The Use of Health Economics in the Design and Analysis of Adaptive Clinical Trials

Author: Laura Flight

Supervisors: Professor Steven Julious, Professor Alan Brennan,

Professor Susan Todd

A thesis submitted in partial fulfilment of the requirements for the degree of doctor of philosophy

University of Sheffield

Faculty of Medicine, Dentistry & Health

School of Health and Related Research

April 2020

To my wonderful grandparents Ron & Dot Flight, and Vincent & Rita Clarke

Acknowledgements

Firstly, I would like to thank my supervisors. Steven – thank you for your unfaltering guidance, encouragement and patience, which has helped me develop from internship to PhD. Alan – thank you for reminding me of the gift of a PhD when I was battling literature reviews and broken code. Sue – thank you for making the intimidating theory understandable and instilling a fresh outlook in me with each visit.

I am indebted to the public advisors that worked with me over the last four years. You have been kind, patient and honest friends, and ultimately made me a better researcher.

I have been supported by numerous researchers within ScHARR and beyond. Specifically, I would like to thank: Daniel Hind for his advice as I ventured into the world of qualitative research; Martin Forster for a welcoming and interesting research visit and subsequently helping to bring together the ENACT collaboration; Nicky Welton and Howard Thom for providing new perspectives that helped me tackle Chapter 8; Matthew Bradley for helping to recruit participants to the qualitative study; Rebecca Palmer, Nicholas Latimer, Tim Chater and Liz Cross for their support with the CACTUS case study; Tim Skalland for sharing R code for adjusted secondary confidence intervals; and Tracey Young for helping me juggle the PhD and work over the last 18 months.

The Medical Statistics Group have been my support network throughout this whole process. I want to thank you all for the coffee breaks, burritos and chats that made coming to work enjoyable even when it felt like R was out to get me. A special thank you to Beckie Simpson and Jane Candlish for the conference adventures and proofreading.

I would like to thank my family who have always celebrated my successes and, more importantly, supported me through challenges. Mum and Dad, thank you for everything.

To my friends Alice, Henry and Lucy, thank you for bringing joy to my life away from work, and for the fantastic holidays!

Finally, I want to thank my partner Rami for keeping me laughing throughout and providing Diet Coke and Haribo when things got desperate. You're the best.

Declaration

I declare that this thesis is my own work. This work has not been previously been presented for an award at this, or any other, university.

Laura Flight is funded by a National Institute for Health Research (NIHR) Doctoral Research Fellowship for this project. This article presents independent research funded by the NIHR. The views expressed are those of the authors and not necessarily those of the NHS, the NIHR, the Department of Health and Social Care or the University of Sheffield.

The CACTUS Pilot trial was funded by the NIHR under its Research for Patient Benefit (RfPB) Programme (Grant Reference Number PB-PG-1207-14097). This study was also supported by the Stroke and Telehealth themes of the South Yorkshire Collaboration for Leadership in applied health research and care (CLAHRC). NIHR CLAHRC for South Yorkshire acknowledges funding from the NIHR. The study also received support from North of Tyne PCT. The views expressed are those of the authors and not necessarily those of the National Health Service, the NIHR, or the Department of Health.

The Big CACTUS project was funded by the NIHR Health Technology Assessment Programme (12/21/01). Additional funding support was provided by the Tavistock Trust for Aphasia. The views and opinions expressed are those of the authors and do not necessarily reflect those of the Health Technology Assessment Programme, the NIHR, the National Health Service, the Department of Health and Social Care, the Tavistock Trust for Aphasia, or the Stroke Association.

Publications

Publication of work in Chapter 3 **Flight L**, Arshad F, Barnsley R, Patel K, Julious S, Brennan A, Todd S (2019) A Review of Clinical Trials With an Adaptive Design and Health Economic Analysis. Value in Health, 22(4), 391-398. https://doi.org/10.1016/j.jval.2018.11.008

Publication of work in Chapter 4 **Flight L**, Julious S, Brennan A, Todd S, Hind D. How can health economics be used in the design and analysis of adaptive clinical trials? A qualitative analysis. Trials. 2020 Dec;21(1):1-2. https://doi.org/10.1186/s13063-020-4137-2

Relevant to PhD Research

Hatfield I, Allison A, **Flight L**, Julious SA, Dimairo M. Adaptive designs undertaken in clinical research: a review of registered clinical trials. Trials. 2016 Dec;17(1):150.

Pallmann P, Bedding AW, Choodari-Oskooei B, Dimairo M, **Flight L**, Hampson LV *et al* (2018) Adaptive designs in clinical trials: why use them, and how to run and report them. BMC Medicine, 16.

Outside PhD Research

Flight L, Allison A, Dimairo M, Lee E, Mandefield L, Walters SJ (2016) Recommendations for the analysis of individually randomised controlled trials with clustering in one arm – a case of continuous outcomes. BMC Medical Research Methodology, 16.

Flight L, Julious SA (2016) Practical guide to sample size calculations: non-inferiority and equivalence trials. Pharmaceutical Statistics, 15(1), 80-89.

Flight L, Julious SA (2016) Practical guide to sample size calculations: superiority trials. Pharmaceutical Statistics, 15(1), 75-79.

Flight L, Julious SA (2016) Practical guide to sample size calculations: an introduction. Pharmaceutical Statistics, 15(1), 68-74.

Walters SJ, Henriques-Cadby IBDA, Bortolami O, **Flight L**, Hind D *et al* (2017) Recruitment and retention of participants in randomised controlled trials: a review of trials funded and

published by the United Kingdom Health Technology Assessment Programme. BMJ Open, 7(3).

Candlish J, Teare MD, Dimairo M, **Flight L**, Mandefield L, Walters SJ (2018) Appropriate statistical methods for analysing partially nested randomised controlled trials with continuous outcomes: a simulation study.. BMC Medical Research Methodology, 18(1), 105-105.

Hykin P, Prevost AT, Sivaprasad S, Vasconcelos , Murphy C, Kelly J, Ramu1 J, Alshreef A, **Flight L**, *et al* (2020) A Multicentre Phase III Double-masked Randomised Controlled Non-Inferiority Trial comparing the clinical and cost effectiveness of intravitreal therapy with ranibizumab (Lucentis) vs aflibercept (Eylea) vs bevacizumab (Avastin) for Macular Oedema due to Central Retinal Vein Occlusion (LEAVO). Health Technology Assessment Journal. In Press February 2020.

Pennington P, Hernadez-Alva M, Hykin P, Sivaprasad S, **Flight L**, Alshreef A *et al* (2020) Mapping from visual acuity to EQ-5D, EQ-5D with vision bolt-on and VFQ-UI in patients with macular edema in the LEAVO trial. Value in Health. In Press April 2020.

Other Outputs

Press Release. Health economics outcomes are rarely considered in adaptive clinical trial designs. Available online at www.ispor.org/heor-resources/news/view/2019/04/30/report-health-economics-outcomes-are-rarely-considered-in-adaptive-clinical-trial-designs. Last Accessed 17.02.2020.

Cytel Blog Post. Adaptive Design and Health Economic Analysis: Interview with Laura Flight. Available online at www.cytel.com/blog/adaptive-design-health-economics. Last Accessed 17.02.2020.

