev])

(i.e. it calculates the treatment effect for intervention in the current and all previous episodes vs. control in the current and all previous episodes).

The Stata code used to implement these mixed-effects models is shown in table 4.3. All mixed-effects models were estimated using restricted maximum likelihood [52, 53]. For each simulation study I also included the relevant independence estimator from chapter 3 for comparison (e.g. for the policy-benefit estimand I include the same independence policy-benefit estimator used in chapter 3, and similarly for the added-benefit estimand).

4.3.3 Performance measures

As before, my main criterion for evaluating estimators was bias (as defined in chapter 3). I also evaluated coverage of the 95% confidence intervals (defined in chapter 3), and precision [27]. I evaluated the precision of the mixed-effects models against the independence estimators. The relative % increase in precision for a mixed-effects model vs. an independence estimator is defined as:

$$100\left(\frac{V\left(\hat{\beta}_{IND}\right)}{V\left(\hat{\beta}_{MM}\right)}-1\right)$$

Where $\hat{\beta}_{IND}$ represents the independence estimator, and $V\left(\hat{\beta}_{MM}\right)$ and $V\left(\hat{\beta}_{IND}\right)$ represent the empirical variance estimates of the mixed-effects model and independence estimators respectively (and are estimated as the variance of the treatment effect estimates across all replications). Therefore, values >1 indicate that mixed-effects models are giving more precise estimates than independence estimators.

4.3.4 Simulation study 1: patients enrolled for all episodes they experience

Full details are provided in chapter 3; I briefly summarise the key information here. This simulation study is based on a trial of 300 patients; 150 patients experience one episode during the trial period, and 150 experience two episodes. I used the same six different data generating mechanisms as in chapter 3 (apart from the size of the ICC, as described above), which are:

1. Constant treatment effect
2. Treatment effect varies across episode
3. Treatment effect varies across patients with different values of M_i
4. Treatment effect carries forward into the 2nd episode
5. Treatment becomes less effective on re-use
6. Treatment effect varies across episodes, across patients with different values of M_i, carries forward, and becomes less effective on re-use

Continuous outcomes were generated based on model 3.13 from chapter 3:

$$Y_{ij}=\alpha+\beta_{trt}Z_{ij}+\beta_{ep}X_{ep_{ij}}+\beta_{M}X_{M_i}+\beta_{TRTxEP}Z_{ij}X_{ep_{ij}}+\beta_{TRTxM}Z_{ij}X_{M_i}+\gamma Z_{i,j-1}+\delta Z_{ij}Z_{i,j-1}+\mu_i+\varepsilon_{ij}$$

where $X_{ep_{ij}}$ is an indicator variable for episode 2, and X_{M_i} is an indicator variable for patients with $M_i = 2$. For each scenario, I set $\alpha = 0$, $\beta_{trt} = 3$, $\beta_{ep} = 1$, $\beta_M = 1$, $\sigma^2_\mu = 7.5$ and $\sigma^2_\varepsilon = 2.5$. I generated μ_i and ε_{ij} independently; based on the chosen variances, the intraclass correlation between episodes from the same patient is 0.75 (conditional on the other variables in the data generating model).

Values for other parameters are shown in table 3.15 in chapter 3.

4.3.5 Simulation study 2a: some patients do not re-enrol for their 2nd episode

Full details are provided in chapter 3; I briefly summarise the key information here. As before, this simulation study is based on a trial of 300 patients; 150 patients experience one episode during the trial period, and 150 experience two episodes. All patients are enrolled for their first episode, but a subset of patients who experience two episodes do not re-enrol for their second episode. Therefore, $N_T = 300$ and $M_{T(1)} = 150$, however $M_{T(2)} < 150$ and $M_T < 450$; the exact values of $M_{T(2)}$ and M_T vary across simulation replications.

I used the same six data generating mechanisms, and the same five non-enrolment scenarios as in chapter 3 (apart from the size of the ICC, as described above). The data generating mechanisms were listed in the previous section, and the non-enrolment scenarios are:

1. Non-enrolment depends on previous treatment allocation
2. Non-enrolment depends on previous treatment allocation and previous outcome
3. Non-enrolment depends on previous treatment allocation and baseline prognosis at episode 2
4. Non-enrolment is differential between treatment groups based on previous outcome
5. Non-enrolment is differential between treatment groups based on baseline prognosis at episode 2

I simulated data by first generating outcomes for all 450 episodes (regardless of whether they were enrolled in the trial) using model 3.14 from chapter 3:

$$Y_{ij} = \alpha + \beta_{trt} Z_{ij} + \beta_{ep} X_{ep_{ij}} + \beta_M X_{M_i} + \beta_{TRTxEP} Z_{ij} X_{ep_{ij}} + \beta_{TRTxM} Z_{ij} X_{M_i} + \gamma Z_{i,j-1} + \delta Z_{ij} Z_{i,j-1} + \beta_{X_{PL}} X_{PL_i} + \beta_{X_{EL}} X_{EL_{ij}} + \mu_i + \varepsilon_{ij}$$

where $X_{ep_{ij}}$ is an indicator variable for episode 2, X_{M_i} is an indicator variable for patients with $M_i = 2$, and X_{PL_i} and $X_{EL_{ij}}$ are unobserved binary covariates, with

X_{PL_i} being a patient-level covariate which does not vary across episodes, and $X_{EL_{ij}}$ being an episode-level covariate which can vary across episodes for the same patient. Values for each parameter are shown in tables 3.16 and 3.17 in chapter 3 (except for σ^2_μ and σ^2_ε, which are 7.5 and 2.5 respectively, as described in the previous section).

I then generated an indicator for each episode to denote whether it was enrolled in the trial or not using model 3.15 from chapter 3:

$$P\left(R_{i2}=0\right)=\alpha^{R_2}+\gamma^{R_2}Z_{i,j-1}+\beta^{R_2}_{X_{PL}}X_{PL_i}+\beta^{R_2}_{X_{EL}}X_{EL_{i2}}+ \\ \delta^{R_2}_{Xpl}Z_{i,j-1}X_{PL_i}+\delta^{R_2}_{Xel}Z_{i,j-1}X_{EL_{i2}}$$

where R_{i2} denotes whether the patient was enrolled at episode 2 (where 1=enrolled, 0=not enrolled).

I then performed analysis only on the subset of enrolled episodes. As before, I set $\alpha^{R_2}=0.05$ and $\gamma^{R_2}=0.10$ for all scenarios; values for the parameters $\beta_{X_{PL}}$, $\beta_{X_{EL}}$, $\delta^{R_2}_{Xpl}$, and $\delta^{R_2}_{Xel}$ are shown in table 3.17 in chapter 3.

4.3.6 Simulation study (results)

4.3.6.1 Simulation study 1: patients enrolled for all episodes they experience

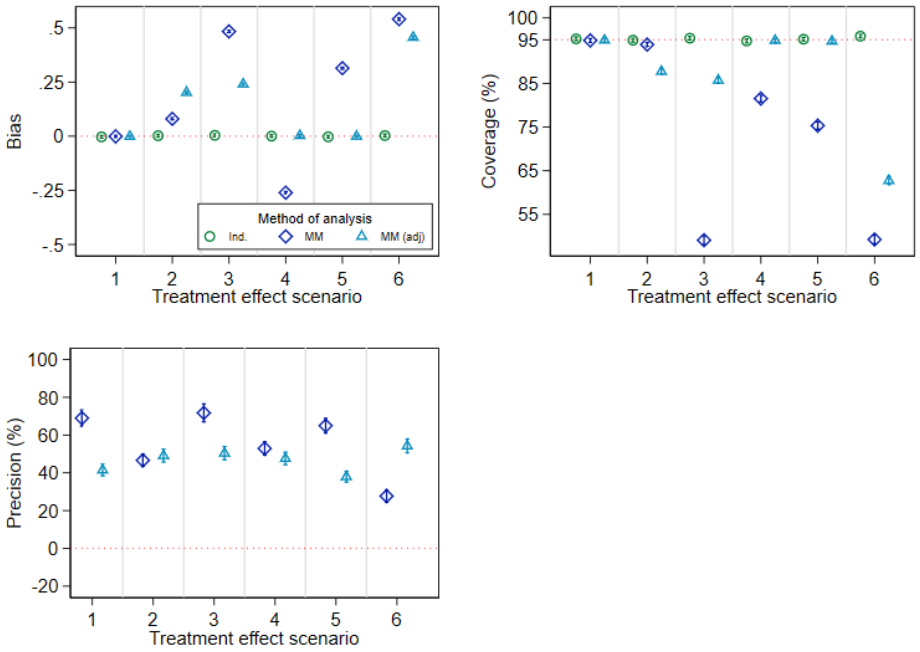
Results for the per-episode added-benefit estimand are shown in figure 4.1. Mixed-effects models had much higher precision than independence estimators across all scenarios. However, they also had high levels of bias in most settings. The unadjusted mixed-effects model was biased in all settings except when the treatment effect was constant (scenario 1); the adjusted model removed bias in scenarios 4 and 5, but was still biased in all other settings. This led to severe under-coverage in most settings.

Results for the per-episode policy-benefit estimand are shown in figure 4.2. Mixed-effects models had higher precision than independence estimators, but were biased when the treatment effect varied by episode (scenarios 2 and 6). Coverage was good except in scenarios where the treatment effect estimate was biased.

4.3.6.2 Simulation study 2a: some patients do not re-enrol for their 2nd episode

Results for the per-episode added-benefit estimand are shown in figures 4.3 to 4.5. Mixed-effects models were biased across most scenarios. Under certain non-enrolment scenarios, bias occurred even when the treatment effect was constant. Interestingly, bias was minimal under non-enrolment scenario 3. This is likely because inclusion of $X_{EL_{ij}}$ in the data generating model for the outcome acts to reduce the ICC by a substantial amount. Because the amount of bias depends on the size of the ICC, a reduction in the ICC leads to lower bias in these scenarios. The reason the inclusion of $X_{EL_{ij}}$ serves to reduce the ICC is that it is an additional source of unexplained variation which is

Figure 4.1: Bias, coverage, and precision of mixed-effects models compared to independence estimators for per-episode added-benefit effect in simulation study 1 (no non-enrolment). Ind=independence estimator. MM=mixed-effects model. MM (adj)=mixed-effects model adjusted for treatment history. Error bars are 95% confidence intervals based on Monte Carlo standard errors.



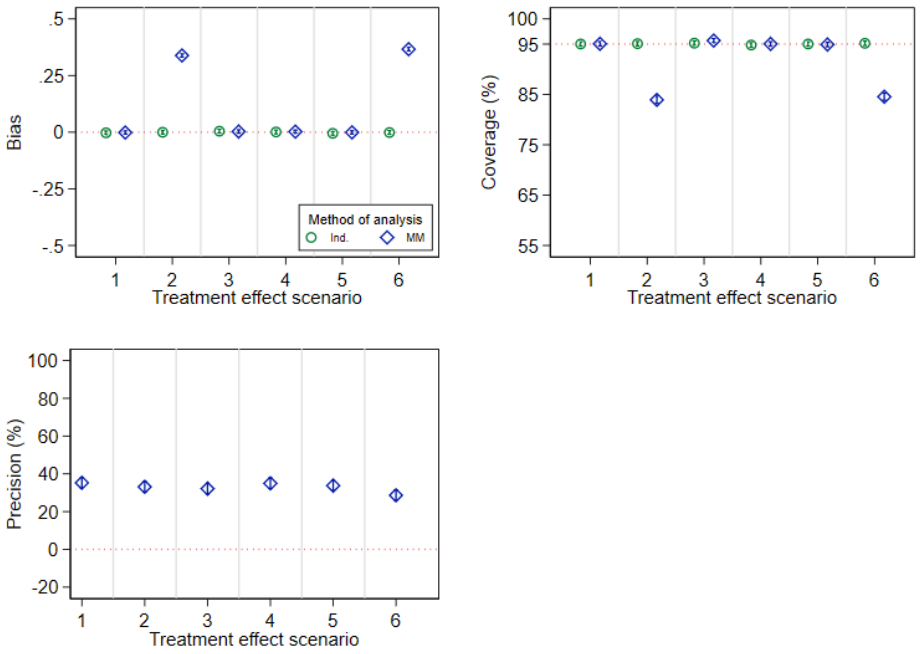
subsumed into the estimate of σ^2_ε. It contributes to σ^2_ε rather than σ^2_μ because it is not a patient-level factor (i.e. it varies across episodes and is independent within patients). The inclusion of $X_{EL_{ij}}$ also serves to decrease the ICC in non-enrolment scenario 5, although bias is not reduced by as much, possibly because the non-enrolment is differential between treatment groups in this scenario.

Similarly, the inclusion of X_{PL_i} serves to increase the ICC in non-enrolment scenarios 2 and 4, because the added variability due to X_{PL_i} is subsumed in σ^2_μ (because X_{PL_i} is a patient level factor). Therefore, bias is higher in these scenarios than in non-enrolment scenario 1, where X_{PL_i} is not included.

Coverage was valid except when estimators were biased, in which case it was too low. As before, mixed-effects models had higher precision than independence estimators. However, there was minimal difference under non-enrolment scenarios 3 and 5. This is also due to the decrease in the ICC from $X_{EL_{ij}}$, as the increase in precision from mixed-effects models depends on the size of the ICC [51].

Results for the per-episode policy-benefit estimand are shown in figures 4.6 to 4.8. The mixed-effects model was biased across a number of scenarios. Interestingly it was less biased than independence estimators in non-enrolment scenario 4 under treatment effect scenarios 1, 4, and 5, though in all other scenarios, independence estimators have

Figure 4.2: Bias, coverage, and precision of mixed-effects models to independence estimators for per-episode policy-benefit effect in simulation study 1 (no non-enrolment). Ind=independence estimator. MM=mixed-effects model. Error bars are 95% confidence intervals based on Monte Carlo standard errors.



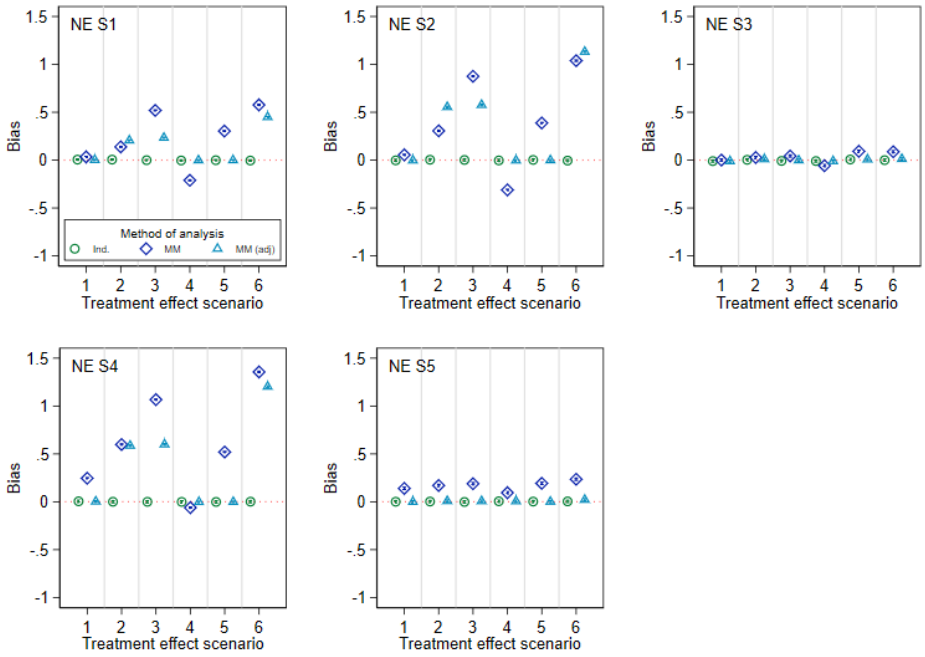
equivalent or smaller bias. It is not clear why mixed-effects models were less biased in these scenarios. One possible explanation is that the bias for independence estimators comes from bias in the estimate of γ; it may be that mixed-effects models are able to obtain valid estimates of γ in these settings due to the within-patient comparison, and so are less biased.

Under non-enrolment scenario 5 all methods are similarly biased. Coverage of mixed-effects models was valid except in scenarios where there was bias. Mixed-effects models were much more precise than independence estimators.

4.3.6.3 Discussion

In this chapter I evaluated the use of mixed-effects models with a random intercept in re-randomisation trials. As expected, these methods offered much higher precision compared to independence estimators. However, this came at the expense of increased bias in most settings. The unadjusted mixed-effects model was biased for the added-benefit estimand in all settings except those where the treatment effect was constant across patients and episodes, and non-enrolment was non-differential between treatment arms. Adjusting for treatment history reduced bias for some, but not all scenarios. Mixed-effects models performed better for the policy-benefit estimand, but were still biased in a number of scenarios. They were generally more biased than independence

Figure 4.3: Bias of mixed-effects models compared to independence estimators for per-episode added-benefit effect in simulation study 2a (with non-enrolment). NE=Non-enrolment scenario. Ind=independence estimator. MM=mixed-effects model. MM (adj)=mixed-effects model adjusted for treatment history. Error bars are 95% confidence intervals based on Monte Carlo standard errors.



estimators except in a couple of instances.

I therefore recommend that mixed-effects models not be used in re-randomisation trials, as their potential risks far outweigh their potential benefits. This is particularly the case as their gain in precision is most pronounced when the ICC is high, however this is also when they are most biased. This result follows on from the informative cluster size literature, which has shown that mixed-effects models are not appropriate when cluster size is informative [45].

Therefore, independence estimators are a preferred option, as they are either unbiased or less biased in most scenarios. As such, independence estimators should be the default method of analysis in re-randomisation trials.

Figure 4.4: Coverage of mixed-effects models compared to independence estimators for per-episode added-benefit effect in simulation study 2a (with non-enrolment). NE=Non-enrolment scenario. Ind=independence estimator. MM=mixed-effects model. MM (adj)=mixed-effects model adjusted for treatment history. Error bars are 95% confidence intervals based on Monte Carlo standard errors.

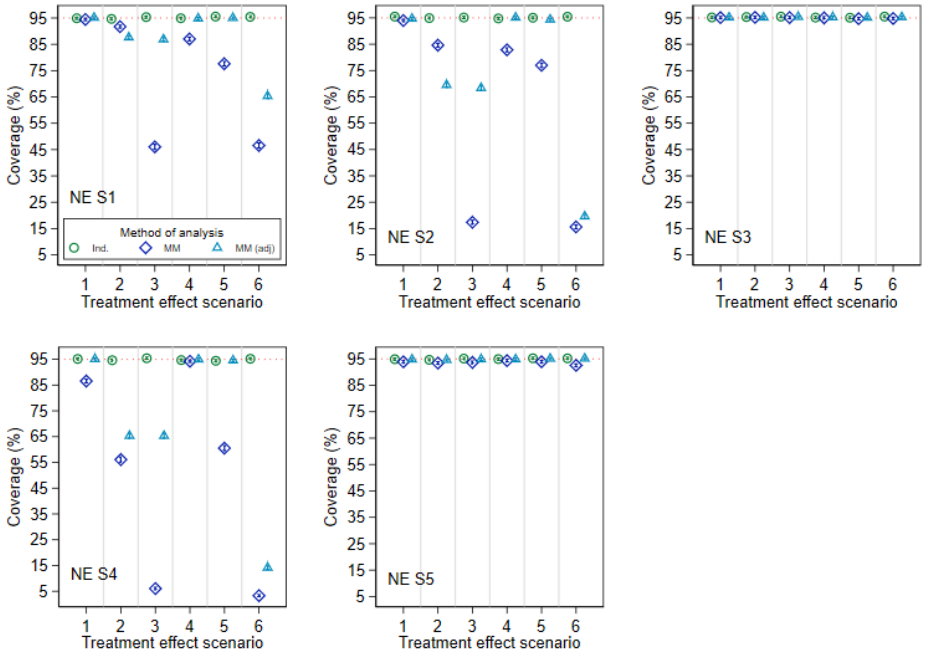


Figure 4.5: Precision of mixed-effects models compared to independence estimators for per-episode added-benefit effect in simulation study 2a (with non-enrolment). NE=Non-enrolment scenario. MM=mixed-effects model. MM (adj)=mixed-effects model adjusted for treatment history. Error bars are 95% confidence intervals based on Monte Carlo standard errors.

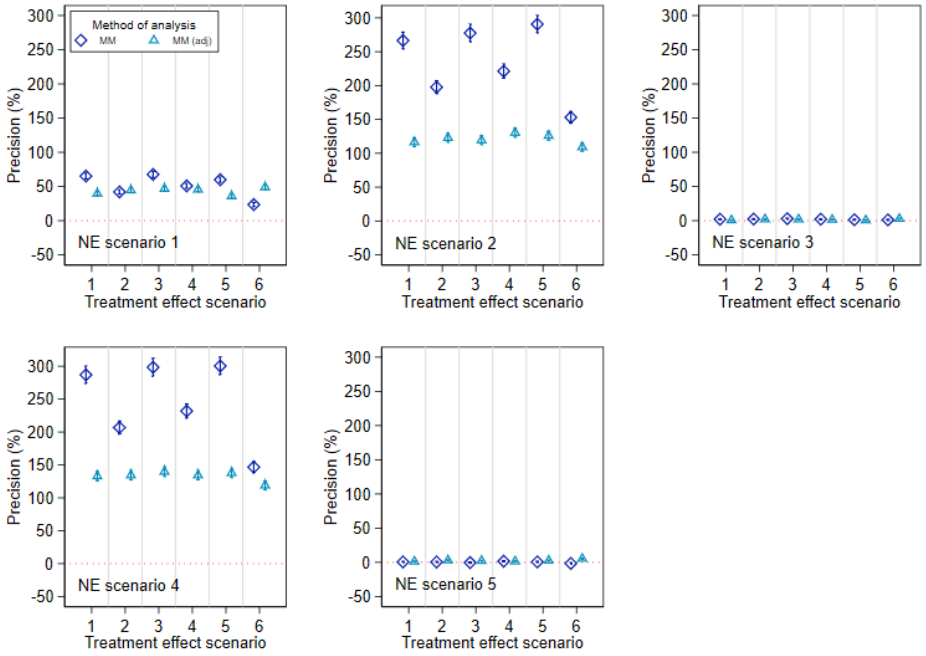


Figure 4.6: Bias of mixed-effects models compared to independence estimators for per-episode policy-benefit effect in simulation study 2a (with non-enrolment). NE=Non-enrolment scenario. Ind=independence estimator. MM=mixed-effects model. Error bars are 95% confidence intervals based on Monte Carlo standard errors.