Conferences Presentations

Invited Oral Presentations

Flight L, Julious S, Brennan A, Todd S (2016) Efficient Study Designs – Adaptive Designs and Economic Evaluation, Society for Clinical Trials, Montreal, Canada, 16 - 18 May 2016.

Flight L. (2017) Adaptive Study Designs and Health Economic Evaluations Society for Clinical Trials/International Clinical Trials Methodology Conference, Liverpool , UK, 7-10 May 2017 Organised, chaired and presented at this invited session.

Flight L, Julious S, Brennan A, Todd S (2018) Using Health Economics in the Design, Monitoring and Analysis of Adaptive Clinical Trials, Royal Statistical Society, Cardiff, UK, 4 - 6 September 2018.

Flight L. (2019) Public Involvement in Methodological Research. NIHR Statistics Group Opportunities for Statisticians to be a PI. Bristol, UK, 9 March 2019.

Contributed Oral Presentations

Flight L, Julious S, Brennan A, Todd S (2017) The Use of Adaptive Designs and Health Economic Analyses in Health Technology Assessments. ScHARR Postgraduate Research Conference, Sheffield, UK, 30 May 2017.

Flight L, Julious S, Brennan A, Todd S (2018) The Impact of an Adaptive Design on a Health Economic Analysis. Health Economic Study Group. Bristol, UK, 20 - 22 June 2018.

Flight L, Julious S, Brennan A, Todd S (2018) Using Health Economics in the Design, Monitoring and Analysis of Adaptive Clinical Trials. Research Students' Conference in Probability and Statistics, Sheffield, UK, 24 - 26 July 2018.

Flight L, Hind D, Julious S, Brennan A, Todd S (2019) Considerations concerning the use of health economics in the design and analysis of adaptive clinical trials - a qualitative study. International Clinical Trials Methodology Conference Brighton, UK, 6 - 9 October 2019.

Poster Presentations

Flight L, Julious S, Brennan A, Todd S (2016) How does looking at data early in a clinical trial impact value for money? NIHR Trainee Meeting, Leeds, UK, 6 - 7 December 2016.

Flight L, Julious S, Brennan A, Todd S (2016) How does looking at data early in a clinical trial impact value for money? NIHR Trainee Meeting, Leeds, UK, 24 May 2018.

Flight L, Julious S, Brennan A, Todd S (2019) Methods for Choosing the Best Design of a Clinical Trial: Implementing Expected Value of Sample Information and Adjusting for Bias Induced By a Group Sequential Design Society for Medical Decision Making, Portland, USA, 21 - 23 October 2019.

Outside PhD Research

Flight L, Baxter S, Khan S (2019) Public involvement beyond clinical research studies. International Clinical Trials Methodology Conference, Brighton, UK, 6 - 9 October 2019.

Flight L, Alshreef A, Hykin PG, Sivaprasad S, Pennington B, Kelly J *et al* (2019) Intravitreal Therapy With Ranibizumab Vs Aflibercept Vs Bevacizumab For Macular Oedema Due To Central Retinal Vein Occlusion: A Within Trial Cost-Utility Analysis. ISPOR Europe, Copenhagen, 4 - 6 November 2019.

Pennington B, Alshreef A, **Flight L**, Hykin P, Sivaprasad S, Brazier JE, Hernández M (2019) Mapping From Visual Acuity To Utility In Patients With Macular Oedema Due To Central Retinal Vein Occlusion Using Data From The Leavo Trial. ISPOR Europe, Copenhagen, 4 - 6 November 2019.

External Seminars

Flight L (2018) Using Health Economics in the Design, Monitoring and Analysis of Adaptive Clinical Trials, Multi Parameter Evidence Synthesis Group, University of Bristol, 29 June 2018.

Flight L (2019) Using Health Economics in the Design, Monitoring and Analysis of Adaptive Clinical Trials, Department of Health Sciences, University of York, 22 January 2019.

Flight L (2020) The Use of Health Economics in the Design and Analysis of Adaptive Clinical Trials. HRB-TMRN Webinar, 25 June 2020.

Awards

Best talk at Research Students' Conference in Probability and Statistics, Sheffield, July 2018.

Best poster ScHARR Postgraduate Research Conference, Sheffield, May 2018.

Grants

Lead Applicant

Value-based Adaptive Clinical Trial Designs for Efficient Delivery of NIHR Research. Eco-Nomics of Adaptive Clinical Trials (ENACT). NIHR CTU Support Funding variable funding (2019), £79,127, 12 months.

Laura Flight, Prof Alan Brennan, Prof Daniel Hind, Prof Steven Julious, School of Health and Related Research (ScHARR), University of Sheffield. Dr Martin Forster, Department of Economics and Related Studies, University of York. Prof Stephen E Chick, Technology and Operations Area, INSEAD Europe Campus, Fontainebleau, France. Dr Stephen Brealey, Belen Corbacho, Ada Keding, Sarah Ronaldson, Dr Puvanendran Tharmanathan, York Trials Unit, Department of Health Sciences, University of York. Prof Paolo Pertile, Department of Economics, University of Verona, Italy.

Co-Applicant

Adaptive Designs Practical Toolkit (PANDA). NIHR CTU Support Funding variable funding (2018), £129,479, 24 months.

Dr Munya Dimairo, **Laura Flight**, Mike Bradburn, Prof Cindy Cooper, University of Sheffield. Dr Philip Pallmann Centre for Trials Research, Cardiff University. Prof Thomas Jaki Lancaster University. Dr Graham Wheeler Cancer Trials Centre, University College London. Dr Julie Marsh Telethon Kids Institute, University of Western Australia.

Costing Adaptive Trials (CAT): developing best practice costing guidance for CTUs supporting adaptive trials. NIHR CTU Support Funding variable funding (2019) £55,629, 12 months.

Newcastle CTU (NCTU), Sheffield CTRU, Cardiff Centre for Trials Research, Lancaster University, Leeds CTRU, The Institute of Cancer Research CTSU, University of Cambridge.

Abstract

Introduction Adaptive clinical trial designs allow changes to an on-going trial based on early examinations of the data. These designs are increasing in popularity, but it is currently unclear what impact they have on health economic analyses that aim to maximise the health gained for money spent. Additionally, opportunities are potentially being missed to incorporate health economic considerations into the design and analysis of adaptive trials.

Research Question How can health economics be used in the design and analysis of adaptive clinical trials to increase the efficiency of healthcare decision making?

Methods A comprehensive review of trials with an adaptive design and health economic analysis was performed. Healthcare researchers, decision makers and the public were interviewed to understand potential barriers to the use of these methods in practice. The existing theory for the analysis of adaptive trials, focussing on the O'Brien-Fleming and Pocock group sequential designs, was extended to the health economic context to enable accurate within trial and model-based economic analyses. Using the CACTUS case study, the theory of expected value of sample information (EVSI) was extended to guide the design of adaptive trials.

Results None of the 37 trials identified adjusted their analysis for the adaptive nature of the trial where thought necessary. Stakeholders acknowledged that cost-effectiveness decisions must be made, but that clinical effectiveness should remain the focus of a trial. Health economic analyses are vulnerable to bias following a group sequential design and should be adjusted. EVSI methods can be extended and adjusted to help decision makers choose between fixed sample size and adaptive designs with a different number of analyses and stopping rules.