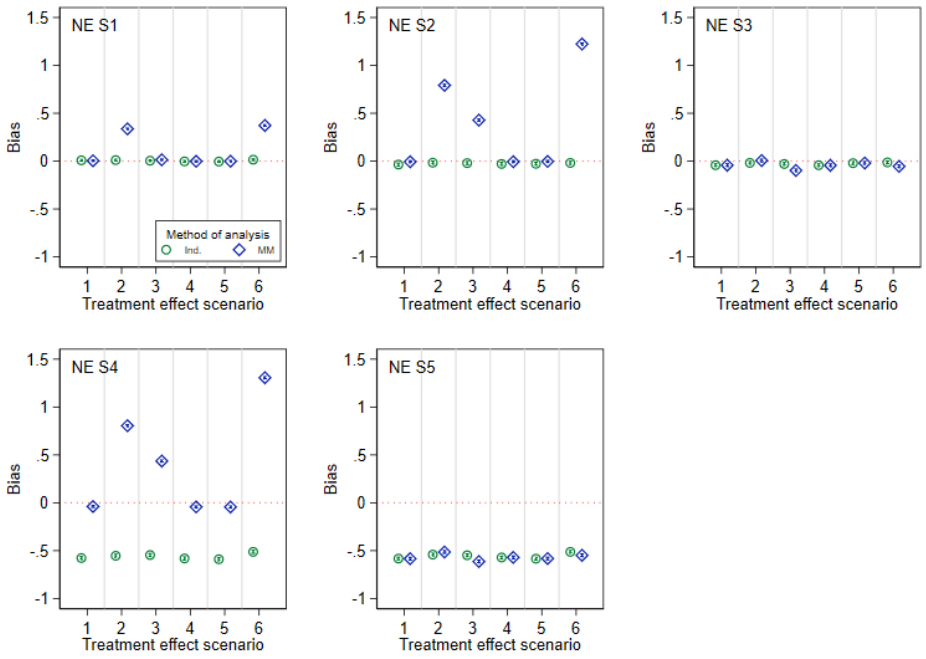


Figure 4.7: Coverage of mixed-effects models compared to independence estimators for per-episode policy-benefit effect in simulation study 2a (with non-enrolment). NE=Non-enrolment scenario. Ind=independence estimator. MM=mixed-effects model. Error bars are 95% confidence intervals based on Monte Carlo standard errors.

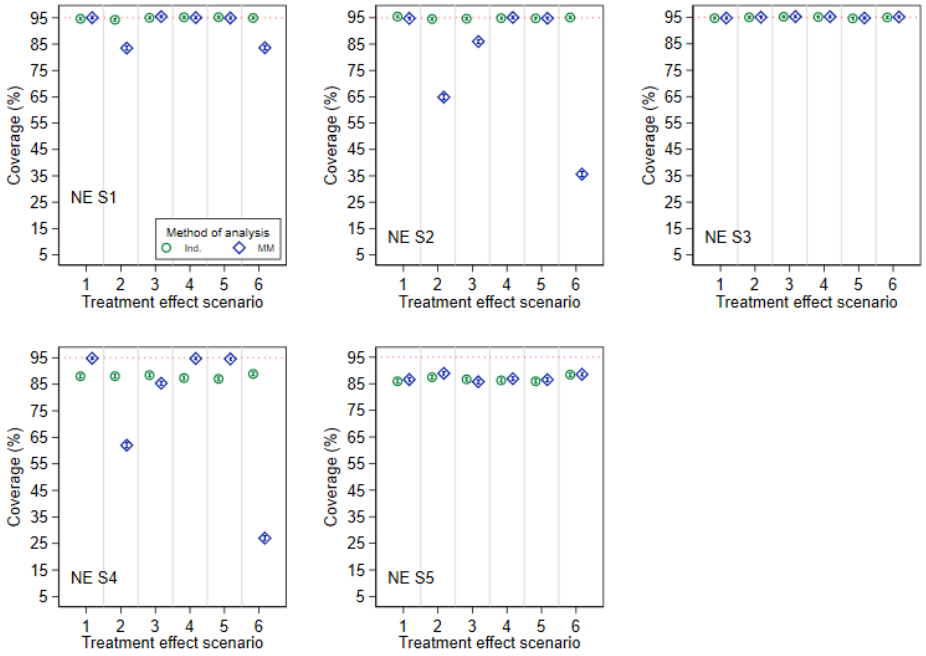
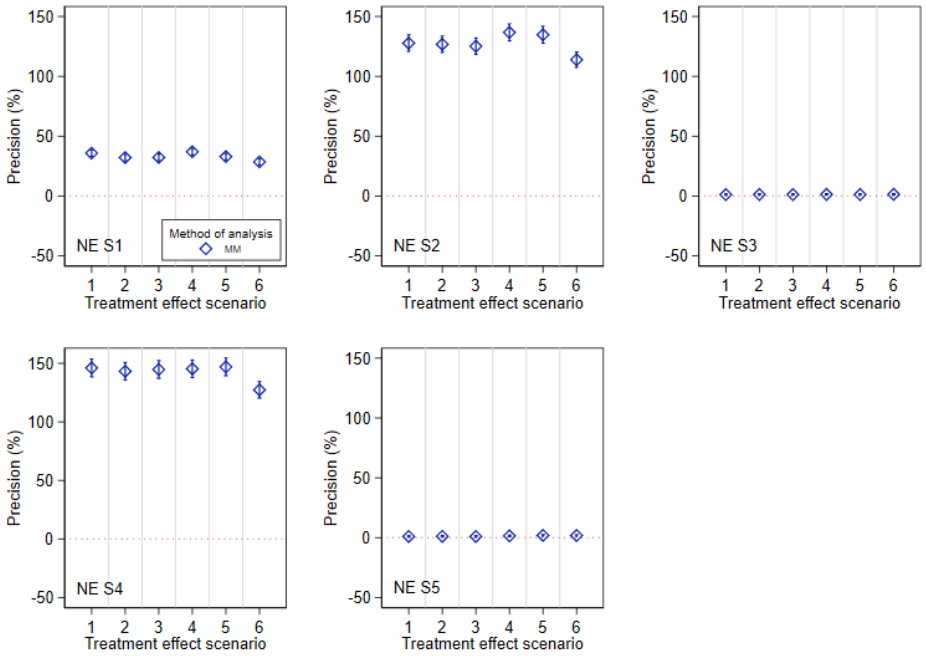


Figure 4.8: Precision of mixed-effects models compared to independence estimators for per-episode policy-benefit effect in simulation study 2a (with non-enrolment). NE=Non-enrolment scenario. MM=mixed-effects model. Error bars are 95% confidence intervals based on Monte Carlo standard errors.



5 *Comparing re-randomisation with other designs*

5.1 Overview of other designs

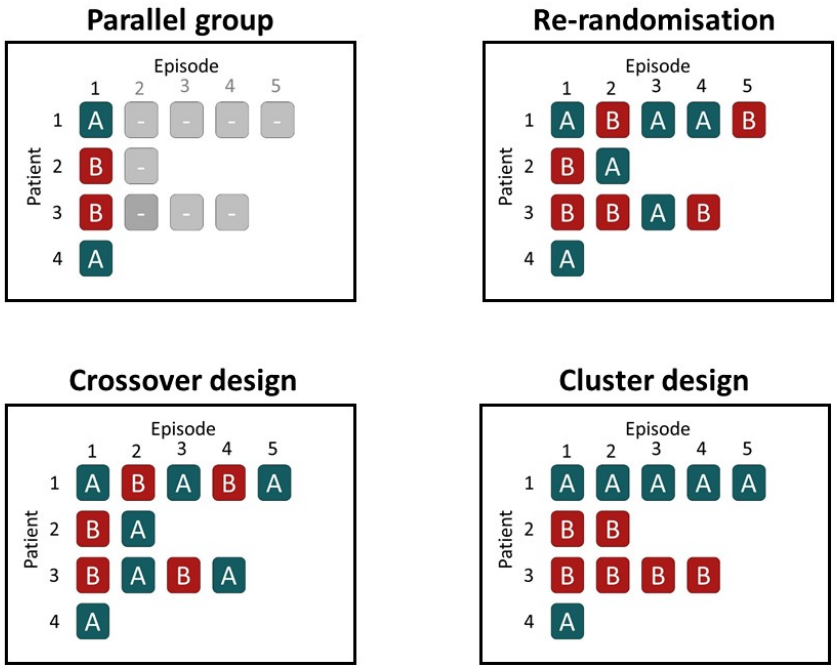
In this chapter, I will compare re-randomisation with other trial designs that could be used in multi-episode settings. The three other designs I consider are (i) parallel group designs; (ii) crossover designs; and (iii) cluster designs. These designs are shown in figure 5.1. I discuss each of these designs below.

5.1.1 Parallel group design

5.1.1.1 Overview

This trial design was briefly discussed in the introduction of this thesis, and its basic outline is shown in figure 5.1. In a parallel group trial, patients can be enrolled for

Figure 5.1: Overview of re-randomisation, cluster, crossover, and parallel group designs



only a single episode; they are not eligible to be re-enrolled for subsequent treatment episodes.

5.1.1.2 Potential benefits

One benefit of the parallel group design is its simplicity; restricting to one episode per patient leads to a much simpler trial. Unlike for a re-randomisation trial, we do not need to worry about which estimand to estimate, as the effect of treatment the first time it is used (the episode 1 effect) is the only estimand available. We also do not need to worry about how to handle the additional episodes in the analysis (for instance, how to weight the episodes or how to specify a causal model for a patient's treatment history).

If our only aim is to estimate the effect of an intervention the first time it is used, then it is unlikely any design will be better than a parallel group trial. Although the re-randomisation design can be used to estimate the episode specific treatment effects, it will usually have lower precision than a parallel group design for the episode 1 effect. For example, for a trial that requires 300 observations, the parallel group design would enrol 300 individual patients whereas a re-randomisation trial may recruit 300 episodes from 200 patients. The parallel group design would then use 300 observations to estimate the episode 1 effect, whereas the re-randomisation trial would use 200 observations for this estimate.

5.1.1.3 Potential downsides

Although the parallel group design is very good at estimating the effect of an intervention the first time it is used, it cannot estimate other treatment effects (e.g. the effect in subsequent episodes or an average effect across all episodes) without making very strong assumptions. The parallel group design will only provide unbiased estimates for other estimands, such as policy-benefit or added-benefit effects, if the treatment effect in the $j = 1$ episodes is the same as the treatment effect in the $j > 1$ episodes. If this is not the case then estimates from a parallel group design will generally be biased. It is worth noting that this assumption cannot be assessed from the data collected in a parallel group design, so the decision to use a parallel group design for these estimands would rely on strong, untestable assumptions.

Another drawback to a parallel group design is the potential loss of precision through exclusion of patient's subsequent episodes. The parallel group trial will either recruit fewer observations over the same time frame as a re-randomisation design, or will recruit an equivalent number of observations over a longer time period. For example, consider a setting where 200 patients experience 300 episodes over a 4 year period (where 100 patients experience one episode and 100 patients experiencing two episodes in that time). Over the 4 year period the parallel group trial would recruit 200 patients while the re-randomisation design would recruit 300 episodes. Alternatively,

the parallel group trial could continue recruiting until it had enrolled 300 patients, but this would likely take an extra 2 years.

5.1.2 Cluster design

5.1.2.1 Overview

The basic outline of the cluster design is shown in figure 5.1. In this design, patients act as the cluster, and are allocated to a treatment arm for all episodes they experience (i.e. instead of randomising episodes to treatments, the cluster design allocates all episodes within a patient to the same treatment). Note that this design differs to cluster trials where a hospital or medical site acts as the cluster, and all patients within the cluster receive the same treatment; instead, the patients themselves are the clusters, and different patients in the same hospital or medical site may be allocated to different treatment strategies.

5.1.2.2 Potential benefits

The design of the cluster trial matches the policy-benefit estimand, as patients are allocated to a treatment policy for all episodes. Therefore estimation of the policy-benefit effect is very simple in a cluster design, and is just calculated as a difference in means between groups (for the per-episode estimand); there is no need to specify a causal model for treatment history as for a re-randomisation trial. Estimation of the policy-benefit effect will therefore be much more robust in a cluster design, as there is no chance of bias from a misspecified causal model.

5.1.2.3 Potential downsides

One potential downside of the cluster design is that it cannot estimate the added-benefit effect, as this requires a comparison between intervention and control patients who share the same treatment history (which cannot happen in a cluster trial). However, this is only a downside if we are interested in the added-benefit effect.

Another downside is that this design does not maintain allocation concealment after the first episode. Because patients are randomised to the same treatment for all episodes, if they receive the intervention in episode 1 they will know that they will also receive the intervention if they re-enrol for their second episode (or third, or fourth, etc). The patient may decide not to re-enrol based on the fact they will get the intervention, or the clinician or research staff may decide not to ask the patient to re-enrol on this basis. This type of selection bias is a common concern in cluster randomised trials which enrol patients after randomisation, and can lead to bias in the estimated treatment effects [54, 55, 56, 57]. The lack of allocation concealment may be less of an issue in double-blinded trials, or in trials where patients and the research staff enrolling patients are blinded [56].

5.1.3 Crossover design

5.1.3.1 Overview

The basic outline of the crossover design is shown in figure 5.1. Patients are allocated to a treatment arm for their first episode, and then receive the opposite treatment in their second episode (i.e. they cross over to the other treatment) [58, 59, 60].

5.1.3.2 Potential benefits

The main benefit of the crossover design is that it provides a within-patient comparison for all patients with >1 episode. This will lead to much higher precision compared to designs which allow some patients with >1 episode to be allocated to the same treatment for both episodes (i.e. the re-randomisation and cluster designs).

5.1.3.3 Potential downsides

The main issue with the crossover design is that it is not designed to match any particular estimand. For instance, the design of a re-randomisation trial matches the added-benefit estimand and the design of a cluster trial matches the policy-benefit estimand, so these trials can provide unbiased estimates of these estimands under simple assumptions. However, the design of a crossover trial does not match either of these estimands (or any other useful estimand that I can think of) and so the crossover trial will only be unbiased for these estimands under quite strong assumptions (i.e. that treatment history has no impact on the outcome or treatment effect in the current episode). The reason for this is that estimation of the policy-benefit effect requires patients who receive the same treatment in consecutive episodes, and estimation of the added-benefit estimand requires a comparison between intervention and control patients who share the same treatment history; neither of these requirements is met in a crossover design. Furthermore, the assumption that treatment history does not affect current episode cannot be assessed from the data in a crossover trial, and so using a crossover design for these estimands will rely on strong untestable, assumptions.

Another downside is that, like the cluster design, crossover trials do not maintain allocation concealment after the first episode (as patients and research staff know patients will receive the opposite treatment for their next episode). Therefore crossover designs are also susceptible to selection bias; however, this selection bias may act in a different manner than in cluster designs, as patients and clinicians will know they will receive the opposite treatment in their next episode, rather than the same treatment.

5.2 Mathematical derivation of bias

In this section I will evaluate the asymptotic bias of the cluster and crossover designs in multi-episode settings, for both independence estimators and mixed-effects models with a random-intercept for patient (hereafter referred to as 'mixed-effects models'). I will evaluate the cluster design against the per-episode policy-benefit estimand, and

the crossover design against the per-episode added-benefit estimand; I chose not to evaluate the crossover design against the policy-benefit estimand as there is such an obvious mismatch between the two that it is clear the crossover trial should not be used for this estimand. I also do not evaluate the bias of the parallel group design as it is obvious that it will be biased in almost all settings where the treatment effect is not constant across patients and episodes.

As before, I will restrict the setting to a trial with a 1:1 allocation ratio, where patients experience a maximum of two episodes, and I will assume that treatment allocation does not affect the occurrence of subsequent episodes.

I will use the same method of derivation as in chapters 3 and 4 (i.e. based on the between- and within-patient estimation components $\hat{\beta}_{B_1}$, $\hat{\beta}_{B_2}$, and $\hat{\beta}_W$), and will use the same notation; for instance, in the section for crossover designs, the parameter ϑ^* will refer to the expected value of the estimated intraclass correlation coefficient based on the mixed-effects model applied to the crossover design.

I will use the same seven scenarios as in chapters 3 and 4:

- S1 – Constant treatment effect
- S2 – Treatment effect varies across episode
- S3 – Treatment effect varies across value of M_i
- S4 – Treatment effect carries forward into the second episode
- S5 – Treatment becomes less effective on re-use
- S6 – Constant treatment effect, differential non-enrolment based on outcome in previous episode
- S7 – Constant treatment effect, differential non-enrolment based on expected outcome in current episode

Further explanation for each scenario (along with a formula for the data generating mechanism) can be found in chapters 3 and 4.

In section 5.2.1 I will discuss the number of patients in each treatment sequence for the cluster and crossover designs, then in section 5.2.2 I will discuss the estimation components $\hat{\beta}_{B_1}$, $\hat{\beta}_{B_2}$ and $\hat{\beta}_W$. In sections 5.2.3-5.2.6 I will derive estimators for independence and mixed-effects models for cluster and crossover trials, and will evaluate the expected value of these estimators under the seven scenarios listed above in sections 5.2.7-5.2.13. In section 5.2.14 I will compare the expected values of these estimators against the estimand for each scenario to assess bias.

Table 5.1: Asymptotic number of patients in each treatment sequence in a cluster trial with 1:1 allocation ratio

	Treatment allocation		**Number of patients at each episode**	
Sequence	**Episode 1**	**Episode 2**	**Episode 1**	**Episode 2**
$Z = (1)$	1	-	$\frac{N_T}{2}(1-p)$	-
$Z = (1,1)$	1	1	$\frac{N_T p}{2}$	$\frac{N_T p}{2}$
$Z = (0)$	0	-	$\frac{N_T}{2}(1-p)$	-
$Z = (0,0)$	0	0	$\frac{N_T p}{2}$	$\frac{N_T p}{2}$

5.2.1 Number of patients at each episode for different sequences

As in chapter 3, we need to calculate the number of patients and episodes in each treatment sequence for the cluster and crossover designs, as these are used to calculate the components $\hat{\beta}_{B_1}$, $\hat{\beta}_{B_2}$ and $\hat{\beta}_W$, which are used to derive the expected values of estimators for both the cluster and crossover designs.

The number of episodes for each treatment sequence in a re-randomisation trial that occur asymptotically under a 1:1 allocation ratio were previously shown in table 3.1 in chapter 3. The number of episodes that occur in each treatment sequence in a cluster or crossover trial is shown in tables 5.1 and 5.2.

As a brief reminder, N_T is the total number of patients enrolled in the trial, M_T the total number of episodes, M_i is the number of episodes for which patient i is enrolled in the trial, and $M_{T(j)}$ denotes the number of patients for whom $M_i = j$. Then, $p = \frac{M_{T(2)}}{N_T}$ is the proportion of patients who are re-enrolled for a second episode.

The crossover design omits treatment sequences $Z = (0,0)$ and $Z = (1,1)$, and so has four possible sequences: Z is one of (0), (1), (0,1), or (1,0). There are $N_T(1-p)$ patients enrolled for a single episode; therefore, there are $\frac{N_T}{2}(1-p)$ patients for treatment sequences $Z = (0)$ and $Z = (1)$ respectively. There are $N_T p$ patients enrolled for two episodes; therefore, there are $\frac{N_T p}{2}$ patients for treatment sequences $Z = (0,1)$ and $Z = (1,0)$.

Similarly, the cluster design omits sequences $Z = (0,1)$ and $Z = (1,0)$, and so has $\frac{N_T}{2}(1-p)$ patients for treatment sequences $Z = (0)$ and $Z = (1)$ respectively, and $\frac{N_T p}{2}$ patients for treatment sequences $Z = (0,0)$ and $Z = (1,1)$.

5.2.2 Estimation components for cluster and crossover designs

Now I derive the components $\hat{\beta}_{B_1}$, $\hat{\beta}_{B_2}$ and $\hat{\beta}_W$ for the cluster and crossover designs, as these will be used to derive the expected value of the estimators. Some further background into these components is available in chapter 3.