Conclusions Cost-effectiveness considerations are unavoidable in publicly funded healthcare systems with limited budgets. Adaptive trials provide an appealing alternative to costly fixed sample size designs in appropriate scenarios. This thesis extends existing theory to maintain an accurate health economic analysis following an adaptive trial and to guide their design. Recommendations are made to maximise the opportunities to incorporate health economic considerations into the design and analysis of adaptive trials whilst also maintaining the accuracy of healthcare decision making.

Plain English Summary

In the United Kingdom (UK) the National Institute for Health Research (NIHR) make decisions about which research should be given money. All types of research compete for this limited funding and so the NIHR need to weigh up what the new research will tell them and how much it will cost.

The National Institute for Health and Care Excellence (NICE) in the UK, decide which treatments are given to patients on the National Health Service (NHS). NICE often rely on information from clinical trials. In a clinical trial treatments are compared in groups of people who are likely to receive the treatment if it is to be provided by the NHS. This information helps NICE decide whether a treatment improves health, is clinically effective and whether it is going to be value for money, is cost-effective.

In my research, I am interested in clinical trials that use an adaptive design. An adaptive design allows researchers to look at the information collected during a trial, rather than waiting until the trial has ended. This early information is used to make changes to the trial, such as stopping early or adding new treatments to be compared. This can save time and money as well as get the best treatments to patients as soon as possible.

My research aimed to understand how value for money could be used with adaptive designs to make better use of the limited money available for conducting research and funding treatments on the NHS. I wanted to show how this could work in practice to make sure this approach gives NICE accurate information to decide which treatments should be made available and that the public, researchers and decision makers were happy with the new approach.

I looked at how researchers currently assess whether a treatment and research is value for money before and during adaptive designs. Despite researchers having looked at how these methods might work in theory, I found very few trials were using these methods in practice.

To understand why this was the case, I asked researchers, decision makers and the public what they thought about potentially using value for money in clinical trials with an adaptive design. I found that, to those who took part, it was important that the aim of a clinical trial

should be to show that a treatment works (is clinically effective) and that value for money considerations, while important, should not be the focus. Researchers and decision makers felt that they needed more training to use this approach in their trials. They also thought there would need to be changes to how research is funded to allow more time and money to incorporate value for money when designing and running adaptive trials.

Previous research has shown that adaptive designs can cause estimates of how well the treatment works to be over exaggerated or under exaggerated if the right methods are not used. I was concerned that this could also affect how we show value for money. Using the wrong methods could mean the wrong decisions are made about which treatments are made available to patients on the NHS. To understand this I imagined running a clinical trial with an adaptive design and calculated whether the treatment was cost-effective using methods that account for the adaptive design and methods that do not. I showed that in some situations using methods that did not allow for the adaptive design could make a treatment seem cost-effective when really it is not. Therefore, I recommend researchers present results that use these adjustment methods in all adaptive trials.

I also wanted to think about how we could use value for money to help choose the best design for adaptive clinical trials. I took existing methods known as value of information analysis that help us to choose a cost-effective design for a clinical trial and thought about how they could work for adaptive designs. To reflect the importance of showing a treatment works to the public, researchers and decision makers I made sure the focus of the trial was to show clinical effectiveness and that value for money was only thought about before the trial began. I have made recommendations to help researchers apply this approach accurately in the design of their own trials, advising them on how to adjust the methods to allow for the adaptive nature of the trial and to calculate the costs of running the trial.

Overall, my research shows that the public, researchers and decision makers are willing to think about value for money in clinical trials with an adaptive design, but this is not carried out at the moment. I have shown that, by applying existing methods in a new way and thinking carefully about how an adaptive design might affect value for money calculations, the public and the NHS can continue to benefit from using adaptive designs and potentially maximise limited budgets so that more research can be conducted.

Contents

List of Figures

List of Tables

List of Abbreviations

ACE	Adaptive designs CONSORT Extension
AD	Adaptive Design
BAMLE	Bias Adjusted Maximum Likelihood Estimate
BNF	British National Formulary
CACTUS	Computer Treatment Compared to Usual Stimulation
CAT	Costing Adaptive Trials
CHEERS	Consolidated Health Economic Evaluation Reporting Standards
CI	Confidence Interval
CONSORT	Consolidated Standards of Reporting Trials
ConVOI	Collaborative Network for Value of Information
COREQ	Consolidated Criteria for Reporting Qualitative Research
CSLT	Computer Speech and Language Therapy
CTU	Clinical Trials Unit
DMEC	Data monitoring and ethics committee
EEACT	Economic Evaluation Alongside a Clinical Trial
EMA	European Medicines Agency
ENACT	EcoNomics of Adaptive Clinical Trials
ENBS	Expected Net Benefit of Sampling
EVPI	Expected Value of Perfect Information
EVPPI	Expected Value of Partially Perfect Information
EVSI	Expected Value of Sample Information
FDA	Food and Drug Administration
GCSE	General Certificate in Secondary Education
GDHT	Goal Directed Haemodynamic Therapy
GP	General Practice
GRIPP2	Guidance for Reporting Involvement of Patients and the Public 2
GSD	Group Sequential Design
HEDMAP	Health Economic and Decision Modelling Analysis Plan
HEEAD	Health Economic Evaluation of Adaptive Designs
HTA	Health Technology Assessment
ICER	Incremental Cost Effectiveness Ratio
INB	Incremental Net Benefit
ISPOR	The Professional Society for Health Economics and Outcomes Research
MAMS	Multi-Arm Multi-Stage
MLE	Maximum Likelihood Estimate
MRC	Medical Research Council
NA	Not Applicable
NHS	National Health Service
NICE	National Institute for Health and Care Excellence
NIHR	National Institute for Health Research
OPTIMA	Optimal Personalised Treatment of early breast cancer using Multiparameter Analysis
PANDA	Practical Adaptive and Novel Designs and Analysis (Toolkit)
PRESSURE-2	Pressure RElieving Support SUrfaces: a Randomised Evaluation 2

PSA	Probabilistic Sensitivity Analysis
PSSRU	Personal Social Services Research Unit
QALY	Quality Adjusted Life Year
RANE	Rapid Assessment of Need for Evidence
RATPAC	Randomised Assessment of Treatment using Panel Assay of Cardiac markers
RMSE	Root Mean Square Error
SAVI	Sheffield Accelerated Value of Information
SD	Standard Deviation
SLT	Speech and Language Therapist
STAMPEDE	Systemic Therapy for Advanced Metastatic Prostate cancer:Evaluation of Drug Efficacy
SMO	Sample Mean Ordering
UK	United Kingdom
US	United States
VOIA	Value of Information Analysis
WHO	World Health Organisation

Chapter 1

Introduction

1.1 Introduction

Around the world, countries are faced with limited healthcare budgets for funding health technologies, such as new drugs, and health research, such as clinical trials. Conducting efficient research is a priority (National Institute of Health Research, 2020). Adaptive design clinical trials use data collected as the trial progresses to inform modifications to the trial (Gallo *et al.*, 2006). They have the potential to directly benefit patients and the healthcare provider ethically and financially, providing an efficient alternative for conducting clinical trials (Bretz *et al.*, 2009). The methods of health economics facilitate the comparison of the costs and benefits of alternative research designs and health technologies aiding efficient healthcare decision making (Drummond *et al.*, 2015).

In the United Kingdom (UK) healthcare is largely provided by the National Health Service (NHS). The NHS is funded by taxpayers and is freely available at the point of use (National Health Service, 2016). The decision making body for allocating resources for health technologies is the National Institute for Health and Care Excellence (NICE). NICE make recommendations about which healthcare technologies are funded on the NHS (National Institute for Health and Care Excellence, 2014). Through technology appraisal committees comprised of experts, NICE assess evidence of a health technology's clinical and cost-effectiveness (National Institute for Health and Care Excellence, 2013a). Clinical effectiveness is used to demonstrate whether a treatment works, for example does a new blood pressure drug lower a patient's blood pressure. Cost-effectiveness is used to demonstrate whether a treatment is value for money, calculated by balancing the costs against the benefits it gives.

To inform their funding recommendations, organisations such as the National Institute for Health Research (NIHR) are tasked with allocating funding for health research. It is desirable to fund research that aids decision making about which technologies should be provided by the NHS. Research that reduces uncertainty in this decision making is a cost-effective way to allocate limited resources. The cost-effectiveness of research demonstrates whether it is worthwhile conducting a piece of research (such as a clinical trial) by considering what is already known, how much can be learnt from the research and how much it will cost. Independent reviewers including experts in the clinical area, methodologists and the public evaluate each funding application. A decision about whether to fund the application is then made by a committee of experts and the public (National Institute for Health Research, 2017a). The committee and reviewers consider the need for evidence, value for money and scientific rigour (National Institute for Health Research, 2017b).

Clinical trials play an important role in providing the evidence base for decision making. A randomised controlled trial is considered to be the 'gold standard' for research and top of the hierarchy of evidence (Guyatt *et al.*, 1995). Traditionally clinical trials are 'fixed' in their design, known throughout this thesis as a fixed sample size design. At the start of the trial, the design and analysis methods are pre-specified and outlined in a protocol. Once a trial is underway, other than for safety monitoring, the data are not examined until the trial has ended (Committee for Medicinal Products for Human Use, 2006). Clinical trials can be large, costly and slow to conduct and so researchers have looked to alternative approaches such as adaptive design clinical trials.

Adaptive design clinical trials have been proposed as an alternative approach; however, there are potentially issues when only clinical effectiveness is used in the design and analysis of these trials. Resources may be wasted if; cost-effectiveness considerations are not used to inform the design of a trial; a trial ends with insufficient evidence to show cost-effectiveness; a trial continues unnecessarily when there is early evidence that the technology is not cost-effective; or incorrect decisions are made on cost-effectiveness grounds due to bias introduced into the health economic analysis.

1.2 Research Question

This thesis aims to answer the question:

How can health economics be used in the design and analysis of adaptive clinical trials to increase the efficiency of healthcare decision making?

1.3 Research Aims

A multi-disciplinary approach is taken with the following aims:

1. To review the current use of health economics in the design and analysis of adaptive clinical trials in the research literature and in practice.

2. To understand stakeholder views towards the use of health economics in the design and analysis of adaptive clinical trials.

3. To explore the potential for an adaptive design to impact the health economic analysis following a clinical trial.

4. To extend existing health economic methods to guide the design of an adaptive design whilst appropriately accounting for the adaptive nature of the trial design.

1.4 Public Involvement

Public involvement has been embedded throughout this thesis with the support of a public advisory panel. The following sections briefly summarise the involvement of the group. The contributions of the group are highlighted in subsequent chapters.

The aim of including members of the public in this thesis was to represent their views in the development of the research, as the ultimate beneficiaries. Any changes to the methods used to estimate clinical and cost-effectiveness will impact decision making and consequently the treatments made available to patients. If unreliable methods are used then poor decisions may be made. Consequently, treatments may not be made available to patients as money is being wasted on costly alternatives.

The advisory panel was formed following an information session held in April 2016. The information session allowed interested members of the public to find out more information about the research and being a member of the advisory panel. Invitations were sent to existing public involvement groups and to public involvement representatives working on trials in the Sheffield Clinical Trials Research Unit. These groups were targeted, as their members were likely to have some experience of clinical trials or research. This experience was thought to be advantageous to participating in the research; however, it was not a pre-requisite.

Members of the group met annually over the four-year duration of the PhD, with additional meetings organised when required. Annual reports were sent to the group between meetings to up date them on progress and the group were invited to share any thoughts between meetings via email. Further discussion of the involvement and contributions of the public are outlined in Section 1.5, the relevant chapters throughout the thesis and reported in accordance with GRIPP 2 (Guidance for Reporting Involvement of Patients and the Public) guidelines in Chapter 9 (Staniszewska *et al.*, 2017).

1.5 Outline of Thesis

This section gives a description of the thesis structure and a summary of work carried out in each chapter.

Chapter 2 introduces the methods of adaptive clinical trials and health economic analysis that underpin subsequent chapters. This includes a summary of literature relating to the design and analysis of adaptive clinical trials. The potential for an adaptive clinical trial to affect a statistical analysis is summarised, outlining adjustment methods for the primary and secondary point estimates and confidence intervals. Current health economic methods for evaluating the cost-effectiveness of an intervention and the cost-effectiveness of a research design in the context of fixed sample size designs are outlined. Existing literature on the use of health economics in adaptive clinical trials is summarised.

Chapter 3 describes a review of clinical trials with an adaptive design and health economic analysis recorded in sources including `clinicaltrials.gov` and the NIHR Health Technology Assessment journal. Information from each trial is extracted and summarised relating to

their use of health economics in their design, analysis and reporting to establish current practice.

Chapter 4 describes a qualitative study exploring the views of members of the public, researchers and decision makers on the use of health economics in the design and analysis of adaptive clinical trials. Findings are summarised into three themes relating to the ethical, methodological and practical considerations and used to inform work conducted in subsequent chapters. The public advisory panel supported the design, recruitment, interpretation of results and reporting of the qualitative study.

Chapter 5 outlines key criteria for a suitable case study to explore the use of health economics in the design and analysis of adaptive clinical trials. The CACTUS pilot trial, health economic analysis and Big CACTUS trial are introduced, outlining how they met the key criteria. A description of how the original health economic model was reproduced in `R` is provided.

Chapter 6 extends the existing theory for the adjustment of analyses following a group sequential design, introduced in Chapter 2, to the context of a health economic analysis. This is considered for a within trial analysis where health economic outcomes include costs, quality adjusted life years and incremental net benefit. A model-based analysis is also considered where trial data are used to inform model parameters. The theory is extended based on the CACTUS case study introduced in Chapter 5.

Chapter 7 takes the theory developed in Chapter 6 and explores the extent to which the health economic analysis following a group sequential design is affected by the adaptive nature of the trial. A simulation study is used to assess the extent to which the stopping rule, number of interim analyses and correlation between primary and health economic outcomes affect the accuracy of the within trial and model-based health economic analysis. Adjusted and unadjusted analysis methods are compared to assess how well existing adjustment methods account for potential biases. The work in this chapter reflected the advice of the public advisory group to build on current practice.