The cluster design uses two between-patient estimation components. The first, denoted by $\hat{\beta}_{B_1}$, is based on treatment sequences $Z = (0)$ and $Z = (1)$; the second,

Table 5.2: Asymptotic number of patients in each treatment sequence in a crossover trial with 1:1 allocation ratio

	Treatment allocation		**Number of patients at each episode**	
Sequence	**Episode 1**	**Episode 2**	**Episode 1**	**Episode 2**
$Z = (1)$	1	-	$\frac{N_T}{2}(1-p)$	-
$Z = (1,0)$	1	0	$\frac{N_T p}{2}$	$\frac{N_T p}{2}$
$Z = (0)$	0	-	$\frac{N_T}{2}(1-p)$	-
$Z = (0,1)$	0	1	$\frac{N_T p}{2}$	$\frac{N_T p}{2}$

denoted by $\hat{\beta}_{B_2}$, is based on treatment sequences $Z = (0,\ 0)$ and $Z = (1,\ 1)$. They are shown below:

$$\hat{\beta}_{B_1} = \frac{1}{\frac{N_T}{2}(1-p)} \left(\sum_{i \in Z=(1)} Y_{i1} - \sum_{i \in Z=(0)} Y_{i1} \right)$$

$$\begin{aligned} \hat{\beta}_{B_2} &= \frac{\sum_{i \in Z=(1,1)} (Y_{i1} + Y_{i2})}{\frac{N_T p}{2} + \frac{N_T p}{2}} - \frac{\sum_{i \in Z=(0,0)} (Y_{i1} + Y_{i2})}{\frac{N_T p}{2} + \frac{N_T p}{2}} \\ &= \frac{1}{N_T p} \left(\sum_{i \in Z=(1,1)} (Y_{i1} + Y_{i2}) - \sum_{i \in Z=(0,0)} (Y_{i1} + Y_{i2}) \right) \end{aligned}$$

i.e. $\hat{\beta}_{B_1}$ is the mean of all episodes in treatment sequence $Z = (1)$ vs. the mean of all episodes in treatment sequence $Z = (0)$, and $\hat{\beta}_{B_2}$ is the mean of all episodes in treatment sequence $Z = (1,\ 1)$ vs. the mean of all episodes in treatment sequence $Z = (0,\ 0)$ (where the denominators are shown in table 5.1, and were derived in the previous section).

The crossover design uses one between-patient estimation component, and one within-patient estimation component. The between-patient component, $\hat{\beta}_{B_1}$, is the same as that used for the cluster design above; the within-patient estimation component, denoted by $\hat{\beta}_W$, is based on treatment sequences $Z = (0,\ 1)$ and $Z = (1,\ 0)$. It is shown below:

$$\begin{aligned} \hat{\beta}_W &= \frac{\sum_{i \in Z=(1,0)} (Y_{i1} - Y_{i2})}{\frac{N_T p}{2} + \frac{N_T p}{2}} - \frac{\sum_{i \in Z=(0,1)} (Y_{i1} - Y_{i2})}{\frac{N_T p}{2} + \frac{N_T p}{2}} \\ &= \frac{1}{N_T p} \left(\sum_{i \in Z=(1,0)} (Y_{i1} - Y_{i2}) - \sum_{i \in Z=(0,1)} (Y_{i1} - Y_{i2}) \right) \end{aligned}$$

i.e. $\hat{\beta}_W$ is the mean of intervention episodes for patients on treatment sequences $Z = (0,\ 1)$ and $Z = (1,\ 0)$ vs. the mean of control episodes for patients on treatment sequences $Z = (0,\ 1)$ and $Z = (1,\ 0)$.

Note that the component $\hat{\beta}_{B_1}$ is equivalent to that used in the derivations for the re-randomisation design; however, calculation of the components $\hat{\beta}_{B_2}$ and $\hat{\beta}_W$ differs to how the components for the re-randomisation design are calculated, as they include the term $\frac{N_T p}{2}$ in the denominator rather than the term $\frac{N_T p}{4}$ as for the re-randomisation design. This is because the re-randomisation design has two additional treatment sequences, and so fewer patients with two episodes are allocated to each of the possible sequences. However, this difference does not change the value of $\hat{\beta}_{B_1}$, $\hat{\beta}_{B_2}$ or $\hat{\beta}_W$ for either the cluster or crossover design compared to the re-randomisation design; this is because although the denominator is different, so is the numerator (as the summation goes to $\frac{N_T p}{2}$ rather than to $\frac{N_T p}{4}$ as in the re-randomisation design), and so the difference cancels out. This means the expected values of the components $\hat{\beta}_{B_1}$, $\hat{\beta}_{B_2}$ and $\hat{\beta}_W$ are identical for each of the three designs (cluster, crossover, re-randomisation) for each scenario. Values of these components are shown in table 5.3 for reference.

5.2.3 Independence estimators for cluster designs

The independence estimator for the cluster design, $\hat{\beta}_{CL,ind}$, can be calculated as a difference in means for all intervention episodes vs. control episodes, and can be re-arranged as:

$$\hat{\beta}_{CL,ind} = \frac{\sum_{ij} Y_{ij} Z_{ij}}{\sum_{ij} Z_{ij}} - \frac{\sum_{ij} Y_{ij} (1 - Z_{ij})}{\sum_{ij} (1 - Z_{ij})}$$

$$= \frac{\frac{N_T}{2} (1 - p) \hat{\beta}_{B_1} + N_T p \hat{\beta}_{B_2}}{\frac{N_T}{2} (1 - p) + N_T p} = \frac{(1 - p) \hat{\beta}_{B_1} + 2p \hat{\beta}_{B_2}}{1 + p}$$

Taking the expectation leads to:

$$E\left(\hat{\beta}_{CL,ind}\right) = \frac{(1 - p) E\left(\hat{\beta}_{B_1}\right) + 2pE\left(\hat{\beta}_{B_2}\right)}{1 + p}$$

5.2.4 Mixed-effects models for cluster designs

The estimated treatment effect for a mixed-model ($\hat{\beta}_{CL,MM}$) for the cluster design is the same as for the re-randomisation design, but with the term $\hat{\beta}_W$ omitted (i.e. it is a weighted average of each of its components, with weights equal to the inverse variance from each component [51]):

$$\hat{\beta}_{CL,MM} = \frac{\frac{\hat{\beta}_{B_1}}{V\left(\hat{\beta}_{B_1}\right)} + \frac{\hat{\beta}_{B_2}}{V\left(\hat{\beta}_{B_2}\right)}}{\frac{1}{V\left(\hat{\beta}_{B_1}\right)} + \frac{1}{V\left(\hat{\beta}_{B_2}\right)}}$$

where $V\left(\hat{\beta}_{B_1}\right)$ and $V\left(\hat{\beta}_{B_2}\right)$ represent the estimated variances; in this chapter (as in chapter 4), I use $V()$ to denote $\hat{V}()$ for simplicity.

Table 5.3: Expected values of between- and within-patient estimation components for cluster, crossover, and re-randomisation designs under different treatment effect mechanisms.

Scenario	$E(\hat{\beta}_{B_1})$	$E(\hat{\beta}_{B_2})$	$E(\hat{\beta}_W)$
S1 – Constant treatment effect	β	β	β
S2 – Treatment effect varies across episode	β_1	$\frac{1}{2}(\beta_1 + \beta_2)$	$\frac{1}{2}(\beta_1 + \beta_2)$
S3 – Treatment effect varies across value of M_i	β_1	β_2	β_2
S4 – Treatment effect carries forward into second episode	β	$\beta + \frac{\gamma}{2}$	$\beta - \frac{\gamma}{2}$
S5 – Treatment becomes less effective on re-use	β	$\beta + \frac{\delta}{2}$	β
S6 – Constant treatment effect, differential non-enrolment based on outcome in previous episode	$\beta_{trt} + \beta_{X_{PL}} \frac{(p_{01} - p_{00})}{2(1-p)}$	$\beta_{trt} + \beta_{X_{PL}} \frac{(p_{00} - p_{01})}{2p}$	β_{trt}
S7 – Constant treatment effect, differential non-enrolment based on expected outcome in current episode	β_{trt}	$\beta_{trt} + \frac{\beta_{X_{EL}}}{4}\left(\frac{p_{00} - p_{01}}{p}\right)$	$\beta_{trt} - \frac{\beta_{X_{EL}}}{4}\left(\frac{p_{00} - p_{01}}{p}\right)$

To work out the expected value of $\hat{\beta}_{CL,MM}$ we need to derive the estimated variances of its components. The estimated variance of $\hat{\beta}_{B_1}$ is the same as under the re-randomisation design:

$$V\left(\hat{\beta}_{B_1}\right) = \frac{4\hat{\sigma}^2}{N_T\left(1-p\right)}$$

The estimated variance of $\hat{\beta}_{B_2}$ under the cluster design is:

$$V\left(\hat{\beta}_{B_2}\right) = V\left(\frac{1}{N_T p}\left(\sum_{i\in Z=(1,1)}\left(Y_{i1}+Y_{i2}\right) - \sum_{i\in Z=(0,0)}\left(Y_{i1}+Y_{i2}\right)\right)\right)$$

$$= \frac{1}{\left(N_T p\right)^2}\left(V\left(\sum_{i\in Z=(1,1)}\left(Y_{i1}+Y_{i2}\right)\right) + V\left(\sum_{i\in Z=(0,0)}\left(Y_{i1}+Y_{i2}\right)\right)\right)$$

$$= \frac{1}{\left(N_T p\right)^2}\left(2\sum_{i\in Z=(1,1)}\left(V\left(Y_{i1}\right)+V\left(Y_{i2}\right)+2COV\left(Y_{i1},Y_{i2}\right)\right)\right)$$

$$= \frac{1}{\left(N_T p\right)^2}\left((2)\left(\frac{N_T p}{2}\right)\left(2\hat{\sigma}^2+2\hat{\vartheta}\hat{\sigma}^2\right)\right)$$

$$= \frac{2\hat{\sigma}^2\left(1+\hat{\vartheta}\right)}{N_T p}$$

Then, the denominator of $\hat{\beta}_{CL,MM}$ is:

$$\frac{1}{V\left(\hat{\beta}_{B_1}\right)} + \frac{1}{V\left(\hat{\beta}_{B_2}\right)} = \frac{N_T\left(1-p\right)}{4\hat{\sigma}^2} + \frac{N_T p}{2\hat{\sigma}^2\left(1+\hat{\vartheta}\right)}$$

$$= \frac{N_T\left(1-p\right)\left(1+\hat{\vartheta}\right)+2N_T p}{4\hat{\sigma}^2\left(1+\hat{\vartheta}\right)} = \frac{N_T\left(1+p+\hat{\vartheta}-\hat{\vartheta}p\right)}{4\hat{\sigma}^2\left(1+\hat{\vartheta}\right)}$$

And the numerator of $\hat{\beta}_{CL,MM}$ is:

$$\frac{\hat{\beta}_{B_1}}{V\left(\hat{\beta}_{B_1}\right)} + \frac{E\hat{\beta}_{B_2}}{V\left(\hat{\beta}_{B_2}\right)} = \frac{N_T\left(1-p\right)\left(1+\hat{\vartheta}\right)\hat{\beta}_{B_1}+2N_T p\hat{\beta}_{B_2}}{4\hat{\sigma}^2\left(1+\hat{\vartheta}\right)}$$

Putting these together, we get:

$$\hat{\beta}_{CL,MM} = \frac{\left(1 - p + \hat{\vartheta} - \hat{\vartheta}p\right)\hat{\beta}_{B_1} + 2p\hat{\beta}_{B_2}}{\left(1 + p + \hat{\vartheta} - \hat{\vartheta}p\right)}$$

And taking the expectation (using a first order Taylor series expansion) we get:

$$E\left(\hat{\beta}_{CL,MM}\right) = \frac{(1 - p + \vartheta^* - \vartheta^* p)\, E\left(\hat{\beta}_{B_1}\right) + 2pE\left(\hat{\beta}_{B_2}\right)}{(1 + p + \vartheta^* - \vartheta^* p)}$$

5.2.5 Independence estimators for crossover designs

The independence estimator for the crossover design, $\hat{\beta}_{CO,ind}$, can be calculated as a difference in means for all intervention episodes vs. control episodes, and can be re-arranged as:

$$\hat{\beta}_{CO,ind} = \frac{\sum_{ij} Y_{ij} Z_{ij}}{\sum_{ij} Z_{ij}} - \frac{\sum_{ij} Y_{ij}\,(1 - Z_{ij})}{\sum_{ij}(1 - Z_{ij})} = \frac{\frac{N_T}{2}(1 - p)\,\hat{\beta}_{B_1} + N_T p\hat{\beta}_W}{\frac{N_T}{2}(1 - p) + N_T p}$$

$$= \frac{(1 - p)\,\hat{\beta}_{B_1} + 2p\hat{\beta}_W}{1 + p}$$

Taking the expectation leads to:

$$E\left(\hat{\beta}_{CO,ind}\right) = \frac{(1 - p)\, E\left(\hat{\beta}_{B_1}\right) + 2pE\left(\hat{\beta}_W\right)}{1 + p}$$

5.2.6 Mixed-effects models for crossover designs

The estimated treatment effect for a mixed-model $\hat{\beta}_{CO,MM}$ for the crossover design is the same as for the re-randomisation design, but with the term $\hat{\beta}_{B_1}$ omitted (i.e. it is a weighted average of each of its components, with weights equal to the inverse variance from each component [51]):

$$\hat{\beta}_{CO,MM} = \frac{\frac{\hat{\beta}_{B_1}}{V\left(\hat{\beta}_{B_1}\right)} + \frac{\hat{\beta}_W}{V\left(\hat{\beta}_W\right)}}{\frac{1}{V\left(\hat{\beta}_{B_1}\right)} + \frac{1}{V\left(\hat{\beta}_W\right)}}$$

To work out the expected value of $\hat{\beta}_{CO,MM}$ we need to derive the estimated variances of its components. The estimated variance of $\hat{\beta}_{B_1}$ is the same as under the re-randomisation design:

$$V\left(\hat{\beta}_{B_1}\right) = \frac{4\hat{\sigma}^2}{N_T\,(1 - p)}$$

The estimated variance of $\hat{\beta}_W$ under the crossover design is:

$$V\left(\hat{\beta}_W\right) = V\left(\frac{1}{N_T p}\left(\sum_{i\in Z=(1,0)}\left(Y_{i1}-Y_{i2}\right) - \sum_{i\in Z=(0,1)}\left(Y_{i1}-Y_{i2}\right)\right)\right)$$

$$= \frac{1}{\left(N_T p\right)^2}\left(V\left(\sum_{i\in Z=(1,0)}\left(Y_{i1}-Y_{i2}\right)\right) + V\left(\sum_{i\in Z=(0,1)}\left(Y_{i1}-Y_{i2}\right)\right)\right)$$

$$= \frac{1}{\left(N_T p\right)^2}\left(2\sum_{i\in Z=(1,0)}\left(V\left(Y_{i1}\right)+V\left(Y_{i2}\right)-2COV\left(Y_{i1},Y_{i2}\right)\right)\right)$$

$$= \frac{1}{\left(N_T p\right)^2}\left((2)\left(\frac{N_T p}{2}\right)\left(2\hat{\sigma}^2 - 2\hat{\vartheta}\hat{\sigma}^2\right)\right)$$

$$= \frac{2\hat{\sigma}^2\left(1-\hat{\vartheta}\right)}{N_T p}$$

Then, the denominator for $\hat{\beta}_{CO,MM}$ is:

$$\frac{1}{V\left(\hat{\beta}_{B_1}\right)} + \frac{1}{V\left(\hat{\beta}_W\right)} = \frac{N_T\left(1-p\right)}{4\hat{\sigma}^2} + \frac{N_T p}{2\hat{\sigma}^2\left(1-\hat{\vartheta}\right)}$$

$$= \frac{N_T\left(1-p\right)\left(1-\hat{\vartheta}\right) + 2N_T p}{4\hat{\sigma}^2\left(1-\hat{\vartheta}\right)} = \frac{N_T\left(1+p-\hat{\vartheta}+\hat{\vartheta}p\right)}{4\hat{\sigma}^2\left(1-\hat{\vartheta}\right)}$$

And the numerator for $\hat{\beta}_{CO,MM}$ is:

$$\frac{\hat{\beta}_{B_1}}{V\left(\hat{\beta}_{B_1}\right)} + \frac{\hat{\beta}_W}{V\left(\hat{\beta}_W\right)} = \frac{N_T\left(1-p\right)\left(1-\hat{\vartheta}\right)\hat{\beta}_{B_1} + 2N_T p\hat{\beta}_W}{4\hat{\sigma}^2\left(1-\hat{\vartheta}\right)}$$

Putting these together, we get:

$$\hat{\beta}_{CO,MM} = \frac{\left(1-p-\hat{\vartheta}+\hat{\vartheta}p\right)\hat{\beta}_{B_1} + 2p\hat{\beta}_W}{\left(1+p-\hat{\vartheta}+\hat{\vartheta}p\right)}$$

And, taking the expectation using a first order Taylor series expansion, we get:

$$E\left(\hat{\beta}_{CO,MM}\right) = \frac{\left(1-p-\vartheta^*+\vartheta^* p\right)E\left(\hat{\beta}_{B_1}\right) + 2pE\left(\hat{\beta}_W\right)}{\left(1+p-\vartheta^*+\vartheta^* p\right)}$$

5.2.7 S1: constant treatment effect

In this scenario (and all other scenarios), the expected value of the different estimators is calculated by substituting in the expected values for the components $\hat{\beta}_{B_1}$, $\hat{\beta}_{B_2}$ and $\hat{\beta}_W$ from table 5.3 into the formulas for the expected value of each estimator derived in the previous sections.

5.2.7.1 Cluster design, independence estimator

The expected value is:

$$E\left(\hat{\beta}_{CL,ind}\right) = \frac{(1-p)\,\beta + 2p\beta}{1+p} = \beta$$

5.2.7.2 Cluster design, mixed-effects model

The expected value is:

$$E\left(\hat{\beta}_{CL,MM}\right) = \frac{(1-p+\vartheta^*-\vartheta^* p)\,\beta + 2p\beta}{(1+p+\vartheta^*-\vartheta^* p)} = \beta$$

5.2.7.3 Crossover design, independence estimator

The expected value is:

$$E\left(\hat{\beta}_{CO,ind}\right) = \frac{(1-p)\,\beta + 2p\beta}{1+p} = \beta$$

5.2.7.4 Crossover design, mixed-effects model

The expected value is:

$$E\left(\hat{\beta}_{CO,MM}\right) = \frac{(1-p-\vartheta^*+\vartheta^* p)\,\beta + 2p\beta}{(1+p-\vartheta^*+\vartheta^* p)} = \beta$$

5.2.8 S2: treatment effect varies across episode

5.2.8.1 Cluster design, independence estimator

The expected value is:

$$E\left(\hat{\beta}_{CL,ind}\right) = \frac{(1-p)\,\beta_1 + 2p\left(\frac{\beta_1+\beta_2}{2}\right)}{1+p} = \frac{\beta_1 + p\beta_2}{1+p}$$

5.2.8.2 Cluster design, mixed-effects model

The expected value is:

$$E\left(\hat{\beta}_{CL,MM}\right) = \frac{\left(1-p+\vartheta^*-\vartheta^* p\right)\beta_1 + 2p\left(\frac{\beta_1+\beta_2}{2}\right)}{\left(1+p+\vartheta^*-\vartheta^* p\right)}$$

$$= \frac{\left(1+\vartheta^*-\vartheta^* p\right)\beta_1 + p\beta_2}{\left(1+p+\vartheta^*-\vartheta^* p\right)}$$

5.2.8.3 Crossover design, independence estimator

The expected value is:

$$E\left(\hat{\beta}_{CO,ind}\right) = \frac{\left(1-p\right)\beta_1 + 2p\left(\frac{\beta_1+\beta_2}{2}\right)}{1+p} = \frac{\beta_1+p\beta_2}{1+p}$$

5.2.8.4 Crossover design, mixed-effects model

The expected value is:

$$E\left(\hat{\beta}_{CO,MM}\right) = \frac{\left(1-p-\vartheta^*+\vartheta^* p\right)\beta_1 + 2p\left(\frac{\beta_1+\beta_2}{2}\right)}{\left(1+p-\vartheta^*+\vartheta^* p\right)}$$

$$= \frac{\left(1-\vartheta^*+\vartheta^* p\right)\beta_1 + p\beta_2}{\left(1+p-\vartheta^*+\vartheta^* p\right)}$$

5.2.9 S3: treatment effect varies across patients with different values of M_i

5.2.9.1 Cluster design, independence estimator

The expected value is:

$$E\left(\hat{\beta}_{CL,ind}\right) = \frac{\left(1-p\right)\beta_1 + 2p\beta_2}{1+p}$$

5.2.9.2 Cluster design, mixed-effects model

The expected value is:

$$E\left(\hat{\beta}_{CL,MM}\right) = \frac{\left(1-p+\vartheta^*-\vartheta^* p\right)\beta_1 + 2p\beta_2}{\left(1+p+\vartheta^*-\vartheta^* p\right)}$$

5.2.9.3 Crossover design, independence estimator

The expected value is:

$$E\left(\hat{\beta}_{CO,ind}\right) = \frac{\left(1-p\right)\beta_1 + 2p\beta_2}{1+p}$$

5.2.9.4 Crossover design, mixed-effects model

The expected value is:

$$E\left(\hat{\beta}_{CO,MM}\right) = \frac{\left(1 - p - \vartheta^* + \vartheta^* p\right)\beta_1 + 2p\beta_2}{\left(1 + p - \vartheta^* + \vartheta^* p\right)}$$

5.2.10 S4: treatment effect carries forward into the second episode

5.2.10.1 Cluster design, independence estimator

The expected value is:

$$E\left(\hat{\beta}_{CL,ind}\right) = \frac{\left(1 - p\right)\beta + 2p\left(\beta + \frac{\gamma}{2}\right)}{1 + p}$$

$$= \frac{\left(1 + p\right)\beta + p\gamma}{1 + p} = \beta + \frac{p\gamma}{1 + p}$$

5.2.10.2 Cluster design, mixed-effects model

The expected value is:

$$E\left(\hat{\beta}_{CL,MM}\right) = \frac{\left(1 - p + \vartheta^* - \vartheta^* p\right)\beta + 2p\left(\beta + \frac{\gamma}{2}\right)}{\left(1 + p + \vartheta^* - \vartheta^* p\right)}$$

$$= \frac{\left(1 + p + \vartheta^* - \vartheta^* p\right)\beta + p\gamma}{\left(1 + p + \vartheta^* - \vartheta^* p\right)} = \beta + \frac{p\gamma}{\left(1 + p + \vartheta^* - \vartheta^* p\right)}$$