Chapter 8 extends existing methods for assessing the cost-effectiveness of proposed fixed sample size designs to the adaptive design setting. Based on suggestions from stakeholders in Chapter 4 and advice from the public advisory group, clinical effectiveness is the focus of the

adaptive decision making once the trial is running. The methods are applied to the CACTUS case study to guide the design of a trial with appropriate adjustments for the adaptive nature of the design.

Chapter 9 draws together the findings in each chapter and summarises how they fulfil the research question and specific research aims. Recommendations are made that consider ethical, methodological and practical aspects from the perspective of all stakeholders, for the use of health economics in the design and analysis of adaptive trials. The public advisory panel co-wrote the summary and reflection of public involvement in the thesis and helped to inform the recommendations made.

Chapter 2

Background

2.1 Introduction

Healthcare decision making is dominated by the need to use scarce resources as efficiently as possible. Evidence of clinical and cost-effectiveness is key. Clinical trials play an important role in providing this evidence; however, they can be costly and slow to conduct. Alternative, more efficient designs have been developed and include adaptive design clinical trials. Additionally, the use of health economics to establish the cost-effectiveness of research and health technologies can be used to inform the efficient allocation of healthcare budgets for research and healthcare interventions. This thesis considers the methods of adaptive clinical trials and health economic analysis together to increase the efficiency of clinical trials and healthcare decision making.

2.2 Chapter Aims

This chapter introduces the adaptive clinical trial design, highlighting available methods and their advantages and limitations. The group sequential design, a specific type of adaptive trial, and approaches to their design and analysis are summarised. The potential for an adaptive design to affect the analysis of primary and secondary outcomes is explained.

The methods of health economics and their role in healthcare decision making are summarised. Economic evaluation methods are described for establishing the short-term and long-term cost-effectiveness of a health technology. Bayesian decision theory is introduced and the methods of value of information analysis described in the context of fixed sample size design clinical

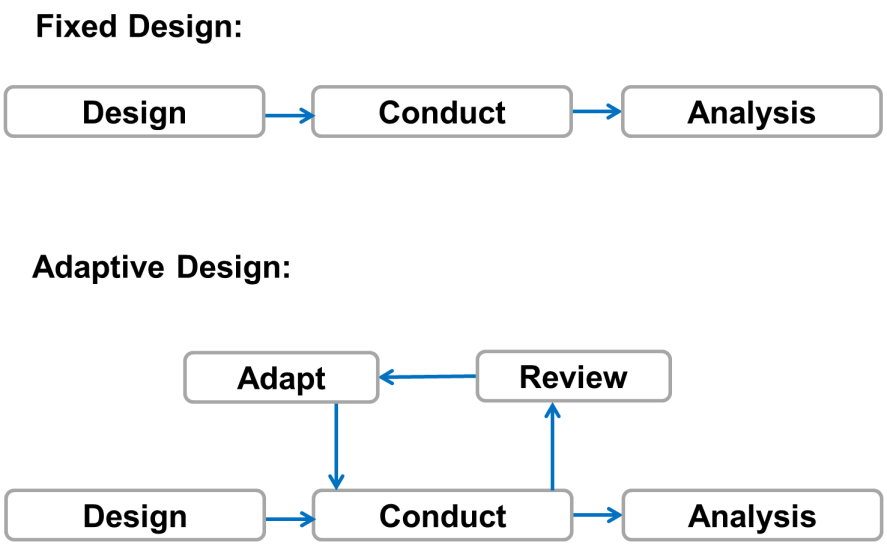


FIGURE 2.1: Illustration of an adaptive clinical trial design compared to a fixed sample size design taken from Pallmann *et al.*, 2018. This article is distributed under the terms of the Creative Commons Attribution 4.0 International License (http://creativecommons.org/licenses/by/4.0/)), which permits unrestricted use, distribution, and reproduction in any medium

trials. Literature describing the use of adaptive designs and health economics together is summarised.

2.3 Adaptive Design Clinical Trials

Adaptive design clinical trials use data collected as a trial progresses to inform modifications to the trial, without compromising the validity or integrity of the study (Gallo *et al.*, 2006). Rather than waiting until the trial has ended, the accumulating data are used to update the uncertain assumptions made before the trial began (Chow *et al.*, 2011). This information is used to make changes or modifications to the trial as illustrated in Figure 2.1.

While many clinical trials will have some element of monitoring to ensure patient safety, adaptive designs are a formalised process with pre-specified methods and rules guiding modifications (Chow *et al.*, 2012a). Throughout this thesis the definition of an adaptive trial is taken from the CONSORT extension for adaptive designs (Dimairo *et al.*, 2019b):

'A clinical trial design that offers pre-planned opportunities to use accumulating trial data to modify aspects of an ongoing trial while preserving the validity and integrity of that trial.'

2.3.1 Advantages and Limitations

Adaptive clinical trials can save time and resources as well as prevent patients from being needlessly randomised (Bretz *et al.*, 2009; Chow *et al.*, 2012b; Pallmann *et al.*, 2018). They offer greater flexibility, provide the opportunity to update uncertain assumptions made before the trial began and may allow a treatment to move through the development process more quickly (Chow *et al.*, 2011).

The potential benefit of these designs is evident from the re-analysis of the emergency medicine clinical trial RATPAC (Randomised Assessment of Treatment using Panel Assay of Cardiac markers) (Goodacre *et al.*, 2011). This multi-centre, fixed sample size design trial evaluated the use of point-of-care marker panels in the treatment of patients with suspected acute myocardial infarction. The original trial concluded that the treatment was clinically effective but was not cost-effective when compared to usual care (Goodacre *et al.*, 2011; Fitzgerald *et al.*, 2011). If RATPAC had used an adaptive design it could have potentially stopped one year earlier, saving approximately £250,000 and requiring 1,521 fewer patients (Sutton *et al.*, 2012).

Adaptive designs are not without limitations. Examining the data multiple times at interim analyses has the potential to introduce bias (Pallmann *et al.*, 2018). It is crucial, therefore, that appropriate methods are employed to avoid this issue, a topic which is discussed further in Section 2.4 in relation to group sequential designs. Adaptive methods should not be used solely because they are seen to be innovative and should only be used with careful justification, consideration and planning (Stevely *et al.*, 2015; Wason *et al.*, 2019). For example, in a trial with a short recruitment period but a long follow-up time before observing the primary outcome will be close to ending recruitment by the time an interim analysis takes place and modifications are made to the trial (Pocock, 1983).

2.3.2 Current Use and Perceptions

In a review by Hartford *et al.*, 2018 of 150,074 trials reported in TrialTrove between 2000 and 2015 only 2,508 (2%) contained an adaptive design search term. Dimairo *et al.*, 2015 highlight a number of practical concerns acting as barriers to the routine use of these designs. These

include a lack of capacity and financial support for the development of adaptive designs. These trial designs are more complex and require specific expertise compared to fixed sample size trials, which can be time consuming and costly especially if this is to take place before a trial is funded.