5.2.10.3 Crossover design, independence estimator

The expected value is:

$$E\left(\hat{\beta}_{CO,ind}\right) = \frac{\left(1 - p\right)\beta + 2p\left(\beta - \frac{\gamma}{2}\right)}{1 + p} = \beta - \frac{p\gamma}{1 + p}$$

5.2.10.4 Crossover design, mixed-effects model

The expected value is:

$$E\left(\hat{\beta}_{CO,MM}\right) = \frac{\left(1 - p - \vartheta^* + \vartheta^* p\right)\beta + 2p\left(\beta - \frac{\gamma}{2}\right)}{\left(1 + p - \vartheta^* + \vartheta^* p\right)}$$

$$= \frac{\left(1 + p - \vartheta^* + \vartheta^* p\right)\beta - p\gamma}{\left(1 + p - \vartheta^* + \vartheta^* p\right)} = \beta - \frac{p\gamma}{\left(1 + p - \vartheta^* + \vartheta^* p\right)}$$

5.2.11 S5: treatment becomes less effective on re-use

5.2.11.1 Cluster design, independence estimator

The expected value is:

$$E\left(\hat{\beta}_{CL,ind}\right) = \frac{(1-p)\,\beta + 2p\left(\beta + \frac{\delta}{2}\right)}{1+p}$$

$$= \frac{(1+p)\,\beta + p\delta}{1+p} = \beta + \frac{p\delta}{1+p}$$

5.2.11.2 Cluster design, mixed-effects model

The expected value is:

$$E\left(\hat{\beta}_{CL,MM}\right) = \frac{(1-p+\vartheta^*-\vartheta^* p)\,\beta + 2p\left(\beta + \frac{\delta}{2}\right)}{(1+p+\vartheta^*-\vartheta^* p)}$$

$$= \frac{(1+p+\vartheta^*-\vartheta^* p)\,\beta + p\delta}{(1+p+\vartheta^*-\vartheta^* p)} = \beta + \frac{p\delta}{(1+p+\vartheta^*-\vartheta^* p)}$$

5.2.11.3 Crossover design, independence estimator

The expected value is:

$$E\left(\hat{\beta}_{CO,ind}\right) = \frac{(1-p)\,\beta + 2p\beta}{1+p} = \beta$$

5.2.11.4 Crossover design, mixed-effects model

The expected value is:

$$E\left(\hat{\beta}_{CO,MM}\right) = \frac{(1-p-\vartheta^*+\vartheta^* p)\,\beta + 2p\beta}{(1+p-\vartheta^*+\vartheta^* p)} = \beta$$

5.2.12 S6: constant treatment effect, non-enrolment scenario 4

5.2.12.1 Cluster design, independence estimator

The expected value is:

$$E\left(\hat{\beta}_{CL,ind}\right) = \frac{(1-p)\left(\beta_{trt} + \beta_{X_{PL}}\frac{(p_{01}-p_{00})}{2(1-p)}\right) + 2p\left(\beta_{trt} + \beta_{X_{PL}}\frac{(p_{00}-p_{01})}{2p}\right)}{1+p}$$

$$= \frac{(1-p+2p)\,\beta_{trt} + \beta_{X_{PL}}\left(\frac{(1-p)(p_{01}-p_{00})}{2(1-p)} + 2p\frac{(p_{00}-p_{01})}{2p}\right)}{1+p}$$

$$= \beta_{trt} + \beta_{X_{PL}} \frac{((p_{01} - p_{00}) + 2(p_{00} - p_{01}))}{2(1+p)}$$

$$= \beta_{trt} + \beta_{X_{PL}} \frac{p_{00} - p_{01}}{2(1+p)}$$

5.2.12.2 Cluster design, mixed-effects model

The expected value is:

$$E\left(\hat{\beta}_{CL,MM}\right) = \frac{(1 - p + \vartheta^* - \vartheta^* p)\left(\beta_{trt} + \beta_{X_{PL}} \frac{(p_{01} - p_{00})}{2(1-p)}\right) + 2p\left(\beta_{trt} + \beta_{X_{PL}} \frac{(p_{00} - p_{01})}{2p}\right)}{(1 + p + \vartheta^* - \vartheta^* p)}$$

$$= \frac{(1 - p + \vartheta^* - \vartheta^* p + 2p)\beta_{trt} + \beta_{X_{PL}}\left((1 - p + \vartheta^* - \vartheta^* p)\frac{(p_{01} - p_{00})}{2(1-p)} + (p_{00} - p_{01})\right)}{(1 + p + \vartheta^* - \vartheta^* p)}$$

$$= \frac{(1 + p + \vartheta^* - \vartheta^* p)\beta_{trt} + \beta_{X_{PL}}\left((1 - p + \vartheta^* - \vartheta^* p)\frac{(p_{01} - p_{00}) + 2(1-p)(p_{00} - p_{01})}{2(1-p)}\right)}{(1 + p + \vartheta^* - \vartheta^* p)}$$

$$= \beta_{trt} + \beta_{X_{PL}} \frac{((1 - p + \vartheta^* - \vartheta^* p)(p_{01} - p_{00}) - 2(1-p)(p_{01} - p_{00}))}{2(1-p)(1 + p + \vartheta^* - \vartheta^* p)}$$

$$= \beta_{trt} + \beta_{X_{PL}} \frac{(p_{01} - p_{00})(1 - p + \vartheta^* - \vartheta^* p - 2 + 2p)}{2(1-p)(1 + p + \vartheta^* - \vartheta^* p)}$$

$$= \beta_{trt} + \beta_{X_{PL}} \frac{(p_{01} - p_{00})(-1 + p + \vartheta^* - \vartheta^* p)}{2(1-p)(1 + p + \vartheta^* - \vartheta^* p)}$$

$$= \beta_{trt} + \beta_{X_{PL}} \frac{(p_{00} - p_{01})(1 - p - \vartheta^* + \vartheta^* p)}{2(1-p)(1 + p + \vartheta^* - \vartheta^* p)}$$

5.2.12.3 Crossover design, independence estimator

The expected value is:

$$E\left(\hat{\beta}_{CO,ind}\right) = \frac{(1-p)\left(\beta_{trt} + \beta_{X_{PL}} \frac{(p_{01} - p_{00})}{2(1-p)}\right) + 2p\beta_{trt}}{1+p}$$

$$= \frac{(1-p)\beta_{trt} + 2p\beta_{trt} + \beta_{X_{PL}} \frac{(p_{01} - p_{00})}{2}}{1+p} = \beta_{trt} + \frac{\beta_{X_{PL}}(p_{01} - p_{00})}{2(1+p)}$$

5.2.12.4 Crossover design, mixed-effects model

The expected value is:

$$E\left(\hat{\beta}_{CO,MM}\right) = \frac{(1-p-\vartheta^*+\vartheta^* p)\left(\beta_{trt}+\beta_{X_{PL}}\frac{(p_{01}-p_{00})}{2(1-p)}\right)+2p\beta_{trt}}{(1+p-\vartheta^*+\vartheta^* p)}$$

$$= \frac{(1-p-\vartheta^*+\vartheta^* p)\,\beta_{trt}+2p\beta_{trt}+\beta_{X_{PL}}\frac{(1-p-\vartheta^*+\vartheta^* p)(p_{01}-p_{00})}{2(1-p)}}{(1+p-\vartheta^*+\vartheta^* p)}$$

$$= \frac{(1+p-\vartheta^*+\vartheta^* p)\,\beta_{trt}+\beta_{X_{PL}}\frac{(1-p-\vartheta^*+\vartheta^* p)(p_{01}-p_{00})}{2(1-p)}}{(1+p-\vartheta^*+\vartheta^* p)}$$

$$= \beta_{trt}+\frac{\beta_{X_{PL}}(1-p-\vartheta^*+\vartheta^* p)(p_{01}-p_{00})}{2(1-p)(1+p-\vartheta^*+\vartheta^* p)}$$

5.2.13 S7: constant treatment effect, non-enrolment scenario 5

5.2.13.1 Cluster design, independence estimator

The expected value is:

$$E\left(\hat{\beta}_{CL,ind}\right) = \frac{(1-p)\,\beta_{trt}+2p\left(\beta_{trt}+\frac{\beta_{X_{EL}}}{4}\left(\frac{p_{00}-p_{01}}{p}\right)\right)}{1+p}$$

$$= \frac{(1+p)\,\beta_{trt}+\beta_{X_{EL}}\left(\frac{p_{00}-p_{01}}{2}\right)}{1+p} = \beta_{trt}+\beta_{X_{EL}}\frac{p_{00}-p_{01}}{2(1+p)}$$

5.2.13.2 Cluster design, mixed-effects model

The expected value is:

$$E\left(\hat{\beta}_{CL,MM}\right) = \frac{(1-p+\vartheta^*-\vartheta^* p)\,\beta_{trt}+2p\left(\beta_{trt}+\frac{\beta_{X_{EL}}}{4}\left(\frac{p_{00}-p_{01}}{p}\right)\right)}{(1+p+\vartheta^*-\vartheta^* p)}$$

$$= \frac{(1-p+\vartheta^*-\vartheta^* p+2p)\,\beta_{trt}+\beta_{X_{EL}}\left(\frac{p_{00}-p_{01}}{2}\right)}{(1+p+\vartheta^*-\vartheta^* p)}$$

$$= \beta_{trt}+\beta_{X_{EL}}\frac{p_{00}-p_{01}}{2(1+p+\vartheta^*-\vartheta^* p)}$$

5.2.13.3 Crossover design, independence estimator

The expected value is:

$$E\left(\hat{\beta}_{CO,ind}\right) = \frac{(1-p)\,\beta_{trt} + 2p\left(\beta_{trt} - \frac{\beta_{X_{EL}}}{4}\left(\frac{p_{00}-p_{01}}{p}\right)\right)}{1+p}$$

$$= \frac{(1+p)\,\beta_{trt} - \beta_{X_{EL}}\left(\frac{p_{00}-p_{01}}{2}\right)}{1+p} = \beta_{trt} - \frac{\beta_{X_{EL}}\left(p_{00}-p_{01}\right)}{2\left(1+p\right)}$$

5.2.13.4 Crossover design, mixed-effects model

The expected value is:

$$E\left(\hat{\beta}_{CO,MM}\right) = \frac{(1-p-\vartheta^*+\vartheta^* p)\,\beta_{trt} + 2p\left(\beta_{trt} - \frac{\beta_{X_{EL}}}{4}\left(\frac{p_{00}-p_{01}}{p}\right)\right)}{(1+p-\vartheta^*+\vartheta^* p)}$$

$$= \frac{(1+p-\vartheta^*+\vartheta^* p)\,\beta_{trt} - \beta_{X_{EL}}\left(\frac{p_{00}-p_{01}}{2}\right)}{(1+p-\vartheta^*+\vartheta^* p)}$$

$$= \beta_{trt} - \frac{\beta_{X_{EL}}\left(p_{00}-p_{01}\right)}{2\left(1+p-\vartheta^*+\vartheta^* p\right)}$$

5.2.14 Bias in cluster and crossover designs

5.2.14.1 Cluster designs

The independence and mixed-effects model estimators are shown against the true value of the policy-benefit estimand for scenarios 1-7 in table 5.4. Mixed-effects models for cluster designs are biased in all scenarios considered except when the treatment effect is constant. Independence estimators are unbiased in all scenarios when there is no non-enrolment; they are biased for scenarios 6 and 7, where non-enrolment is differential based on the outcome in the previous episode or on the baseline prognosis in the current episode respectively.

5.2.14.2 Crossover designs

The independence and mixed-effects model estimators are shown against the true value of the added-benefit estimand for scenarios 1-7 in table 5.5. Mixed-effects models for crossover designs are biased in all scenarios considered except when the treatment effect is constant. Independence estimators are biased when treatment history affects the current episode (scenario 4, treatment effect carries forward; and scenario 5, treatment becomes less effective on re-use), or under scenarios 6 and 7, where there is differential non-enrolment. They are unbiased for scenarios 1-3 (when treatment effect is constant, or varies by episode or by value of M_i).

Table 5.4: Bias of the cluster design for the per-episode policy-benefit estimand. Derivation of these estimands is shown in chapter 2.

Scenario	Estimand	Independence estimator	Mixed-effects model
S1 – Constant treatment effect	β	β	β
S2 – Treatment effect varies across episode	$\frac{\beta_1+p\beta_2}{1+p}$	$\frac{\beta_1+p\beta_2}{1+p}$	$\frac{(1+\vartheta^*-\vartheta^*p)\beta_1+p\beta_2}{(1+p+\vartheta^*-\vartheta^*p)}$
S3 – Treatment effect varies across value of M_i	$\frac{(1-p)\beta_1+2p\beta_2}{(1+p)}$	$\frac{(1-p)\beta_1+2p\beta_2}{1+p}$	$\frac{(1-p+\vartheta^*-\vartheta^*p)\beta_1+2p\beta_2}{(1+p+\vartheta^*-\vartheta^*p)}$
S4 – Treatment effect carries forward into the second episode	$\beta+\frac{p\gamma}{(1+p)}$	$\beta+\frac{p\gamma}{1+p}$	$\beta+\frac{p\gamma}{(1+p+\vartheta^*-\vartheta^*p)}$
S5 – Treatment becomes less effective on re-use	$\beta+\frac{p\delta}{(1+p)}$	$\beta+\frac{p\delta}{1+p}$	$\beta+\frac{p\delta}{(1+p+\vartheta^*-\vartheta^*p)}$
S6 – Constant treatment effect, differential non-enrolment based on outcome in previous episode	β_{trt}	$\beta_{trt}+\beta_{X_{PL}}\frac{p_{00}-p_{01}}{2(1+p)}$	$\beta_{trt}+\beta_{X_{PL}}\frac{(p_{00}-p_{01})(1-p-\vartheta^*+\vartheta^*p)}{2(1-p)(1+p+\vartheta^*-\vartheta^*p)}$
S7 – Constant treatment effect, differential non-enrolment based on expected outcome in current episode	β_{trt}	$\beta_{trt}+\beta_{X_{EL}}\frac{p_{00}-p_{01}}{2(1+p)}$	$\beta_{trt}+\beta_{X_{EL}}\frac{p_{00}-p_{01}}{2(1+p+\vartheta^*-\vartheta^*p)}$

5.3 Simulation study (design)

In this section I describe the methods used to implement a simulation study to compare the performance of the four different designs (re-randomisation, parallel group, crossover, and cluster). For the parallel group design, I evaluate two different scenarios, which are based on enrolling a different number of patients. The number of individual patients in the first scenario, labelled PG(big), is based on matching the total number of episodes enrolled by the re-randomisation, cluster, and crossover designs. The number of individual patients in the second scenario, labelled PG(small), is based on matching the total number of individual patients enrolled by the re-randomisation, cluster, and crossover designs. For example, in a setting where 300 individual patients experience 450 episodes, PG(big) would enrol 450 individual patients whereas PG(small) would enrol 300 individual patients.

As in chapters 3 and 4, the main purpose of this simulation study is to evaluate the designs under smaller sample sizes and with more complicated data generating

Table 5.5: Bias of the crossover design for the per-episode policy-benefit estimand. Derivation of these estimands is shown in chapter 2.

Scenario	Estimand	Independence estimator	Mixed-effects model
S1 – Constant treatment effect	β	β	β
S2 – Treatment effect varies across episode	$\frac{\beta_1+p\beta_2}{1+p}$	$\frac{\beta_1+p\beta_2}{1+p}$	$\frac{(1-\vartheta^*+\vartheta^*p)\beta_1+p\beta_2}{(1+p-\vartheta^*+\vartheta^*p)}$
S3 – Treatment effect varies across value of M_i	$\frac{(1-p)\beta_1+2p\beta_2}{(1+p)}$	$\frac{(1-p)\beta_1+2p\beta_2}{1+p}$	$\frac{(1-p-\vartheta^*+\vartheta^*p)\beta_1+2p\beta_2}{(1+p-\vartheta^*+\vartheta^*p)}$
S4 – Treatment effect carries forward into the second episode	β	$\beta - \frac{p\gamma}{1+p}$	$\beta - \frac{p\gamma}{(1+p-\vartheta^*+\vartheta^*p)}$
S5 – Treatment becomes less effective on re-use	$\beta + \frac{p\delta}{2(1+p)}$	β	β
S6 – Constant treatment effect, differential non-enrolment based on outcome in previous episode	β_{trt}	$\beta_{trt} + \frac{\beta_{X_{PL}}(p_{01}-p_{00})}{2(1+p)}$	$\beta_{trt} + \frac{\beta_{X_{PL}}(1-p-\vartheta^*+\vartheta^*p)(p_{01}-p_{00})}{2(1-p)(1+p-\vartheta^*+\vartheta^*p)}$
S7 – Constant treatment effect, differential non-enrolment based on expected outcome in current episode	β_{trt}	$\beta_{trt} - \frac{\beta_{X_{EL}}(p_{00}-p_{01})}{2(1+p)}$	$\beta_{trt} - \frac{\beta_{X_{EL}}(p_{00}-p_{01})}{2(1+p-\vartheta^*+\vartheta^*p)}$

mechanisms than were considered in the mathematical derivations in the previous section.

I describe the estimands, methods of analysis, and performance measures, and data generating mechanisms below. I used 10,000 replications for all simulation scenarios.

5.3.1 Estimands

In this simulation study, I focus on the per-episode added-benefit and the per-episode policy-benefit estimands. I decided not to evaluate the per-patient estimands as I felt the per-episode estimands were sufficient to illustrate the pros and cons of each design (for instance, the bias exhibited in the crossover design when treatment history affects current episode will be there regardless of whether it is a per-episode or per-patient estimand).

I did not evaluate each trial design against each estimand; for instance, I did not assess whether the cluster design gave unbiased estimates for the added-benefit

estimand, as this trial is essentially designed to estimate the policy-benefit effect, and so will naturally not perform well against the added-benefit estimand. Similarly, I did not assess the crossover design against the policy-benefit estimand.

Therefore, for the per-episode added-benefit estimand I evaluated the (i) re-randomisation design; (ii) parallel group design; and (iii) crossover design. For the per-episode policy-benefit estimand, I evaluated the (a) re-randomisation design; (b) parallel group design; and (c) cluster design. A summary of this is given in table 5.6.

In the simulation study I generated data so that the same set of episodes were enrolled for each of the re-randomisation, cluster, and crossover designs, and so that the treatment history was the same for each design (further detail on this is given in the data generating sections below). This implies that each design can be compared against the same value of the estimand (as the specific estimand value can depend on the exact set of episodes enrolled in the trial and the distribution of the treatment history in the trial). For example, different distributions of the treatment history will lead to different values of the added-benefit estimands under data generating model 2.2 (where the treatment effect carries forward); forcing the treatment history to be the same for each trial design means that we can evaluate results from each design against the same value of the estimand. Note that it is only possible to force the re-randomisation, cluster, and crossover designs to have the same treatment history because of the maximum limit of two episodes in each trial, as this means the treatment history is based only on the episode 1 treatment allocation, which is independent of trial design.

I also evaluated the parallel group design against the same estimand values that were used for the other designs (i.e. against an estimand that was based on the set of episodes enrolled in the re-randomisation, cluster, and crossover designs, rather than the set of episodes enrolled in the parallel group design). This was to allow the parallel group design to be compared directly against the other designs, to see how well it was able to estimate the added-benefit and policy-benefit estimands in multi-episode settings.

5.3.2 Methods of analysis

I analysed data from the parallel group design using a linear regression model with treatment allocation as a covariate. I analysed data from the re-randomisation design using independence estimators (i.e. the same approach as in chapter 3); the one difference to chapter 3 was that I included an indicator variable for episode number in the model for the added-benefit estimand (the model for the policy-benefit estimand already included this term).

For the cluster and crossover designs, I used independence estimators with cluster-robust standard errors, with patients acting as the cluster. I also included an indicator variable for episode number for each of these designs. The Stata code for each estimator

Table 5.6: Summary of which trial designs are evaluated against which estimands

Trial design	**Estimand(s) against which the trial is evaluated**
Parallel group	• Per-episode added-benefit • Per-episode policy-benefit
Re-randomisation	• Per-episode added-benefit • Per-episode policy-benefit
Cluster	• Per-episode policy-benefit
Crossover	• Per-episode added-benefit

is provided in table 5.7.

The reason I included an indicator variable for episode number in the added-benefit model was to facilitate a more fair comparison with the parallel group design. Variation in outcomes across episodes affected re-randomisation, cluster, and crossover designs, but not parallel designs, which would increase the precision of the parallel group design compared to the other designs. Given that this source of variability can be controlled at the analysis stage for the re-randomisation, cluster, and crossover designs, it seemed most fair to do so, in order to compare precision based on the elements which cannot be controlled for in the analysis stage.

I decided not to include mixed-effects models for cluster or crossover designs in this simulation study. This is because we know they can be extremely biased in most settings considered in this thesis (based on the results from the mathematical derivations shown previously), and so comparing each design (re-randomisation, cluster, crossover, parallel group) based on the best method of analysis for that design seemed like the most fair approach.

5.3.3 Performance measures

I will evaluate the different designs in terms of bias, coverage, and the precision compared to re-randomisation trials. Bias, coverage, and precision were defined in chapters 3 and 4.

For bias and coverage, I omit results for the PG(small) design and instead only present results for the PG(big) design. This is because bias and coverage are the

Table 5.7: Stata code to implement independence estimators. 'y' denotes patient outcome, 'z' denotes treatment allocation, 'id' is a unique identifier for patient, 'm_i' denotes the number of episodes for which the patient is enrolled in the trial, 'z_prev' denotes the patient's treatment allocation in their previous episode (and is set to 0 if it is the patient's first episode), 'x_ep' denotes the episode (0=first episode, 1=second episode), and 'prop_1st_ep' and 'prop_2nd_ep' represent the proportion of episodes in the trial which are 1st and 2nd episodes respectively. In order to run the above code in Stata, 'prop_1st_ep', 'prop_2nd_ep', 'prop_has_1ep', and 'prop_has_2ep' must be saved as Stata local macros. All analyses are for per-episode estimands.

Estimator	Stata code
Re-randomisation	
Added-benefit	reg y z x_ep, vce(cluster id)
Policy-benefit	reg y z##z_prev x_ep, vce(cluster id) lincom /// 'prop_1st_ep'*_b[1.z] + /// 'prop_2nd_ep'*(_b[1.z]+_b[1.z_prev] + _b[1.z#1.z_prev])
Parallel group	reg y z
Cluster	reg y z x_ep, vce(cluster id)
Crossover	reg y z x_ep, vce(cluster id)

same in expectation for both designs, and so omitting the PG(small) design does not discard any important information and will make figures easier to read. Conversely, for precision I will include results for both the PG(big) and PG(small) designs, as the precision of the two designs is not the same.

5.3.4 Data generating methods

I based the simulations in this chapter on simulation studies 1 and 2a in chapter 3 (sections 3.4.2 and 3.4.3), with some differences which I discuss below in sections 5.3.4.1 and 5.3.4.2. I did not use simulation study 2b as I felt the results from studies 1 and 2a were sufficiently conclusive to demonstrate when each design is (and is not) appropriate, and that including study 2b would not add any additional information.

5.3.4.1 Simulation study 1: patients enrolled for all episodes they experience

This simulation study is based on a trial of 300 patients; 150 patients experience one episode during the trial period, and 150 experience two episodes.

I used the same six data generating mechanisms as in chapter 3, described in table 5.8. Data for re-randomisation, cluster, and crossover designs was generated based on the following general model for a continuous outcome:

$$Y_{ij} = \alpha + \beta_{trt} Z_{ij} + \beta_{ep} X_{ep_{ij}} + \beta_M X_{M_i} + \beta_{TRTxEP} Z_{ij} X_{ep_{ij}}$$
$$+\beta_{TRTxM} Z_{ij} X_{M_i} + \gamma Z_{i,j-1} + \delta Z_{ij} Z_{i,j-1} + \mu_i + \varepsilon_{ij}$$

For each scenario, I set $\alpha = 0$, $\beta_{trt} = 3$, $\beta_{ep} = 1$, $\beta_M = 1$, $\sigma^2_\mu = 5$ and $\sigma^2_\varepsilon = 5$. I generated μ_i and ε_{ij} independently; based on the chosen variances, the intraclass correlation between episodes from the same patient is 0.50 (conditional on the other variables in the data generating model). Values for other parameters are shown in table 5.8.

For each of the three designs (re-randomisation, cluster, crossover) I generated Z_{i1} (the episode 1 treatment allocation) using simple randomisation (based on a Bernoulli random variable with probability 0.5). I used the same value of Z_{i1} for each of the three designs (i.e. the treatment allocation in episode 1 for a particular patient would be the same under each design). I then generated Z_{i2} (the episode 2 treatment allocation) as follows: (a) for re-randomisation I used a Bernoulli random variable with probability 0.5 (generated independently from Z_{i1}); (b) for the cluster design, I set Z_{i2} to be the same as Z_{i1}; and (c) for the crossover design, I set Z_{i2} to be the opposite of Z_{i1} (i.e. if $Z_{i1} = 0$ then I set $Z_{i2} = 1$, and vice versa).

Data for the two parallel group designs, PG(big) and PG(small), were generated using the following general model for a continuous outcome:

$$Y_i = \alpha + \beta_{trt} Z_i + \beta_M X_{M_i} + \beta_{TRTxM} Z_i X_{M_i} + \varepsilon_i^{PG}$$

Note that this model omits terms associated with episode 2 (i.e. $\beta_{ep} X_{ep_{ij}}$, $\beta_{TRTxEP} Z_{ij} X_{ep_{ij}}$, $\gamma Z_{i,j-1}$, and $\delta Z_{ij} Z_{i,j-1}$); this is because in the parallel group design there is only one episode. The other main difference compared to the data generating model for the re-randomisation, cluster, and crossover designs is the inclusion of the term ε_i^{PG} in place of the terms σ^2_μ and σ^2_ε. ε_i^{PG} is a residual error term which encompasses both within- and between-patient variation, where $\varepsilon_i^{PG} \sim N\left(0,\ \sigma^2_\mu + \sigma^2_\varepsilon\right)$. Therefore, $V\left(\varepsilon_i^{PG}\right) = \sigma^2_\mu + \sigma^2_\varepsilon$ (i.e. the residual variation in the parallel group designs is the same as in the re-randomisation, cluster, and crossover designs). I used the same values for all other parameters as in table 5.8. For PG(small) I used 300 individual patients (equivalent to the number of individual patients enrolled in the re-randomisation, cluster, and crossover designs), and for PG(big) I used 450 individual patients (equivalent to the number of episodes enrolled in the re-randomisation, cluster, and crossover designs).

5.3.4.2 Simulation study 2a: some patients do not re-enrol for their 2nd episode

As before, this simulation study is based on a trial of 300 patients; 150 patients experience one episode during the trial period, and 150 experience two episodes. All

Table 5.8: Simulation parameters for different scenarios (simulation scenario 1). For all scenarios, I set $\alpha = 0$, $\beta_{trt} = 3$, $\beta_{ep} = 1$, $\beta_M = 1$, $\sigma^2_\mu = 5$ and $\sigma^2_\varepsilon = 5$.

Scenario	β_{TRTxEP}	β_{TRTxM}	γ	δ
Scenario 1: Constant treatment effect	0	0	0	0
Scenario 2: Treatment effect varies across episode	1.5	0	0	0
Scenario 3: Treatment effect varies across patients with different values of M_i	0	3	0	0
Scenario 4: Treatment effect carries forward	0	0	1	0
Scenario 5: Treatment becomes less effective on re-use	0	0	0	-3
Scenario 6: Treatment effect varies across episodes, across patients with different values of M_i, carries forward, and becomes less effective on re-use	1.5	3	1	-3

patients are enrolled for their first episode, but a subset of patients who experience two episodes do not re-enrol for their second episode. Therefore, $N_T = 300$ and $M_{T(1)} = 150$, however $M_{T(2)} < 150$ and $M_T < 450$; the exact values of $M_{T(2)}$ and M_T vary across simulation replications.

I used the same six data generating mechanisms, and the same five non-enrolment scenarios as in chapter 3. For re-randomisation, cluster, and crossover designs I simulated data by first generating outcomes for all 450 episodes (regardless of whether they were enrolled in the trial) using the model:

$$Y_{ij} = \alpha + \beta_{trt} Z_{ij} + \beta_{ep} X_{ep_{ij}} + \beta_M X_{M_i} + \beta_{TRTxEP} Z_{ij} X_{ep_{ij}} + \beta_{TRTxM} Z_{ij} X_{M_i} + \gamma Z_{i,j-1} + \delta Z_{ij} Z_{i,j-1} + \beta_{X_{PL}} X_{PL_i} + \beta_{X_{EL}} X_{EL_{ij}} + \mu_i + \varepsilon_{ij}$$

I then generated an indicator for each episode to denote whether it was enrolled in the trial or not using the model:

$$P\left(R_{i2} = 0\right) = \alpha^{R_2} + \gamma^{R_2} Z_{i,j-1} + \beta^{R_2}_{X_{PL}} X_{PL_i} + \beta^{R_2}_{X_{EL}} X_{EL_{i2}} + \delta^{R_2}_{Xpl} Z_{i,j-1} X_{PL_i} + \delta^{R_2}_{Xel} Z_{i,j-1} X_{EL_{i2}}$$

where R_{i2} denotes whether the patient was enrolled at episode 2 (where 1=enrolled, 0=not enrolled).

I then performed analysis only on the subset of enrolled episodes. The different treatment effect and non-enrolment scenarios are described below.

It should be noted that 2nd episode non-enrolment does not affect parallel group trials (as there is no 2nd episode in these designs). However, I present results from parallel group designs under each non-enrolment scenario for comparison. For PG(small) I used 300 individual patients (equivalent to the number of individual patients enrolled in the re-randomisation, cluster, and crossover designs), and for PG(big) I used M_T individual patients (equivalent to the number of episodes enrolled in the re-randomisation, cluster, and crossover designs); note that the number of patients for PG(big) therefore varied across simulation replications. I generated outcomes for the parallel group designs using the model:

$$Y_{ij} = \alpha + \beta_{trt}Z_{ij} + \beta_M X_{M_i} + \beta_{TRTxM}Z_{ij}X_{M_i} + \beta_{X_{PL}}X_{PL_i} + \beta_{X_{EL}}X_{EL_i} + \mu_i + \varepsilon_i^{PG}$$

where ε_i^{PG} was described in the previous section.

I used the same parameter values to generate outcomes as in the previous section, except I included the parameters $\beta_{X_{PL}}$ and $\beta_{X_{EL}}$. To generate non-enrolment, I set $\alpha^{R_2} = 0.05$ and $\gamma^{R_2} = 0.10$ for all scenarios; values for the parameters $\beta_{X_{PL}}$, $\beta_{X_{EL}}$, $\delta^{R_2}_{Xpl}$, and $\delta^{R_2}_{Xel}$ are shown in table 5.9; all other parameter values are the same as in the previous section.

5.4 Simulation study (results)

A summary of when the different designs were biased is given in table 5.10.

5.4.1 Simulation study 1: patients enrolled for all episodes they experience

Results for the per-episode added-benefit estimand are shown in figure 5.2. As shown in chapter 3, the re-randomisation design was unbiased in all settings. The crossover design was biased in scenarios where the treatment history affected the current episode (scenarios 4-6), but was unbiased when treatment history had no effect (scenarios 1-3). The parallel group trial design was biased in all scenarios except when the treatment effect was constant (scenario 1) or when the treatment effect carried forward (scenario 4). All designs had good coverage for settings in which they were unbiased. The crossover design had much higher precision than the re-randomisation design in all scenarios. The parallel group design with a large sample size had similar precision to re-randomisation in most settings, whereas with a small sample size it had lower precision than re-randomisation in all scenarios.

Results for the per-episode policy-benefit estimand are shown in figure 5.3. The re-randomisation and cluster designs were unbiased in all scenarios. The parallel group

Table 5.9: Simulation parameters for different scenarios (simulation scenario 2a). For all scenarios, I set $\alpha = 0$, $\beta_{trt} = 3$, $\beta_{ep} = 1$, $\beta_M = 1$, $\sigma^2_\mu = 5$, $\sigma^2_\varepsilon = 5$, $\alpha^{R_2} = 0.05$ and $\gamma^{R_2} = 0.10$

Scenario	$\beta^{R_2}_{X_{PL}}$	$\beta^{R_2}_{X_{EL}}$	$\beta_{X_{PL}}$	$\beta_{X_{EL}}$	$\delta^{R_2}_{Xpl}$	$\delta^{R_2}_{Xel}$
Scenario 1 – Non-enrolment depends on previous treatment allocation	0	0	0	0	0	0
Scenario 2 – Non-enrolment depends on previous treatment allocation and previous outcome	0.25	0	10	0	0	0
Scenario 3 – Non-enrolment depends on previous treatment allocation and baseline prognosis at episode 2	0	0.25	0	10	0	0
Scenario 4 – Non-enrolment is differential between treatment groups based on previous outcome	0	0	10	0	0.5	0
Scenario 5 – Non-enrolment is differential between treatment groups based on baseline prognosis at episode 2	0	0	0	10	0	0.5

design was biased in all scenarios except when the treatment effect was constant. All designs had good coverage for settings in which they were unbiased. The cluster and parallel group design with a large sample size were both more precise than the re-randomisation design in all scenarios. The parallel group design with a small sample size was more efficient than re-randomisation in most, but not all scenarios.

The reason the policy-benefit estimate from a re-randomisation trial was generally less precise than other designs is likely because this estimator is based on a linear combination of parameters, some of which are estimated with very low levels of precision. For example, the parameter δ is based on the interaction $Z_{ij}Z_{i,j-1}$ which has far fewer patients than the main effect (based on Z_{ij}), and so has higher variance. Conversely, both the cluster and parallel group designs are estimated using a single parameter (which is based only on the term Z_{ij}) and do not require estimation of interactions or combinations of parameters, and so have lower variance.

5.4.2 Simulation study 2a: some patients do not re-enrol for their 2nd episode

Results for the per-episode added-benefit estimand are shown in figures 5.4-5.6. As shown in chapter 3, the re-randomisation design was unbiased across all treatment effect mechanisms and non-enrolment scenarios. For non-enrolment scenarios 1-3 the crossover design was biased in treatment effect scenarios where the treatment history affected the current episode (treatment effect scenarios 4-6), but was unbiased when treatment history had no effect (treatment effect scenarios 1-3); however, in

Table 5.10: Summary of simulation results (scenario 1 and 2a)

Design and target estimand	**Biased when:**
Parallel group	
Per-episode added-benefit	• Treatment effect is not constant
Per-episode policy-benefit	• Treatment effect is not constant • Treatment effect carries forward
Re-randomisation	
Per-episode added-benefit	• NA (unbiased in all settings)
Per-episode policy-benefit	• There is differential non-enrolment between treatment arms based on outcome in previous episode or prognosis in current episode
Cluster	
Per-episode policy-benefit	• There is differential non-enrolment between treatment arms based on outcome in previous episode or prognosis in current episode
Crossover	
Per-episode added-benefit	• Treatment history affects outcome or treatment effect in current episode • There is differential non-enrolment between treatment arms based on outcome in previous episode or prognosis in current episode

Figure 5.2: Bias, coverage, and precision of re-randomisation, crossover, and parallel group designs for the per-episode added-benefit effect in simulation study 1 (no non-enrolment). RR(ind)=re-randomisation design. Crossover(ind)=crossover design. PG(big)=parallel group design with 450 patients. PG(small)=parallel group design with 300 patients. Error bars are 95% confidence intervals based on Monte Carlo standard errors. Precision is calculated relative to the re-randomisation design with independence estimator.

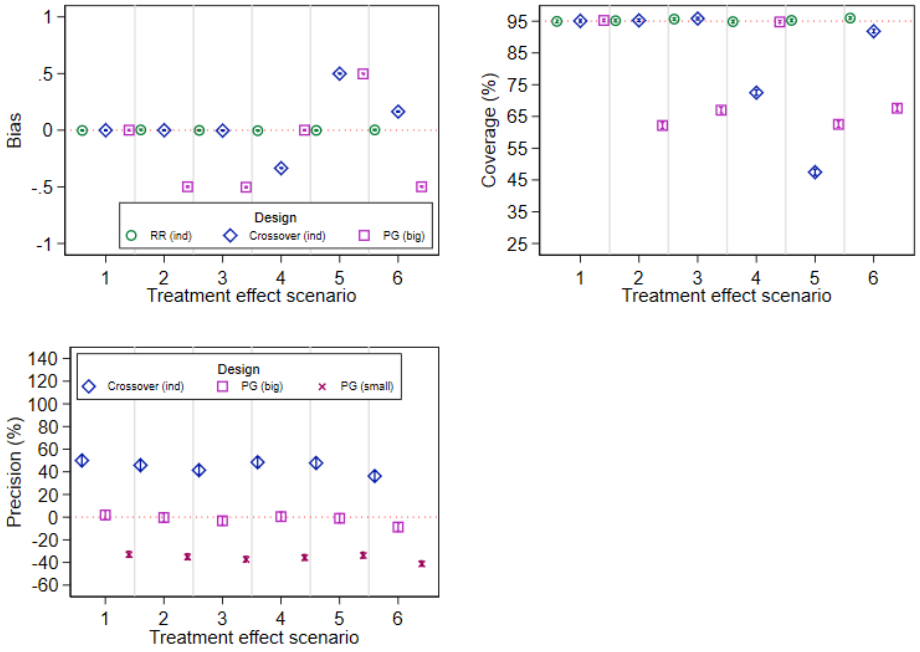
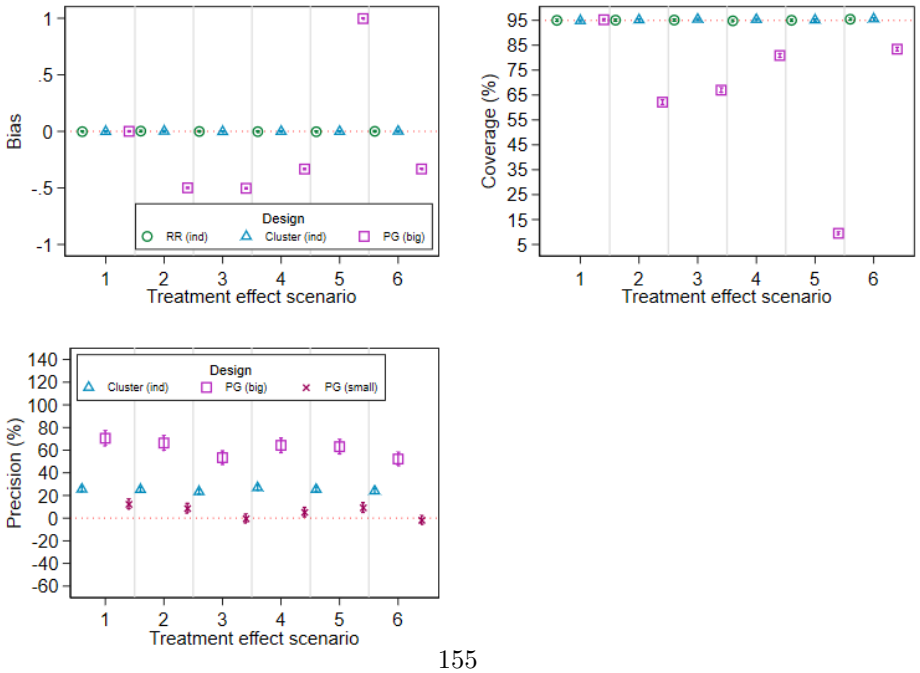


Figure 5.3: Bias, coverage, and precision of re-randomisation, cluster, and parallel group designs for the per-episode policy-benefit effect in simulation study 1 (no non-enrolment).



non-enrolment scenarios 4 and 5 the crossover design was extremely biased for all treatment effect scenarios. All designs had good coverage for settings in which they were unbiased. Results for precision were broadly similar to when there was no non-enrolment, although the precision of the crossover design increased in non-enrolment scenarios 2 and 4 and decreased in non-enrolment scenarios 3 and 5 compared to re-randomisation. The reason for these changes is that the inclusion of X_{PL_i} and $X_{EL_{ij}}$ in the data generating model for the outcome changes both the ICC and the overall variance (with X_{PL_i} increasing the ICC and $X_{EL_{ij}}$ reducing the ICC, and both increasing the overall variance), which affects the precision of each design differently.

Results for the per-episode policy-benefit estimand are shown in figures 5.7-5.9. The re-randomisation and cluster designs were both unbiased in non-enrolment scenarios 1-3, and were both biased in non-enrolment scenarios 4-5. All designs had good coverage for settings in which they were unbiased. Results for precision were similar to when there was no non-enrolment, though as above changes to the ICC due to inclusion of X_{PL_i} and $X_{EL_{ij}}$ in the data generating model for the outcome in non-enrolment scenarios 2-5 has an impact on the precision.

5.5 Discussion

In this chapter I compared the re-randomisation design with three alternate design choices for multi-episode settings; (a) parallel group design; (b) cluster design; and (c) crossover design. There are several key takeaways here. First, mixed-effects models are biased for cluster and crossover designs whenever the treatment effect is not constant, or under certain non-enrolment mechanisms. Therefore, mixed-effects models should not be used with these designs, and independence estimators should be used instead. This is a similar conclusion to that of chapter 4, which evaluated mixed-effects models in re-randomisation trials. It also follows on from the informative cluster size literature, which has shown that mixed-effects models are not appropriate when cluster size is informative [44].

The second takeaway is that the choice of which design to use will depend on the specific aims of the study. If our main aim is to estimate a policy-benefit effect, then either re-randomisation or a cluster design could be used, and each has its own benefits and drawbacks. The benefits of the re-randomisation design are that it maintains allocation concealment across all episodes, and that it can be used to estimates other effects in addition to the policy-benefit effect. Its drawbacks are that it requires specification of a causal model for treatment history in order to estimate the policy-benefit effect, which may lead to bias if the model is wrong. Even if the causal model is correct, it can still be biased for the policy-benefit effect under certain non-enrolment mechanisms. Furthermore, the policy-benefit estimates from the re-randomisation design can be imprecise.

The benefits of the cluster design are that the policy-benefit effect can be estimated

Figure 5.4: Bias of re-randomisation, crossover, and parallel group designs for the per-episode added-benefit effect in simulation study 2a (with non-enrolment). RR(ind)=re-randomisation design. Crossover(ind)=crossover design. PG(big)=parallel group design with M_T patients. PG(small)=parallel group design with 300 patients. Error bars are 95% confidence intervals based on Monte Carlo standard errors. NE sc = non-enrolment scenario.