The number of trials using adaptive methods is, however, increasing from 11 per 10,000 registered trials between 2001 and 2005 to 38 per 10,000 registered trials between 2012 and 2013 (Hatfield *et al.*, 2016). Adaptive designs are most frequently used in Oncology and are more prevalent in North America and Europe (Hatfield *et al.*, 2016; Mistry *et al.*, 2017; Bothwell *et al.*, 2018). This increase can be explained by the change in attitude of regulators, greater training and understanding, as well as a greater appreciation for the need to conduct clinical trials as efficiently as possible (Meurer *et al.*, 2016). Literature identifying the practical limitations of adaptive designs and solutions to overcome these has helped to bridge the gap between methodological development and practical implementation (Quinlan *et al.*, 2010; Coffey *et al.*, 2012; Kairalla *et al.*, 2012; Morgan *et al.*, 2014b; Dimairo *et al.*, 2015).

2.3.3 Types of Adaptive Clinical Trials

The statistical methods for adaptive designs are developing quickly (Kairalla *et al.*, 2012). A variety of options are available for all phases of drug development (Chow *et al.*, 2012b). The following sections describe some of the most common types of adaptive trial designs.

2.3.3.1 Sample Size Re-estimation

In a sample size re-estimation, the sample size calculated before a trial commences is updated using information collected up to an interim analysis (U.S. Food and Drug Administration, 2019). This can be conducted blind to the treatment allocation or unblind. A blinded sample size re-estimation does not require any statistical adjustment, however when unblinded data are used the probability of making a type I error increases (declaring a false positive result). Generally, these methods are only recommended to increase the sample size. For example, prior to the EVIDENCE study (North *et al.*, 2011), the study team identified that there was a lack of information to inform their sample size. They therefore outlined in the trial protocol that they would conduct an interim analysis to re-estimate the required number of patients to achieve sufficient statistical power.

2.3.3.2 Adaptive Randomisation

Trials with adaptive randomisation allow modifications to the randomisation procedure after the trial has commenced based on the allocation of previous participants in the trial (Chow *et al.*, 2012b). This can increase the probability of success of the trial and allow new participants to receive the most promising treatment. For example, the DexFEM trial used adaptive randomisation to adapt the treatment allocation probabilities based on the data collected during the trial. This allowed women subsequently recruited to receive doses that were more informative about the dose-response relationship of oral dexamethasone for amelioration of heavy menstrual bleeding (Warner *et al.*, 2015).

2.3.3.3 Multi-Arm Multi-Stage

Multi-Arm Multi-Stage (MAMS) trials allow multiple treatments to be compared to a single control arm. The multiple stages (interim analyses) increase efficiency by allowing arms to be dropped for futility or even for the whole trial to stop if efficacy can be demonstrated (Pallmann *et al.*, 2018). For example, the Systemic Therapy for Advanced or Metastatic Prostate cancer: Evaluation of Drug Efficacy (STAMPEDE) trial began by exploring five active and one control treatment, each evaluated at three stages; a pilot stage, three intermediate 'activity' stages, and a final 'efficacy' (Sydes *et al.*, 2012). This trial design allowed the comparison of multiple active treatments to a single control arm simultaneously, saving time and resources.

2.4 Group Sequential Trials

In a group sequential design clinical trial, interim analyses are carried out after groups of patients have reached the outcome of interest (Jennison *et al.*, 2000). The interim data are analysed and compared to pre-specified stopping rules to determine whether the trial can stop early. It is common for stopping rules and interim decision making to be based solely on the clinical primary outcome. A trial might stop early for efficacy – the new treatment works much better than the alternative or for futility – the new treatment is not working as expected, or both. Simply, the aim of a stopping rule is to identify when a trial can stop with sufficient confidence that the decision that is made based on the early data would be the same had the study continued (Emerson, 2012).

An example of a group sequential design is given by Tröger *et al.*, 2013. In this trial, an O'Brien-Fleming stopping rule (described in Section 2.4.1) was used to analyse data collected after 220, 320 and 480 patients had been recruited to the trial. The interim analyses were based on the clinical outcome, overall survival, in patients with late-stage pancreatic cancer randomised to either viscum album or no antineoplastic therapy.

The group sequential design is the focus of this thesis as it is one of the most common types of adaptive design used in practice (Stevely *et al.*, 2015; Hatfield *et al.*, 2016; Hartford *et al.*, 2018; Bothwell *et al.*, 2018; U.S. Food and Drug Administration, 2019). The impact of the design on subsequent analyses is well documented with a number methods available to adjust for this, as will be described in Section 2.5. It is felt that group sequential designs are likely to be providing data for many health economic analyses given their common use. There is an urgent need to better understand the impact of these designs on a health economic analysis and a strong knowledge basis to extend the methods of adaptive designs to the health economic context.

2.4.1 Designing a Group Sequential Trial

During a group sequential design the data are examined multiple times which has the potential to increase the chance of declaring a false positive result (making a type I error) (Bretz *et al.*, 2009). Consequently, group sequential stopping rules are designed to control the type I error at the desired level, typically 5%, (Jennison *et al.*, 2000). To control the error the required maximum sample size may increase compared to the fixed sample size design with the same properties but with no early examinations of the data.

At an interim analysis the difference in means (or another chosen statistic) between the intervention and control arm for the outcome of interest (typically a clinical outcome) is calculated. This value is compared to the pre-specified stopping rule. If the estimate crosses the value specified by the stopping rule (often referred to as the stopping boundary), the trial will stop at that point. If the estimate falls within the boundary the trial will continue to the next analysis. This process is repeated at each interim analysis using all accumulated evidence until the trial crosses the boundary or reaches the final analysis.

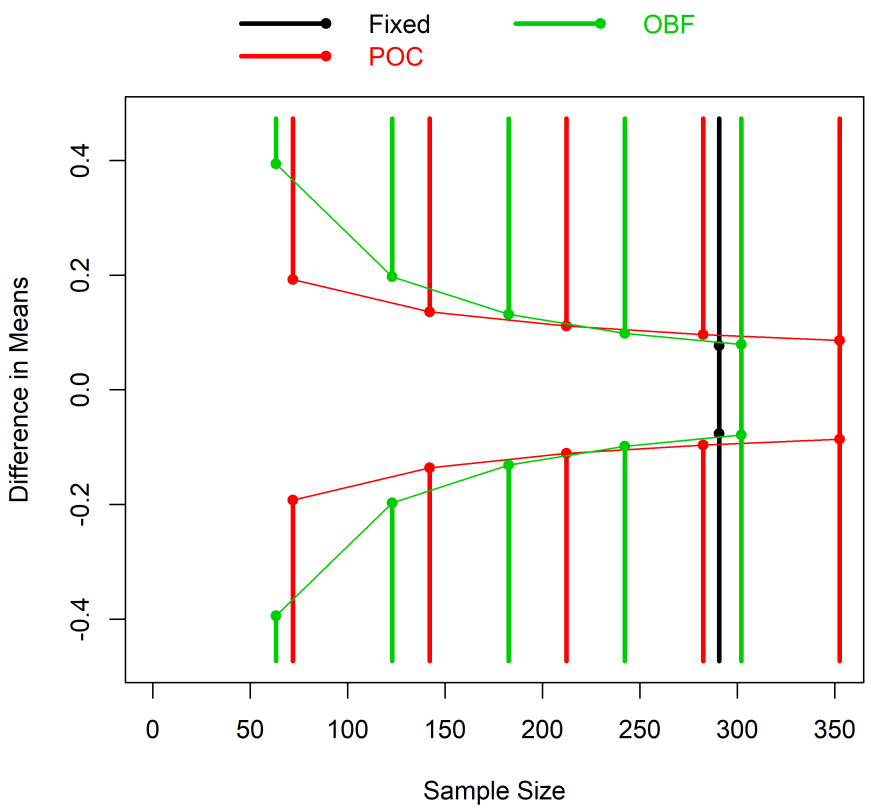


FIGURE 2.2: Illustration of the stopping boundary for an O'Brien-Fleming (green) and Pocock (red) stopping rules for a trial with up to five analyses.