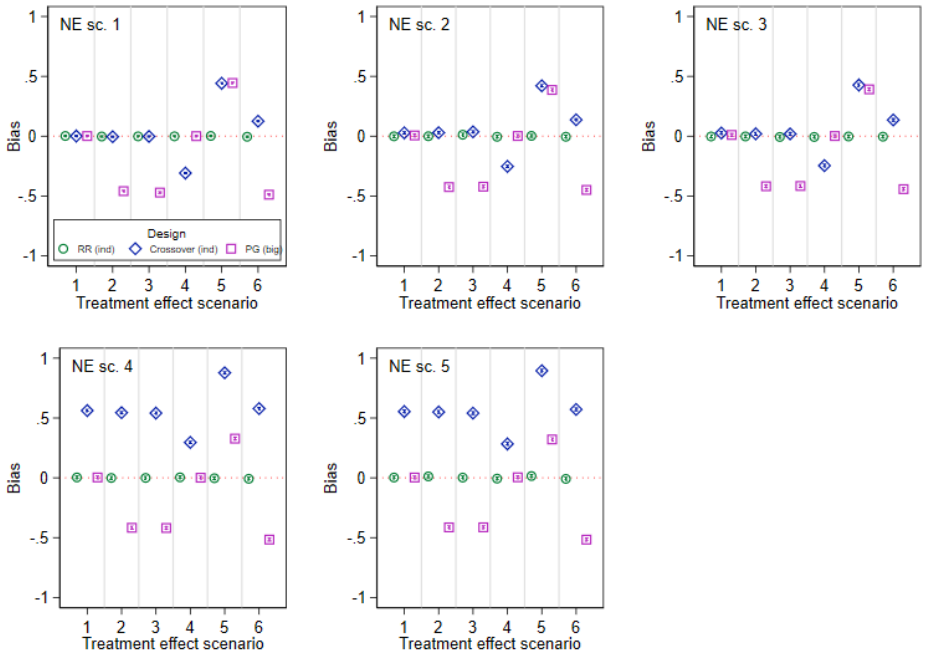


Figure 5.5: Coverage of re-randomisation, crossover, and parallel group designs for the per-episode added-benefit effect in simulation study 2a (with non-enrolment).

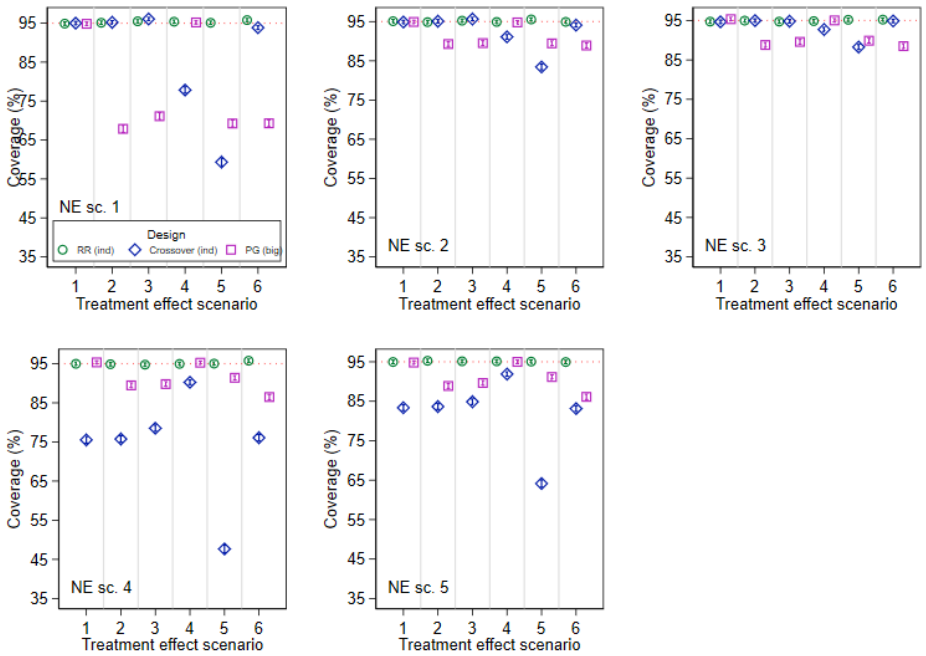
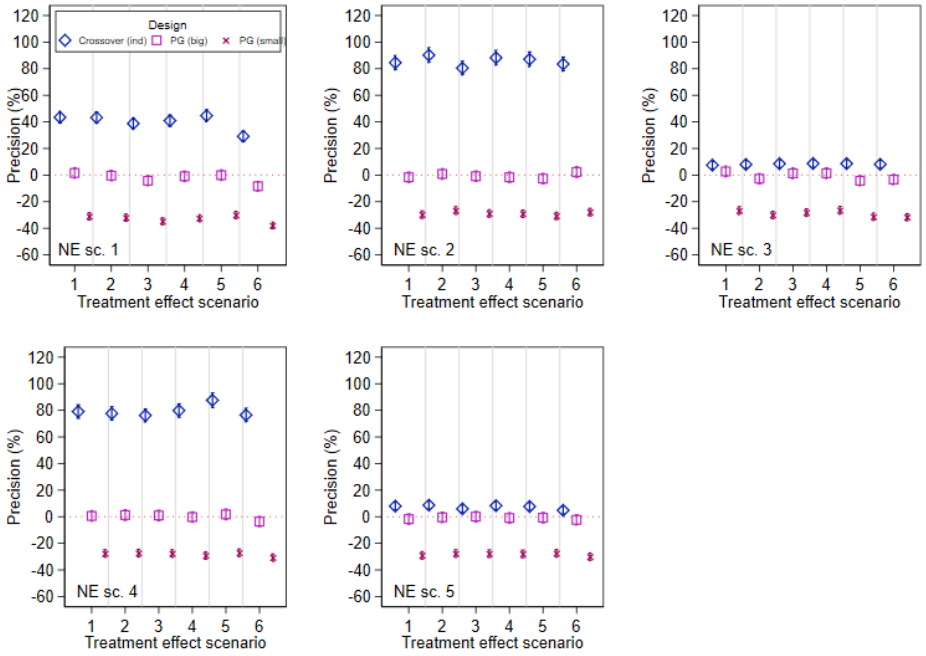


Figure 5.6: Precision of crossover and parallel group compared to re-randomisation designs for the per-episode added-benefit effect in simulation study 2a (with non-enrolment). RR(ind)=re-randomisation design with independence estimator. Crossover(ind)=crossover design with independence estimator. PG(big)=parallel group design with M_T patients. PG(small)=parallel group design with 300 patients. Error bars are 95% confidence intervals based on Monte Carlo standard errors. Precision is calculated relative to the re-randomisation design with independence estimator. NE sc = non-enrolment scenario.



without assuming any sort of causal model for treatment history. It also has higher precision for the policy-benefit effect than the re-randomisation design. Its drawbacks are that it does not maintain allocation concealment past the first episode, and so is at risk of bias from selective enrolment. It can also provide biased estimates under the same non-enrolment scenarios as the re-randomisation design. Furthermore, it is only suited to estimating the policy-benefit effect, so if both the policy-benefit and added-benefit effects are of interest the cluster design may not be suitable. The decision on which design to use may come down to whether the risk of selection bias through the cluster design is thought to be greater than the risk of bias from misspecification of the causal model for treatment history in the re-randomisation design; though as a secondary consideration, it is worth noting that the re-randomisation design can at least ensure unbiased estimates of at least one effect (added-benefit), while the cluster design can offer no such guarantee.

For the added-benefit effect, the crossover design is biased when treatment history affects current episode or under certain non-enrolment scenarios. In comparison, the re-randomisation design is unbiased for the added-benefit effect across all scenarios. Although the crossover design has much higher precision than the re-randomisation

Figure 5.7: Bias of re-randomisation, crossover, and parallel group designs for the per-episode policy-benefit effect in simulation study 2a (with non-enrolment). RR(ind)=re-randomisation design. Crossover(ind)=crossover design. PG(big)=parallel group design with M_T patients. PG(small)=parallel group design with 300 patients. Error bars are 95% confidence intervals based on Monte Carlo standard errors. NE sc = non-enrolment scenario.

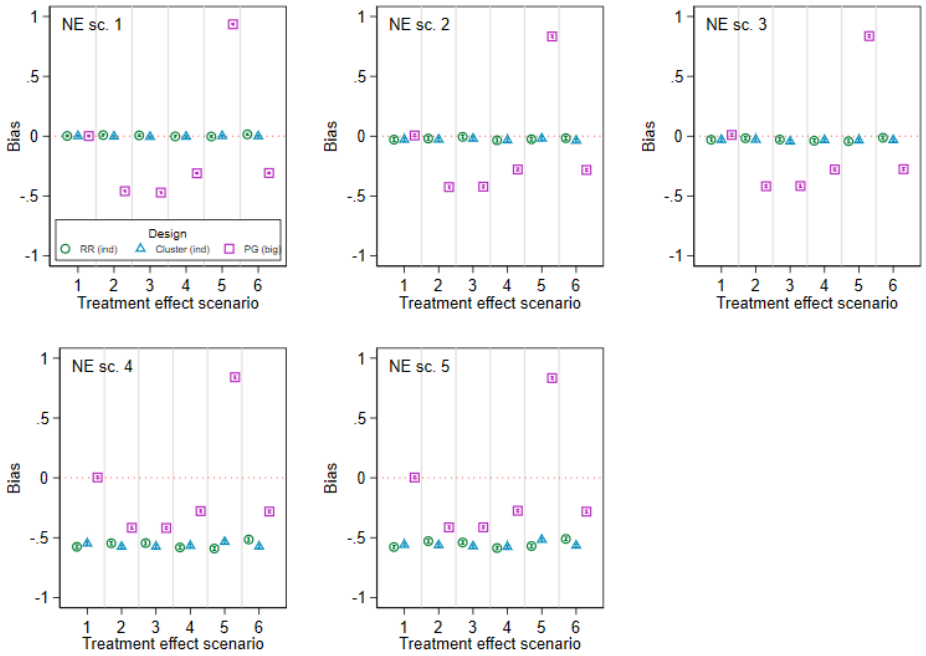


Figure 5.8: Coverage of re-randomisation, crossover, and parallel group designs for the per-episode policy-benefit effect in simulation study 2a (with non-enrolment).

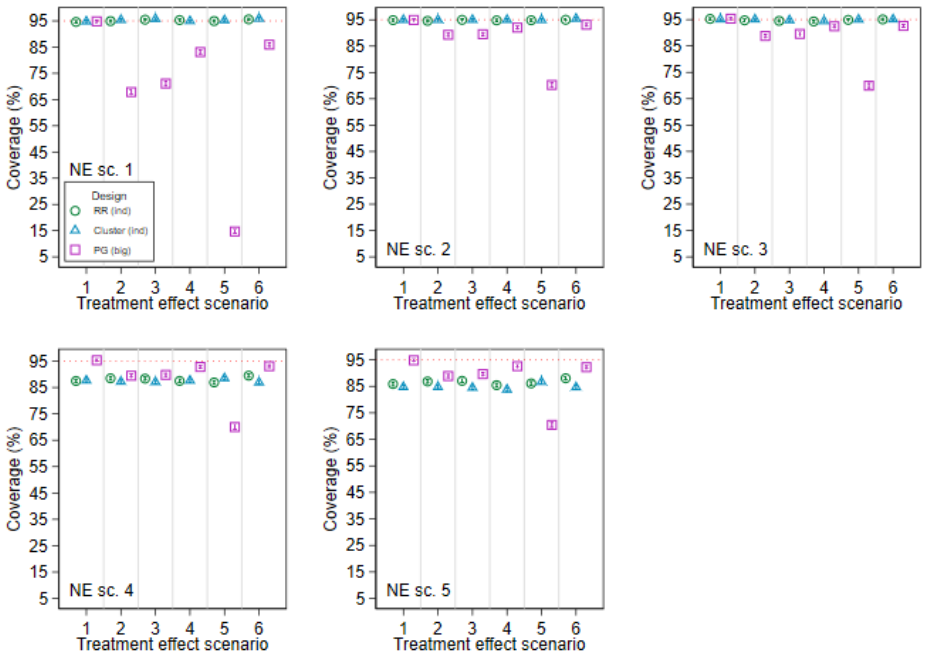
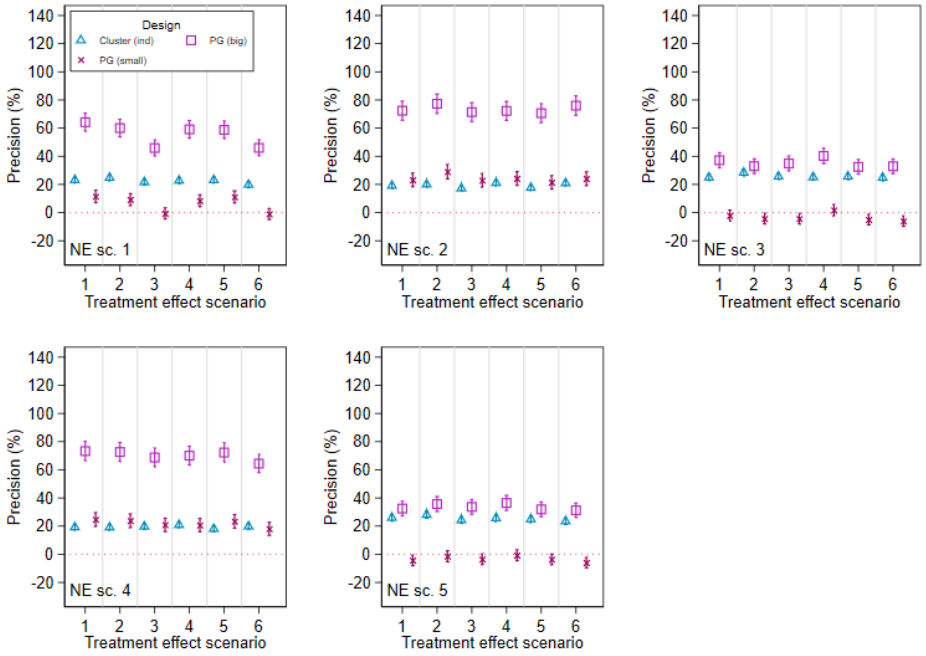


Figure 5.9: Precision of crossover and parallel group compared to re-randomisation designs for the per-episode policy-benefit effect in simulation study 2a (with non-enrolment). RR(ind)=re-randomisation design with independence estimator. Crossover(ind)=crossover design with independence estimator. PG(big)=parallel group design with M_T patients. PG(small)=parallel group design with 300 patients. Error bars are 95% confidence intervals based on Monte Carlo standard errors. Precision is calculated relative to the re-randomisation design with independence estimator. NE sc = non-enrolment scenario.



design, this increased precision is not worth the major risk in bias from using the crossover design. Therefore, I recommend the crossover design not be used, and that re-randomisation should be the design of choice for the added-benefit effect.

The parallel group design is not appropriate for either the added-benefit or policy-benefit effect. However, it is appropriate for the effect of the first time an intervention is used (the episode 1 effect). If the main study aim is to estimate the episode 1 effect, then the parallel group design is a valid choice. An alternative approach in this setting would be to either use a re-randomisation or cluster design that was powered for the episode 1 effect (i.e. enrolled the same number of individual patients as a parallel group trial), but also enrolled patients for their second (or later) episodes. This way, in addition to getting the same episode 1 effect estimate as you would in a parallel group design, you could also answer additional questions about the intervention, such as the episode 2 effect, or the added-benefit or policy-benefit effect across episodes. This approach may not be powered for these secondary aims, but an imprecise estimate of these effects may still be better than no estimate at all. The main downside to this approach is the additional expense from enrolling additional episodes (though it should be noted this design would not require any additional time for recruitment, as

it would finish enrolling the required number of individual patients as a parallel group design).

I therefore recommend that (i) the re-randomisation design be used for the added-benefit effect; (ii) the re-randomisation or cluster design be used for the policy-benefit effect; (iii) the parallel group, cluster, or re-randomisation design be used for the episode 1 effect (provided they are appropriately powered for this effect); (iv) the crossover design not be used in multi-episode settings; and (v) mixed-effects models not be used for designs in multi-episode settings, and independence estimators be used instead.

6 A review of re-randomisation trials in febrile neutropenia

In this chapter I review a set of re-randomisation trials in febrile neutropenia. Febrile neutropenia occurs when neutropenic patients (those with abnormally low neutrophil granulocyte counts) develop fever. It is often a complication for patients with cancer who receive chemotherapy regimens which suppress bone marrow activity. Because chemotherapy is usually given in multiple cycles, patients may develop febrile neutropenia multiple times during the course of their cancer treatment, and each episode of febrile neutropenia would require medical intervention. Standard care for febrile neutropenia is broad-spectrum antibiotics. However, it has been suggested that granulocyte colony-stimulating factor (G-CSF) could be useful in this setting, as it regulates the production of the neutrophil lineage. Re-randomisation could be an appropriate design in the setting of febrile neutropenia, as some patients experience multiple episodes and require treatment for each episode, and the intervention (G-CSF) and patient follow-up duration are typically short-term.

In 2014, Mhaskar *et al* [61] published a Cochrane systematic review and meta-analysis which evaluated the use of G-CSF with antibiotics vs. antibiotics alone in patients with febrile neutropenia. This systematic review included trials published up to 2014. It included 14 trials in total, 9 of which used a parallel group design, and 5 which used a re-randomisation design. The trials which used a parallel group design allowed patients to enrol for a single episode of febrile neutropenia only; they were not allowed to enrol for any subsequent episodes of febrile neutropenia. Conversely, the trials which used a re-randomisation design did allow patients to be re-enrolled and re-randomised for subsequent episodes of febrile neutropenia (one of these trials was described in chapter 1, section 1.4.3).

The systematic review by Mhaskar *et al* [61] presents a unique opportunity to look at how some re-randomisation trials have been designed and analysed, and how the use of re-randomisation may have impacted results. I have four objectives for this review: to evaluate (1) design and analysis characteristics of the re-randomisation trials; (2) the impact re-randomisation had on recruitment; (3) whether re-randomisation led to higher rates of non-compliance or loss-to-follow-up in subsequent episodes; and (4) whether treatment effect estimates from the re-randomisation trials were different to those from the parallel group trials. I summarise the methods below, and further

detail on each of these objectives is given in the sections below. This chapter has been published as an article in the *Journal of Clinical Epidemiology* (*Kahan, B.C., et al., Re-randomization increased recruitment and provided similar treatment estimates as parallel designs in trials of febrile neutropenia. J Clin Epidemiol, 2018. 97: p. 14-19*) [39].

6.1 Methods

Full details of the search strategy, inclusion criteria, and data collection procedure are available in the publication by Mhaskar *et al* [61]. For each article, I extracted relevant information relating to the objectives above. In order to ensure there were no errors in data extraction, another reviewer (Tim Morris, Medical Research Council Clinical Trials Unit at UCL) extracted data in parallel. We compared extractions, and discrepancies were resolved through discussion.

6.1.1 Design and analysis characteristics of re-randomisation trials

I evaluated whether the five re-randomisation trials in the review had been designed according to the two core design requirements described in chapter 1 (section 1.3, table 1.1); (i) patients are only re-enrolled if they have completed the follow-up period from their previous randomisation; and (ii) randomisations for the same patient are performed independently.

I also looked at whether these trials placed any constraints on the number of episodes for which patients could be enrolled (for example, saying patients could be enrolled for a maximum of four episodes), and whether trials explicitly reported the number of episodes that each patient was enrolled for.

I also evaluated how these trials were designed and analysed. In chapter 1, I described previous research showing that the sample size from a parallel group trial could be used for a re-randomisation trial (section 1.6.4). I therefore assessed how many trials based their sample size calculation on that of a parallel group design. For the analysis, I looked at whether they used an independence estimator or used a method that allowed for correlation between episodes from the same patient, such as a mixed-effects model. I also looked at whether they analysed data on a per-episode or per-patient basis; I did not look at whether they used an added-benefit or policy-benefit estimator, as I had not developed these estimands at the time of doing this review. However, this information was retrospectively completed (this is discussed further in section 6.2.1).

6.1.2 Impact of re-randomisation on recruitment

For the five re-randomisation trials I evaluated to what extent recruitment had been increased through the use of re-randomisation. I did this by looking at the number of individual patients enrolled and the number of episodes enrolled in each trial. I then

divided the number of episodes by the number of individual patients; this measure represents the extra number of episodes each trial gained by using a re-randomisation design instead of a parallel group design.

6.1.3 Treatment compliance and loss-to-follow-up in re-randomisation trials

It is plausible that in some settings repeated enrolments in re-randomisation trials may place undue burden on patients due to increased treatment or follow-up burden, and may lead to higher rates of non-compliance or loss-to-follow-up in subsequent enrolments. For instance, this may occur in trials where patients must fill out numerous long questionnaires at each follow-up visit. Due to their length, they may find these questionnaires frustrating to complete, and may become more frustrated each time they are required to complete the questionnaires. Then, the more times they are re-enrolled, the more likely they are at some point to stop filling out the questionnaires.

For the five trials which used re-randomisation, I looked at whether the use of re-randomisation may have led to higher rates of non-compliance or loss-to-follow-up in later episodes. I assessed the number of episodes for which patients did not comply with the treatment protocol, and for outcomes reported by two or more re-randomisation trials, the number of episodes excluded from the analysis due to missing data. I wanted to look at this data separately for each episode (i.e. for the first episode vs. the second episode, etc), to assess whether non-compliance and missing data was higher in later episodes. However, this information was not reported separately by episode number (i.e. first vs. second episode) in any trial, and so I instead looked at the overall rates of non-compliance and loss-to-follow-up across all episodes.

6.1.4 Difference in treatment effect estimates between re-randomisation and parallel group trials

As discussed in chapters 1 and 5, re-randomisation trials and parallel group trials may give different treatment effect estimates. This is because they target different estimands; parallel group trials target the episode 1 effect (the effect of the intervention the first time it is used), whereas re-randomisation trials can target estimands based on all episodes (not just the first). Therefore, we would expect results from parallel group and re-randomisation trials to differ when these estimands differ, e.g. when the treatment effect in episode 1 is different to the effect in subsequent episodes.

The best way to evaluate whether the inclusion of subsequent episodes affects results would be to re-analyse each re-randomisation trial by restricting the analysis to only 1st episodes to see if this changes the estimates. However, this requires individual patient data, which I do not have access to for these trials. An alternative approach is to compare the estimated treatment effects from re-randomisation and parallel group trials. This will give an unbiased estimate of the effect of allowing enrolment of subsequent episodes under the assumption of no confounding (i.e. that there are no confounding factors that influence both a trial's design and its observed treatment

effect). Note that these results would be conditional on the type of analysis implemented in the re-randomisation trials (i.e. the difference in estimates between parallel and re-randomisation trials would depend on whether the re-randomisation trials used per-episode added-benefit vs per-patient policy-benefit estimators, etc).

In order to implement this analysis, I extracted the treatment effect estimate and 95% confidence interval for each outcome reported in each of the 14 trials included in the review. However, I only compared treatment effects between parallel and re-randomisation trials for outcomes that were reported for at least two re-randomisation and two parallel group trials. This was to avoid having results being entirely reliant on a single trial. This led to five outcomes being included in the analysis; overall mortality, infection related mortality, hospital stay > 10 days, duration of neutropenia, and time to recovery from fever.

For each of these five outcomes, I used the Stata package *metareg* to conduct a random-effects meta-regression model to estimate the difference in treatment effect estimates between re-randomisation and parallel group trials. Treatment effect estimates were log(hazard ratio) (log(HR)) for time-to-event outcomes, log(risk ratio) (log(RR)) for binary outcomes, and standardised mean differences for continuous outcomes. A negative difference in effect sizes indicates that re-randomisation trials showed a more beneficial treatment effect than parallel group trials.

I found that two re-randomisation trials did not have any events for overall mortality, and two trials (one re-randomisation, one parallel group) did not have any events for infection-related mortality. These trials were excluded from the analysis of these outcomes, as it was impossible to estimate either a treatment effect or standard error. One parallel group trial included three treatment arms (two active, one control), and involved two treatment comparisons (both active interventions vs. control). The meta-analysis by Mhaskar *et al* [61] included both treatment comparisons in the analysis as separate trials; I therefore used the same approach here.

6.2 Results

6.2.1 Design and analysis characteristics of re-randomisation trials

Results are shown in table 6.1. None of the five trials explicitly stated whether randomisations for the same patient were independent, or that patients were only re-enrolled once the follow-up period from their previous enrolment was complete.

One trial reported placing a limit on the number of times patients could be enrolled (they set a maximum limit of four episodes per patient). None of the other four trials reported using a limit (or reported that no such limit existed). Three trials reported the number of episodes for which each patient was enrolled (in two trials all patients were enrolled for one or two episodes, and in the third most patients were enrolled for one or two treatment episodes and a small proportion were enrolled for three or more

episodes); the other two did not report this information.

Of the three trials that reported a sample size calculation, all three based their calculation on a parallel group design. All five trials analysed the data on a per-episode basis, and each of the five analysed data using an independence estimator.

Although I did not explicitly look at whether trials used added-benefit or policy-benefit estimators, all trials analysed data as if it were a parallel group trial where episodes were individual patients (e.g. using a two-sample t-test on episodes); this method of analysis is essentially equivalent to the per-episode added-benefit estimator in chapter 3 (i.e. analysis model 3.1), implying that all five trials used a method of analysis which targeted the per-episode added-benefit estimand.

6.2.2 Impact of re-randomisation on recruitment

Amongst the five trials using re-randomisation, the median number of individual patients recruited was 40 (range 28 to 112) and the median number of episodes of febrile neutropenia enrolled was 58 (35 to 186). The median increase in the sample size obtained through the use of re-randomisation was 25% (range 16% to 66%), indicating that using a re-randomisation design allowed trials to recruit between 16% and 66% more episodes of febrile neutropenia than they would have under a parallel group design.

6.2.3 Treatment compliance and loss-to-follow-up in re-randomisation trials

Results are shown in table 6.2. The median percent of episodes which were non-compliant with the protocol was 1.7% (range across trials 0% to 8.9%). None of the five outcomes I assessed excluded any episodes from the analysis due to missing data.

6.2.4 Difference in treatment effect estimates between re-randomisation and parallel group trials

Differences in treatment effect estimates between re-randomisation and parallel group trials are shown in figure 6.1. In this figure, the two blue lines denote the point estimate and 95% CI for the re-randomisation and parallel group trials (the top blue line is for re-randomisation), and the red line shows the difference in point estimates between the two designs (re-randomisation vs. parallel group), with a 95% CI for the difference. If the point estimate for the red line is far away from 0, this means that the two designs are showing different estimates; if it is close to 0, it means the two designs are showing similar estimates.

I found that treatment effect estimates for the two designs were similar for each of the five outcomes, and none of the differences were statistically significant. However, confidence intervals were wide, indicating that differences are possible.

Table 6.1: Design and analysis characteristics of re-randomisation trials

	Re-randomisation trials (n=5)
Explicitly stated that randomisations for the same patient were independent	
No	5 (100)
Yes	0 (0)
Explicitly stated that patients were only re-randomised when the follow-up period from their previous enrolment was complete	
No	5 (100)
Yes	0 (0)
Placed limit on maximum number of times each patient could be enrolled in the trial	
No	0 (0)
Yes	1 (20)
Not reported	4 (80)
Based sample size calculation on a parallel group design	
No	0 (0)
Yes	3 (60)
No sample size calculation reported	2 (40)
Analysed data on a per-episode or per-patient basis	
Per-episode	5 (100)
Per-patient	0 (100)
Both	0 (100)
Did analysis account for correlation between treatment episodes from the same patient?	
No (independence estimator)	5 (100)
Yes (e.g. mixed-effects model)	0 (0)
Reported the number of treatment episodes for each patient	
No	2 (40)
Yes	3 (60)

Table 6.2: Compliance and loss-to-follow-up in re-randomisation trials

	Number of trials reporting measure	**Median (range)**
Percent of episodes non-complying with treatment protocol	5	1.7 (0, 8.9)
Percent of episodes excluded from analysis due to missing data		
Overall mortality	5	0 (0, 0)
Infection related mortality	4	0 (0, 0)
Hospitalisation >10 days	3	0 (0, 0)
Duration of grade IV neutropenia	2	0 (0, 0)
Time to recovery from fever	3	0 (0, 0)

6.3 Discussion

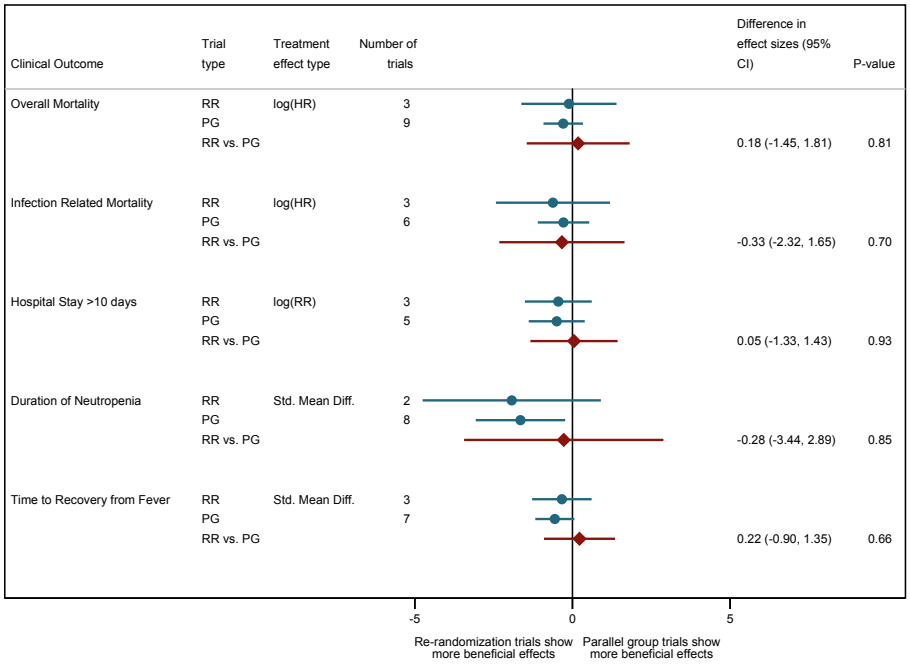
In this chapter I used a Cochrane review of granulocyte colony-stimulating factors for patients with febrile neutropenia which include both parallel group and re-randomisation trials to assess how re-randomisation trials are designed and analysed, what impact re-randomisation had on recruitment, whether it led to higher rates of non-compliance or missing data, and whether treatment estimates differed between the two designs.

I found that investigators tended to design and analyse the re-randomisation trials as if they were parallel group trials. For instance, all trials which reported a sample size calculation used the sample size calculation for a parallel group trial, and recruited the specified number of episodes. As discussed in chapter 1, this is a valid approach to the sample size calculation under certain assumptions. Similarly, all trials analysed data as though it were a parallel group trial, and the episodes were patients. This corresponds to a per-episode added-benefit estimator, which, as shown in chapter 3, is an unbiased estimator for the per-episode added-benefit estimand. Therefore, using a re-randomisation design rather than a parallel group trial design does not necessarily require additional methodological complexity, and can be done in a very simple way.

I found that the five re-randomisation trials recruited between 16-66% more episodes than they would have had they used a parallel group design. There was no missing data in the re-randomisation trials for any of the outcomes considered, and there were very low rates of non-compliance, implying that increased patient burden due to trial re-enrolment was not an issue in this setting. This may be because most outcomes were recorded by the trial team, rather than by the patients themselves.

I did not find any difference in treatment effect estimates between re-randomisation and parallel group trials, though confidence intervals for the estimated differences were wide.

Figure 6.1: Difference in effect sizes between re-randomisation and parallel group trials. The blue lines represent the estimated treatment effect and 95% CI from the re-randomisation and parallel group trials. The red lines represent the difference in the treatment effect estimates between re-randomisation and parallel group trials (and a 95% CI for this difference). If the red line is close to 0, it means that re-randomisation and parallel group trials are providing similar estimates of treatment effect; if it is far away from 0, then re-randomisation and parallel group trials are giving different estimates of treatment effect. The x-axis shows the size of the effect for both the blue and red lines, however the x-axis text (whether re-randomisation or parallel group trials show more beneficial effects) applies only to the red line. RR=re-randomisation, PG=parallel group, CI=confidence interval, Std. Mean Diff.=standardized mean difference. Reproduced from *Kahan BC, Morris TP, Harris E, Pearse R, Hooper R, Eldridge E. Re-randomization increased recruitment and provided similar treatment estimates as parallel designs in trials of febrile neutropenia. Journal of Clinical Epidemiology 97 (2018) 14-19. DOI: https://doi.org/10.1016/j.jclinepi.2018.02.002* with permission.



There were some limitations to this review. There was only a small number of trials available. This led to wide confidence intervals for the differences in treatment effect estimates, meaning that I could not rule out differences between designs. I also focused only on one clinical area, and so these results may not be generalisable to other settings. Furthermore, reporting of key trial characteristics in the re-randomisation trials was often inadequate, which may in part reflect a lack of guidance on good reporting practice at the time these trials were conducted.

Overall, these results suggest that in the setting of febrile neutropenia, use of the re-randomisation design can increase recruitment compared to parallel group designs while providing similar results, without increasing rates of missing data or non-compliance.

7 Discussion

In multi-episode settings some patients may require treatment on more than one occasion. For instance, patients who experience acute sickle cell pain crises will require medication to relieve symptoms each time they present to hospital with a new pain crisis. Randomised trials in multi-episode settings are most often conducted using a parallel group design, where patients are allowed to be enrolled for a single episode only, and are not eligible to be re-enrolled for subsequent treatment episodes. This approach can be inefficient as it discards a portion of the available treatment episodes. Furthermore, results may not be generalisable if the treatment effect in the eligible episodes differs to the effect in the excluded episodes.

The re-randomisation design has been proposed as an alternative approach in multi-episode settings, as it allows patients to be re-enrolled and re-randomised for each new treatment episode they experience. Potential benefits of the re-randomisation design include increased recruitment and efficiency through inclusion of additional episodes, and increased generalisability. However, to date there has been very little methodological work done on re-randomisation trials, and so it is not entirely clear whether these trials can generally provide valid results. As such, there has been relatively slow uptake of these designs in practice.

In this thesis I evaluated the methodological considerations around re-randomisation trials in order to address this issue. In particular, I looked at what treatment effects re-randomisation trials actually estimate, how to analyse these designs to ensure they are unbiased, and in what settings they are preferable to other design options. I summarise the main results of this thesis in section 7.1 below. I then discuss future work related to this topic in section 7.2, and make concluding remarks in section 7.3.

7.1 Summary of thesis

7.1.1 Chapter 2: estimands in multi-episode settings

In chapter 2 I defined a set of estimands that may be of interest in multi-episode settings. I defined four main estimands of interest: (i) per-episode added-benefit; (ii) per-episode policy-benefit; (i) per-patient added-benefit; and (i) per-patient policy-benefit. The per-episode estimands represent an average treatment effect across episodes, while the per-patient estimands represent an average treatment effect across patients. The policy-benefit estimands represent the treatment effect in an episode given a policy

of intervention for all episodes vs. control for all episodes, whereas the added-benefit estimands represent the additional benefit of the intervention in an episode based on the treatment history in previous episodes. These estimands can be helpful in deciding which research question is of most interest, and choosing a study design and analysis method which is best able to answer this question.

7.1.2 Chapter 3: Independence estimators for re-randomisation trials

In chapter 3 I evaluated the use of independence estimators to analyse re-randomisation trials. Independence estimators use a working independence correlation structure to estimate treatment effects. I found that independence estimators are unbiased for the per-episode added-benefit estimand. They are also unbiased for the other estimands in many settings, though can be biased in certain scenarios. They can be biased for the policy-benefit estimands if the causal model for treatment history is misspecified. They can also be biased for the per-patient and policy-benefit estimands under certain non-enrolment scenarios (e.g. when there is differential non-enrolment across treatment groups).

7.1.3 Chapter 4: Mixed-effects models for re-randomisation trials

In chapter 4 I evaluated the use of mixed-effects models to analyse re-randomisation trials. Although mixed-effects models can offer much better precision compared to independence estimators, they are biased for most scenarios in which the treatment effect is not constant across patients or episodes. They can also be biased under some non-enrolment scenarios even when the treatment effect is constant. Therefore, mixed-effects models should not be used in re-randomisation trials. Instead, independence estimators should be the default method of analysis.

7.1.4 Chapter 5: Comparing re-randomisation trials with other design

In chapter 5 I compared the re-randomisation design with cluster, crossover, and parallel group designs in multi-episode settings. I found that although crossover trials are the most efficient design, they are biased in many settings and I recommend they not be used. Parallel group designs are useful if the aim is to estimate the effect of an intervention the first time it is used. However, they are biased for other estimands in many settings. They are also less precise than re-randomisation designs in many settings. Cluster designs can be useful to estimate policy-benefit effects, and do not rely on specifying a causal model for treatment history in order to be unbiased. However, they do not maintain allocation concealment, and can be biased in certain non-enrolment scenarios. Re-randomisation trials are also biased for the policy-benefit effect in the same non-enrolment scenarios as the cluster design, however the re-randomisation design can at least provide an unbiased estimate of at least one treatment effect (per-episode added-benefit) in these scenarios, which is not the case for the cluster design.

7.1.5 Chapter 6: Re-randomisation in febrile neutropenia trials

In chapter 6 I used a Cochrane review of granulocyte colony-stimulating factors for patients with febrile neutropenia which include both parallel group and re-randomisation trials to assess what impact re-randomisation had on recruitment, whether it led to higher rates of non-compliance or missing data, and whether treatment estimates differed between the two designs. I found that the five re-randomisation trials recruited between 16-66% more episodes than they would have had they used a parallel group design. There was no missing data in the re-randomisation trials for any of the outcomes considered, and there were very low rates of non-compliance. I did not find any difference in treatment effect estimates between re-randomisation and parallel group trials, though confidence intervals for these difference were wide.

These data suggest that in this setting, use of the re-randomisation design can increase recruitment compared to parallel group designs while providing similar results, without increasing rates of missing data or non-compliance.

7.2 Future work

Given the scarcity of methodological research about re-randomisation trials, there is a large scope for future work. I restricted the work in my thesis to the simplest setting where patients experience a maximum of two episodes, and treatment allocation does not affect the occurrence of subsequent episodes. It would be useful to extend this work to more complex settings.

For instance, with a larger number of episodes it will be more difficult to specify a causal model for treatment history for the policy-benefit estimands. As a motivating example, one of the trials evaluating granulocyte colony-stimulating factor in febrile neutropenia from the review in chapter 6 had some patients who enrolled for up to four episodes. The reason that specifying a causal model for treatment history is more difficult with four episodes rather than two is because the number of potential variables to include in the model will increase quite quickly with each additional episode. With two episodes we can specify the effect of treatment history using two variables ($Z_{i,j-1}$ and $Z_{ij}Z_{i,j-1}$), which allow the outcome and treatment effect to depend on treatment allocation in the previous episode. However, with three episodes we could specify the effect of treatment history using up to six variables ($Z_{i,j-1}$, $Z_{i,j-2}$, $Z_{i,j-1}Z_{i,j-2}$, $Z_{ij}Z_{i,j-1}$, $Z_{ij}Z_{i,j-2}$, $Z_{ij}Z_{i,j-1}Z_{i,j-2}$), which allows both the outcome and treatment effect to depend on the treatment allocations in the previous two episodes, and for these associations to also depend on whether the intervention was received in both previous episodes or only one of the two episodes. As the number of episodes increases there will very quickly be too many possible variables to reasonably include in the model, and we will need to decide which variables to omit, or whether we can make simplifying assumptions in order to reduce the number of variables. For example, we could assume that only the treatment allocation in the previous episode matters, or that the effect

of treatment history can be accurately captured from a single variable representing the number of previous episodes in which the patient received the intervention. Further research on how best to specify these causal models for treatment history when there are a large number of episodes would be useful.

In this thesis I assumed the treatment allocation had no effect on the occurrence of subsequent episodes. Although this is a reasonable assumption in some settings, there are other settings where it will be false. For example, some trials may have terminating endpoints (i.e. where a treatment success or failure precludes future episodes) such as mortality, or pregnancy in a fertility study. Several of the trials discussed in chapter 1 had terminating endpoints. For example, the trials of granulocyte colony-stimulating factor in febrile neutropenia and albumin for complications from cirrhosis both included mortality as an outcome. It would therefore be useful to extend the estimands defined in chapter 2 to this setting, and evaluate whether the re-randomisation design is able to provide unbiased estimates. This setting is complicated by the fact that some treatment histories are impossible, i.e. that certain episodes may not exist under an alternate treatment history. This presents a challenge for the policy-benefit estimand; if an episode would only exist under one of the two treatment histories of interest (all intervention vs. all control) it is unclear how to define the treatment effect at this episode. One solution might be to use a principal stratification approach [35, 62, 63, 64], where the policy-benefit effect is defined within the subset of episodes that would exist under both treatment histories. However, it is not clear whether this effect is clinically useful, or whether it would be possible to estimate it from re-randomisation or other designs without using unrealistic assumptions.

Conversely, the per-episode added-benefit estimand should be directly applicable to the setting where treatment allocation affects occurrence of future episodes. Because it averages over treatment histories according to their probability of being observed, it is already defined in a way that excludes treatment histories under which an episode would not exist (as these treatment histories have 0 probability of being observed). Until the other estimands are extended to this setting, re-randomisation trials in which treatment allocation may affect occurrence of subsequent episodes (i.e. trials with terminating endpoints) should use the per-episode added-benefit effect to ensure their estimand is defined in a valid way.

In this thesis I primarily focussed on bias, and did not explicitly look at the best way to estimate standard errors in re-randomisation trials. For the independence estimators in the simulation studies in chapters 3-5 I used a robust standard error which allowed for clustering of episodes within patients. However, there is some evidence to suggest that model-based standard errors from independence estimators (which completely ignore clustering) are valid in many settings [9]. It would be useful to directly compare these two approaches to determine whether one is preferable.

Currently there are no formula for sample size calculations available for re-

randomisation trials. It has previously been shown that under certain assumptions the sample size calculation for a parallel group trial could also be used for a re-randomisation design (where the re-randomisation design would recruit the specified number of episodes). However, this approach is only valid for the per-episode added-benefit effect, and assumes a constant treatment effect mechanism. It would be useful to develop a more general approach to sample size calculations that works under any treatment effect mechanism, and to develop sample size formulas for the per-patient and policy-benefit estimands.

7.3 Concluding remarks

The re-randomisation design can be a useful design option in multi-episode settings. It can increase the recruitment rate, allowing trials to finish earlier, and facilitate estimation of different treatment effects, allowing researchers to answer multiple questions about the treatments. It can provide unbiased treatment effect estimates, though for some estimands the effects are only unbiased under certain assumptions; careful consideration of these assumptions should be undertaken to evaluate their plausibility. The re-randomisation design compares favourably with other design options that could be used in multi-episode settings, though depending on the specific research question other designs may be more appropriate in certain instances. In conclusion, re-randomisation can be a safe and useful option, and should be used more often.

Bibliography

[1] C. R. Morris, F. A. Kuypers, L. Lavrisha, M. Ansari, N. Sweeters, M. Stewart, G. Gildengorin, L. Neumayr, and E. P. Vichinsky, "A randomized, placebo-controlled trial of arginine therapy for the treatment of children with sickle cell disease hospitalized with vaso-occlusive pain episodes," *Haematologica*, vol. 98, pp. 1375–1382, Sept. 2013.

[2] J. Stokholm, B. L. Chawes, N. H. Vissing, E. Bjarnadóttir, T. M. Pedersen, R. K. Vinding, A.-M. M. Schoos, H. M. Wolsk, S. Thorsteinsdóttir, H. W. Hallas, L. Arianto, S. Schjørring, K. A. Krogfelt, T. K. Fischer, C. B. Pipper, K. Bønnelykke, and H. Bisgaard, "Azithromycin for episodes with asthma-like symptoms in young children aged 1-3 years: a randomised, double-blind, placebo-controlled trial," *The Lancet Respiratory Medicine*, vol. 4, pp. 19–26, Jan. 2016.