Available stopping rules include Pocock (Pocock, 1977) and O'Brien-Fleming (O'Brien *et al.*, 1979). Each rule has different characteristics and varying impact on the design (such as maximum sample size) and subsequent analyses (Jennison *et al.*, 2000). Figure 2.2 illustrates the different characteristics of the Pocock and O'Brien-Fleming stopping rules compared to the fixed sample size design trial. The difference in sample mean between the intervention and control arm is plotted against the number of participants at an interim analysis. The Pocock stopping rule (red) requires a larger maximum sample size if the trial does not stop early, however there is a lower hurdle for stopping the trial at the first and second interim analyses compared to the O'Brien-Fleming rule (green). The black line gives the sample size for the equivalent fixed sample size design.

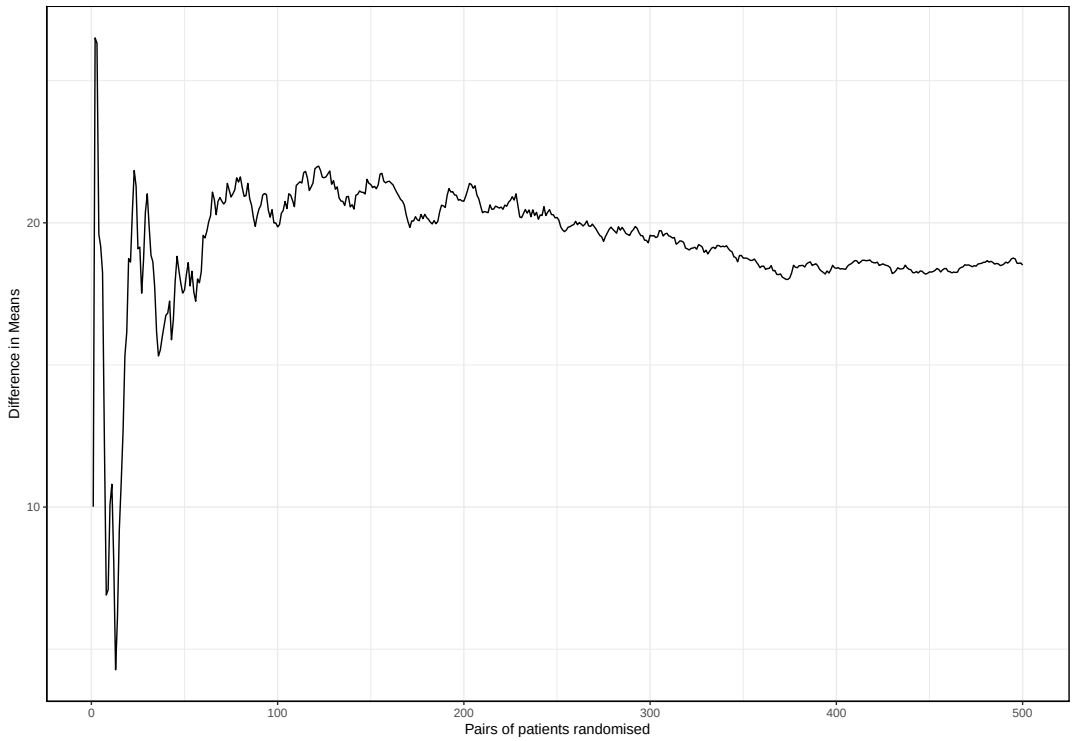


FIGURE 2.3: Illustration of the initial variability in the treatment effect estimate (difference in means between intervention and control) for pairs of patients randomised as the trial progresses and regression to the mean as the sample size increases

2.4.2 Analysing a Group Sequential Trial

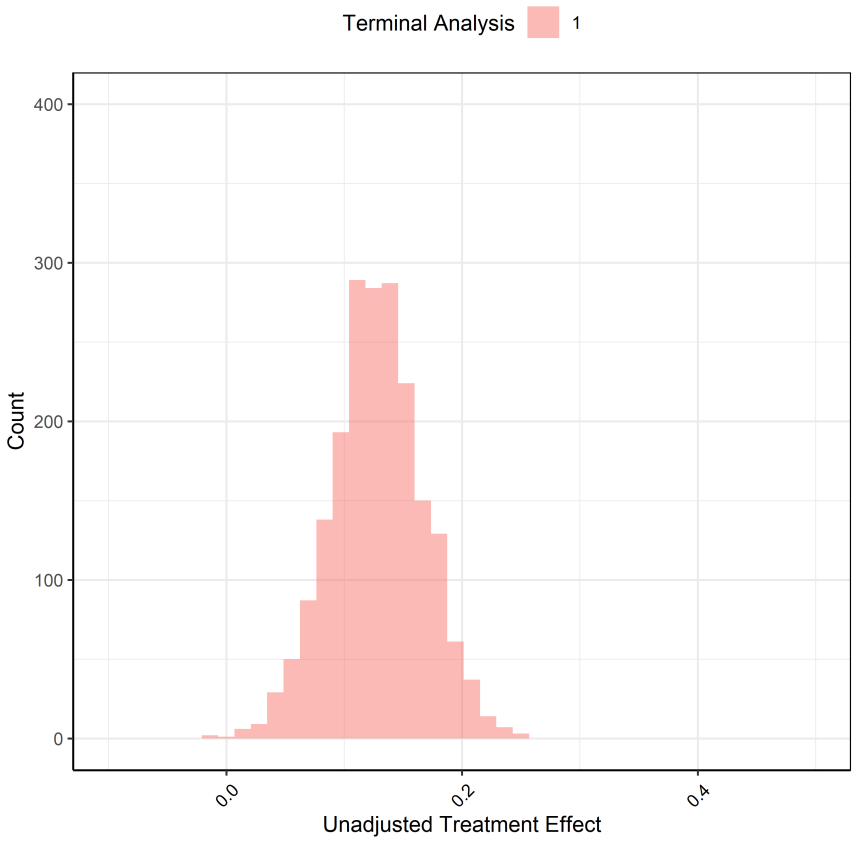
The analysis following group sequential designs requires careful consideration and it is important to account for the adaptive nature of the study design. It is not appropriate to use the same methods as if the trial had a fixed sample size design. When a group sequential design stops early for efficacy the point estimate of the primary outcome may be biased upwards or biased downwards when stopping for futility (Whitehead, 1997). The confidence interval following a group sequential design is also affected and may not have desirable properties such as correct coverage, appropriate length and may not contain the maximum likelihood estimate (MLE) (Jennison *et al.*, 2000).

To understand why this occurs consider the start of a trial where there is a lot of variation in the estimate of the treatment effect. This is explained by the fact that a small number of participants have been randomised at this point (Zhang *et al.*, 2012). This is illustrated in Figure 2.3 where the cumulative difference in mean primary outcome for simulated pairs of patients randomised to either the treatment or control arm of a trial is plotted.