[3] P. L. Mitchell, B. Morland, M. C. Stevens, G. Dick, D. Easlea, L. C. Meyer, and C. R. Pinkerton, "Granulocyte colony-stimulating factor in established febrile neutropenia: a randomized study of pediatric patients," *Journal of Clinical Oncology*, vol. 15, pp. 1163–1170, Mar. 1997.

[4] G. Cho, K. A. Anie, J. Buckton, P. Kiilu, M. Layton, L. Alexander, C. Hemmaway, D. Sutton, C. Amos, C. J. Doré, B. Kahan, and S. Meredith, "SWIM (sickle with ibuprofen and morphine) randomised controlled trial fails to recruit: lessons learnt," *BMJ Open*, vol. 6, p. e011276, June 2016.

[5] R. Harbour and J. Miller, "A new system for grading recommendations in evidence based guidelines," *BMJ*, vol. 323, pp. 334–336, Aug. 2001.

[6] A. M. McDonald, R. C. Knight, M. K. Campbell, V. A. Entwistle, A. M. Grant, J. A. Cook, D. R. Elbourne, D. Francis, J. Garcia, I. Roberts, and C. Snowdon, "What influences recruitment to randomised controlled trials? A review of trials funded by two UK funding agencies," *Trials*, vol. 7, p. 9, Apr. 2006.

[7] B. G. O. Sully, S. A. Julious, and J. Nicholl, "A reinvestigation of recruitment to randomised, controlled, multicenter trials: a review of trials funded by two UK funding agencies," *Trials*, vol. 14, p. 166, June 2013.

[8] B. Carlisle, J. Kimmelman, T. Ramsay, and N. MacKinnon, "Unsuccessful trial accrual and human subjects protections: an empirical analysis of recently closed trials," *Clinical Trials*, vol. 12, pp. 77–83, Feb. 2015.

[9] B. C. Kahan, A. B. Forbes, C. J. Doré, and T. P. Morris, "A re-randomisation design for clinical trials," *BMC Medical Research Methodology*, vol. 15, p. 96, Nov. 2015.

[10] C. Haywood, S. Lanzkron, M. Diener-West, J. Haythornthwaite, J. J. Strouse, S. Bediako, G. Onojobi, M. C. Beach, and I. Investigators, "Attitudes toward clinical trials among patients with sickle cell disease," *Clinical Trials*, vol. 11, pp. 275–283, June 2014.

[11] C. Tudur Smith, H. Hickey, M. Clarke, J. Blazeby, and P. Williamson, "The trials methodological research agenda: results from a priority setting exercise," *Trials*, vol. 15, p. 32, Jan. 2014.

[12] B. C. Kahan, "Using re-randomization to increase the recruitment rate in clinical trials - an assessment of three clinical areas," *Trials*, vol. 17, p. 595, Dec. 2016.

[13] A. J. Dunning and J. Reeves, "Control of type 1 error in a hybrid complete two-period vaccine efficacy trial," *Pharmaceutical Statistics*, vol. 13, pp. 397–402, 2014.

[14] M. Nason and D. Follmann, "Design and analysis of crossover trials for absorbing binary endpoints," *Biometrics*, vol. 66, pp. 958–965, Sept. 2010.

[15] P. F. van der Meer, P. F. Ypma, N. van Geloven, J. A. van Hilten, R. J. van Wordragen-Vlaswinkel, O. Eissen, J. J. Zwaginga, M. Trus, E. A. M. Beckers, P. Te Boekhorst, A. Tinmouth, Y. Lin, C. Hsia, D. Lee, P. J. Norris, R. P. Goodrich, A. Brand, T. Hervig, N. M. Heddle, J. G. van der Bom, and J.-L. H. Kerkhoffs, "Hemostatic efficacy of pathogen-inactivated vs untreated platelets: a randomized controlled trial," *Blood*, vol. 132, pp. 223–231, July 2018.

[16] C. A. DiazGranados, A. J. Dunning, M. Kimmel, D. Kirby, J. Treanor, A. Collins, R. Pollak, J. Christoff, J. Earl, V. Landolfi, E. Martin, S. Gurunathan, R. Nathan, D. P. Greenberg, N. G. Tornieporth, M. D. Decker, and H. K. Talbot, "Efficacy of high-dose versus standard-dose influenza vaccine in older adults," *The New England Journal of Medicine*, vol. 371, pp. 635–645, Aug. 2014.

[17] H. M. Said, S. Gupta, L. K. Vricella, K. Wand, T. Nguyen, and G. Gross, "Effect of ambient light on the time needed to complete a fetal biophysical profile: A randomized controlled trial," *European Journal of Obstetrics, Gynecology, and Reproductive Biology*, vol. 217, pp. 59–65, Oct. 2017.

[18] L. China, S. S. Skene, K. Bennett, Z. Shabir, R. Hamilton, S. Bevan, T. Chandler, A. A. Maini, N. Becares, D. Gilroy, E. H. Forrest, and A. O'Brien, "ATTIRE: Albumin to prevent infection in chronic liver failure: study protocol for an interventional randomised controlled trial," *BMJ Open*, vol. 8, p. e023754, Oct. 2018.

[19] D. B. Rubin, "Estimating causal effects of treatments in randomized and nonrandomized studies," *Journal of Educational Psychology*, vol. 66, no. 5, pp. 688–701, 1974.

[20] M. A. Hernán and J. M. Robins, *Causal Inference*. Boca Raton: Chapman & Hall/CRC, 2020.

[21] K.-Y. Liang and S. L. Zeger, "Longitudinal data analysis using generalized linear models," *Biometrika*, vol. 73, no. 1, pp. 13–22, 1986.

[22] M. S. Pepe and G. L. Anderson, "A cautionary note on inference for marginal regression models with longitudinal data and general correlated response data," *Communications in Statistics - Simulation and Computation*, vol. 23, pp. 939–951, Jan. 1994.

[23] M. Pavlou, S. R. Seaman, and A. J. Copas, "An examination of a method for marginal inference when the cluster size is informative," *Statistica Sinica*, 2013.

[24] L. N. Yelland, T. R. Sullivan, M. Pavlou, and S. R. Seaman, "Analysis of randomised trials including multiple births when birth size is informative," *Paediatric and Perinatal Epidemiology*, vol. 29, pp. 567–575, Nov. 2015.

[25] J. M. Williamson, S. Datta, and G. A. Satten, "Marginal analyses of clustered data when cluster size is informative," *Biometrics*, vol. 59, pp. 36–42, Mar. 2003.

[26] J. M. Wooldridge, *Econometric Analysis of Cross Section and Panel Data*. The MIT Press, 2010.

[27] T. P. Morris, I. R. White, and M. J. Crowther, "Using simulation studies to evaluate statistical methods," *Statistics in Medicine*, vol. 38, pp. 2074–2102, May 2019.

[28] M. Takada, T. Sozu, and T. Sato, "Practical approaches for design and analysis of clinical trials of infertility treatments: crossover designs and the mantel-haenszel method are recommended," *Pharmaceutical Statistics*, vol. 14, pp. 198–204, 2015.

[29] M. H. Gail, S. Wieand, and S. Piantadosi, "Biased estimates of treatment effect in randomized experiments with nonlinear regressions and omitted covariates," *Biometrika*, vol. 71, no. 3, pp. 431–444, 1984.

[30] W. W. Hauck, S. Anderson, and S. M. Marcus, "Should we adjust for covariates in nonlinear regression analyses of randomized trials?," *Controlled Clinical Trials*, vol. 19, no. 3, pp. 249–256, 1998.

[31] L. D. Robinson and N. P. Jewell, "Some surprising results about covariate adjustment in logistic regression models," *International Statistical Review*, pp. 227–240, 1991.

[32] C. A. DiazGranados, A. J. Dunning, C. A. Robertson, H. K. Talbot, V. Landolfi, and D. P. Greenberg, "Effect of previous-year vaccination on the efficacy, immunogenicity, and safety of high-dose inactivated influenza vaccine in older adults," *Clinical Infectious Diseases*, vol. 62, pp. 1092–1099, Feb. 2016.

[33] S. H. Choi, A. H. Waltje, D. L. Ronis, D. Noonan, O. Hong, C. R. Richardson, J. D. Meeker, and S. A. Duffy, "Web-enhanced tobacco tactics with telephone support versus 1-800-QUIT-NOW telephone line intervention for operating engineers: Randomized controlled trial," *Journal of Medical Internet Research*, vol. 16, p. e255, Nov. 2014.

[34] E. Park, S. H. Choi, and S. A. Duffy, "The effect of re-randomization in a smoking cessation trial," *American Journal of Health Behavior*, vol. 40, pp. 667–674, Sept. 2016.

[35] International Conference on Harmonization, "ICH E9 (R1): addendum to statistical principles for clinical trials on choosing appropriate estimands and defining sensitivity analyses in clinical trials," 30 August 2017.

[36] M. Akacha, F. Bretz, D. Ohlssen, G. Rosenkranz, and H. Schmidli, "Estimands and their role in clinical trials," *Statistics in Biopharmaceutical Research*, vol. 9, pp. 268–271, Mar. 2017.

[37] M. Akacha, F. Bretz, and S. Ruberg, "Estimands in clinical trials - broadening the perspective," *Statistics in Medicine*, vol. 36, pp. 5–19, July 2016.

[38] A.-K. Leuchs, J. Zinserling, A. Brandt, D. Wirtz, and N. Benda, "Choosing appropriate estimands in clinical trials," *Therapeutic Innovation & Regulatory Science*, vol. 49, pp. 584–592, Feb. 2015.

[39] B. C. Kahan, T. P. Morris, E. Harris, R. Pearse, R. Hooper, and S. Eldridge, "Re-randomization increased recruitment and provided similar treatment estimates as parallel designs in trials of febrile neutropenia," *Journal of Clinical Epidemiology*, vol. 97, pp. 14–19, May 2018.

[40] E. Benhin, J. N. K. Rao, and A. J. Scott, "Mean estimating equation approach to analysing cluster-correlated data with nonignorable cluster sizes," *Biometrika*, vol. 92, pp. 435–450, June 2005.

[41] E. B. Hoffman, P. K. Sen, and C. R. Weinberg, "Within-cluster resampling," *Biometrika*, vol. 88, pp. 1121–1134, Dec. 2001.

[42] Y. Huang and B. Leroux, "Informative cluster sizes for subcluster-level covariates and weighted generalized estimating equations," *Biometrics*, vol. 67, pp. 843–851, Jan. 2011.

[43] J. M. Neuhaus and C. E. McCulloch, "Estimation of covariate effects in generalized linear mixed models with informative cluster sizes," *Biometrika*, vol. 98, pp. 147–162, Jan. 2011.

[44] S. Seaman, M. Pavlou, and A. Copas, "Review of methods for handling confounding by cluster and informative cluster size in clustered data," *Statistics in Medicine*, vol. 33, pp. 5371–5387, Dec. 2014.

[45] S. R. Seaman, M. Pavlou, and A. J. Copas, "Methods for observed-cluster inference when cluster size is informative: a review and clarifications," *Biometrics*, vol. 70, pp. 449–456, June 2014.

[46] B. C. Kahan, "Bias in randomised factorial trials," *Statistics in Medicine*, vol. 32, pp. 4540–4549, Nov. 2013.

[47] B. C. Kahan, "Accounting for centre-effects in multicentre trials with a binary outcome - when, why, and how?," *BMC Medical Research Methodology*, vol. 14, p. 20, Feb. 2014.

[48] B. C. Kahan and T. P. Morris, "Improper analysis of trials randomised using stratified blocks or minimisation," *Statistics in Medicine*, vol. 31, pp. 328–340, Dec. 2011.

[49] R. H. Keogh and T. P. Morris, "Multiple imputation in Cox regression when there are time-varying effects of covariates," *Statistics in medicine*, vol. 37, pp. 3661–3678, Nov. 2018.

[50] S. J. Walters, R. M. Jacques, I. B. dos Anjos Henriques-Cadby, J. Candlish, N. Totton, and M. T. S. Xian, "Sample size estimation for randomised controlled trials with repeated assessment of patient-reported outcomes: what correlation between baseline and follow-up outcomes should we assume?," *Trials*, vol. 20, no. 1, pp. 1–16, 2019.

[51] S. Rabe-Hesketh and A. Skrondal, *Multilevel and Longitudinal Modeling Using Stata, 3rd Edition*. Stata Press books, StataCorp LP, April 2012.

[52] D. A. Harville, "Maximum likelihood approaches to variance component estimation and to related problems," *Journal of the American Statistical Association*, vol. 72, pp. 320–338, June 1977.

[53] H. D. Patterson and R. Thompson, "Recovery of inter-block information when block sizes are unequal," *Biometrika*, vol. 58, no. 3, pp. 545–554, 1971.

[54] S. Puffer, D. Torgerson, and J. Watson, "Evidence for risk of bias in cluster randomised trials: review of recent trials published in three general medical journals," *BMJ*, vol. 327, pp. 785–789, Oct. 2003.

[55] M. K. Campbell, D. R. Elbourne, and D. G. Altman, "CONSORT statement: extension to cluster randomised trials," *BMJ*, vol. 328, pp. 702–708, Mar. 2004.

[56] S. Eldridge, S. Kerry, and D. J. Torgerson, "Bias in identifying and recruiting participants in cluster randomised trials: what can be done?," *BMJ*, vol. 339, pp. b4006–b4006, Oct. 2009.

[57] D. J. Torgerson, "Contamination in trials: is cluster randomisation the answer?," *BMJ*, vol. 322, pp. 355–357, Feb. 2001.

[58] K. Dwan, T. Li, D. G. Altman, and D. Elbourne, "CONSORT 2010 statement: extension to randomised crossover trials," *BMJ*, p. l4378, July 2019.

[59] B. Jones and M. G. Kenward, *Design and Analysis of Cross-Over Trials*. Taylor & Francis Ltd, 2014.

[60] E. J. Mills, A.-W. Chan, P. Wu, A. Vail, G. H. Guyatt, and D. G. Altman, "Design, analysis, and presentation of crossover trials," *Trials*, vol. 10, Apr. 2009.

[61] R. Mhaskar, O. A. C. Clark, G. Lyman, T. E. A. Botrel, L. M. Paladini, and B. Djulbegovic, "Colony-stimulating factors for chemotherapy-induced febrile neutropenia," *Cochrane Database of Systematic Reviews*, Oct. 2014.

[62] J. L. Zhang and D. B. Rubin, "Estimation of causal effects via principal stratification when some outcomes are truncated by "death"," *Journal of Educational and Behavioral Statistics*, vol. 28, pp. 353–368, Dec. 2003.

[63] C. E. Frangakis and D. B. Rubin, "Principal stratification in causal inference," *Biometrics*, vol. 58, pp. 21–29, Mar. 2002.

[64] Y. Chiba and T. J. VanderWeele, "A simple method for principal strata effects when the outcome has been truncated due to death," *American Journal of Epidemiology*, vol. 173, pp. 745–751, Feb. 2011.

A Calculating expected values in tables 3.9 and 3.11

A.1 Calculating the expected values of outcomes at each episode in table 3.9

The expected values of the outcomes in table 3.9 (chapter 3, section 3.3.10) are obtained by a weighted average. I show this below.

For $E(Y_{ij})$ of sequence $Z=(0)$, $j=1$:

$$E(Y_{ij}) = \frac{\beta_{X_{PL}}(1-p_{01}) + 0(1-p_{00})}{(1-p_{01})+(1-p_{00})} = \beta_{X_{PL}}\frac{(1-p_{01})}{2(1-p)}$$

For $E(Y_{ij})$ of sequence $Z=(0,1)$, $j=1$, and sequence $Z=(0,0)$, $j=1,\ 2$:

$$E(Y_{ij}) = \frac{\beta_{X_{PL}}p_{01} + 0p_{00}}{p_{01}+p_{00}} = \beta_{X_{PL}}\frac{p_{01}}{2p}$$

For $E(Y_{ij})$ of sequence $Z=(0,1)$, $j=2$:

$$E(Y_{ij}) = \frac{(\beta_{trt}+\beta_{X_{PL}})p_{01} + \beta_{trt}p_{00}}{p_{01}+p_{00}} = \beta_{trt} + \beta_{X_{PL}}\frac{p_{01}}{2p}$$

For $E(Y_{ij})$ of sequence $Z=(1)$, $j=1$:

$$E(Y_{ij}) = \frac{(\beta_{trt}+\beta_{X_{PL}})(1-p_{00}) + \beta_{trt}(1-p_{01})}{(1-p_{00})+(1-p_{01})} = \beta_{trt} + \beta_{X_{PL}}\frac{(1-p_{00})}{2(1-p)}$$

For $E(Y_{ij})$ of sequence $Z=(1,0)$, $j=1$, and sequence $Z=(1,1)$, $j=1,\ 2$:

$$E(Y_{ij}) = \frac{(\beta_{trt}+\beta_{X_{PL}})p_{00} + \beta_{trt}p_{01}}{p_{00}+p_{01}} = \beta_{trt} + \beta_{X_{PL}}\frac{p_{00}}{2p}$$

And for $E(Y_{ij})$ of sequence $Z=(1,0)$, $j=2$:

$$E(Y_{ij}) = \frac{\beta_{X_{PL}}p_{00} + 0p_{01}}{p_{00}+p_{01}} = \beta_{X_{PL}}\frac{p_{00}}{2p}$$

A.2 Calculating the expected values of outcomes at each episode in table 3.11

The expected values of the outcomes in table 3.11 (chapter 3, section 3.3.11) are obtained by a weighted average. I show this below.

For $E(Y_{ij})$ of sequence $Z=(0)$, $j=1$:

$$E(Y_{ij}) = \frac{0(1-p_{00}) + 0(1-p_{01}) + \beta_{X_{EL}}(1-p_{00}) + \beta_{X_{EL}}(1-p_{01})}{2(1-p_{00}) + 2(1-p_{01})} = \frac{\beta_{X_{EL}}}{2}$$

(note that $\frac{N_T}{8}$ cancels out in the numerator and denominator)

For $E(Y_{ij})$ for $j=1$ for sequence $Z=(0,1)$ and $Z=(0,0)$:

$$E(Y_{ij}) = \frac{0(p_{00}) + 0(p_{01}) + \beta_{X_{EL}}p_{00} + \beta_{X_{EL}}p_{01}}{2p_{00} + 2p_{01}} = \frac{\beta_{X_{EL}}}{2}$$

For $E(Y_{ij})$ for $j=2$ of sequence $Z=(0,1)$:

$$E(Y_{ij}) = \frac{p_{00}\beta_{trt} + p_{01}(\beta_{trt} + \beta_{X_{EL}}) + p_{00}\beta_{trt} + p_{01}(\beta_{trt} + \beta_{X_{EL}})}{2p_{00} + 2p_{01}}$$

$$= \frac{2\beta_{trt}(p_{00} + p_{01}) + 2\beta_{X_{EL}}p_{01}}{2(p_{00} + p_{01})} = \beta_{trt} + \beta_{X_{EL}}\frac{p_{01}}{p_{00} + p_{01}}$$

For $E(Y_{ij})$ for $j=2$ of sequence $Z=(0,0)$:

$$E(Y_{ij}) = \frac{0(p_{00}) + \beta_{X_{EL}}p_{01} + 0(p_{00}) + \beta_{X_{EL}}p_{01}}{2p_{00} + 2p_{01}} = \beta_{X_{EL}}\frac{p_{01}}{p_{00} + p_{01}}$$

For $E(Y_{ij})$ of sequence $Z=(1)$, $j=1$:

$$E(Y_{ij}) = \frac{\beta_{trt}(1-p_{01}) + \beta_{trt}(1-p_{00}) + (\beta_{trt} + \beta_{X_{EL}})(1-p_{01}) + (\beta_{trt} + \beta_{X_{EL}})(1-p_{00})}{2(1-p_{01}) + 2(1-p_{00})}$$

$$= \frac{2\beta_{trt}(2-p_{01}-p_{00}) + \beta_{X_{EL}}(2-p_{01}-p_{00})}{2(2-p_{01}-p_{00})} = \beta_{trt} + \frac{\beta_{X_{EL}}}{2}$$

For $E(Y_{ij})$ for $j=1$ of sequences $Z=(1,0)$ and $Z=(1,1)$:

$$E(Y_{ij}) = \frac{p_{01}\beta_{trt} + p_{00}\beta_{trt} + p_{01}(\beta_{trt} + \beta_{X_{EL}}) + p_{00}(\beta_{trt} + \beta_{X_{EL}})}{2p_{00} + 2p_{01}} = \beta_{trt} + \frac{\beta_{X_{EL}}}{2}$$

And for $E(Y_{ij})$ of sequence $Z = (1, 0)$, $j = 2$:

$$E(Y_{ij}) = \frac{0(p_{01}) + \beta_{X_{EL}} p_{00} + 0(p_{01}) + \beta_{X_{EL}} p_{00}}{2p_{00} + 2p_{01}} = \beta_{X_{EL}} \frac{p_{00}}{p_{00} + p_{01}}$$

And for $E(Y_{ij})$ of sequence $Z = (1, 1)$, $j = 2$:

$$E(Y_{ij}) = \frac{p_{01}\beta_{trt} + p_{00}(\beta_{trt} + \beta_{X_{EL}}) + p_{01}\beta_{trt} + p_{00}(\beta_{trt} + \beta_{X_{EL}})}{2p_{00} + 2p_{01}} = \beta_{trt} + \beta_{X_{EL}} \frac{p_{00}}{p_{00} + p_{01}}$$