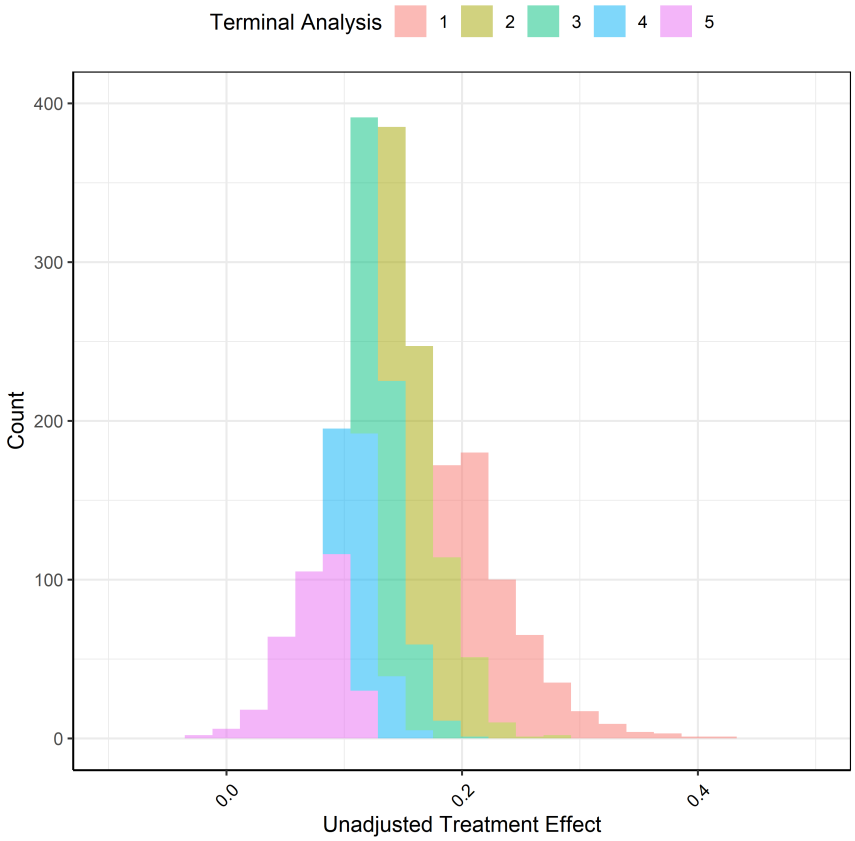
It is possible that the first ten participants have a good response to treatment and so there is a peak in the treatment effect estimate. The next 20 participants may have a poor response to treatment and so there will be a drop in the estimate. In a fixed sample size design trial, these random highs and random lows go undetected, as there are no early examinations of the data. These extreme results are then diluted as more participants are recruited and the treatment effect estimate regresses to the mean (Pallmann *et al.*, 2018). However, in a group sequential design, at an early examination of the data, a random high or low might cross the chosen stopping boundary and the decision made to stop the trial early. This preferential stopping of a trial when an extreme result is observed can introduce bias (Ellenberg *et al.*, 2010).

In the frequentist framework the trial analysis is based on, the hypothetical, repeated sampling of an experiment (Swinscow, 1997). For a fixed sample size design, for large samples, the sampling distribution of the MLE for a continuous primary outcome follows a Normal distribution by the Central Limit Theorem (Swinscow, 1997). This is illustrated in Figure 2.4a for 2,000 simulated trials. This Normality assumption forms the basis for analysis. In a group sequential design, the sampling distribution of the primary outcome MLE becomes skewed, as the opportunity to stop the trial early is introduced (Pinheiro, 1997; Jennison *et al.*, 2000). This is illustrated in Figure 2.4b for 2,000 trials using the Pocock stopping rule with five analyses. Some of the 2,000 simulated trials stop at the first interim, second, third and so on, creating skew in the distribution. This distribution is referred to as the sequential distribution.

The extent to which the point estimate of the primary outcome is biased following a group sequential design will depend on the stopping rule used (Chow *et al.*, 2012b). Pinheiro, 1997 found that the bias was greater for the Pocock stopping rule when compared to the O'Brien-Fleming stopping rule when the treatment effect was small but the opposite was true when the treatment effect was large. They also showed that the bias was reduced the later the interim analyses were conducted in the trial. Emerson *et al.*, 1990 discuss how the stopping rule influences the bias in the subsequent confidence interval. The authors state that generally the behaviour of the unadjusted confidence interval is worse for the Pocock stopping rule than the O'Brien-Fleming stopping rule.



(A) Fixed sample size design



(B) Pocock stopping rule with up to five analyses

FIGURE 2.4: Distribution of maximum likelihood estimate of the treatment effect for 2,000 trials

2.5 Bias Adjustment Methods Following a Group Sequential Trial

Methods are available that adjust the point estimate and confidence interval of the outcomes from a group sequential design to account for the adaptive nature of the trial (Jennison *et al.*, 2000). This helps to ensure inferences made following the trial are as accurate and reliable as possible. This is important in healthcare decision making where the results of a group sequential design may be used to change clinical practice affecting which treatments are made available to patients.

In this thesis the bias adjusted maximum likelihood estimate (BAMLE) approach for adjusting point estimates (Whitehead, 1986a; Whitehead, 1986b) and the sample mean ordering (SMO) approach for adjusting confidence intervals (Emerson *et al.*, 1990) are described in Section 2.5.1 and Section 2.5.4 and used in subsequent chapters. These methods have been chosen for exploration as Emerson *et al.*, 1990 recommend the SMO approach as it gives uniformly average shorter confidence intervals, however, point estimates based on the SMO were no better than Whitehead's BAMLE (Whitehead, 1986a; Emerson *et al.*, 1990).

2.5.1 Adjusted Point Estimate for Primary Outcomes

Whitehead, 1986a's BAMLE method calculates an adjusted estimate of the treatment effect for the primary outcome by subtracting an estimate of the bias. This adjustment, in effect, reduces the bias in the point estimate but will not necessarily eliminate it. The treatment effect for the primary outcome might be the mean difference between the intervention and control arm of a trial. Let

θ_1 be the true treatment effect parameter for the primary outcome,

$\hat{\theta}_1$ be the MLE and,

$\tilde{\theta}_1$ be the BAMLE.

The BAMLE is calculated using (Whitehead, 1986a) (page 578)

$$\tilde{\theta}_1 = \hat{\theta}_1 - b(\theta_1), \tag{2.1}$$

where

$b(\theta_1)$ is the bias.

The bias is calculated using

$$b(\theta_1) = \mathbb{E}_{\theta_1}\left(\hat{\theta}_1\right) - \theta_1, \tag{2.2}$$

where

$\mathbb{E}_{\theta_1}(\cdot)$ is the expected value over θ_1.

In order to solve Equation 2.2 a Newton-Raphson iterative procedure is required as the true value θ_1 is unknown.

2.5.1.1 Newton-Raphson Procedure

The Newton-Raphson procedure can be used to find a solution to non-linear equations such as Equation 2.1 (Smith, 2020). As described by Whitehead, 1986a (page 578), solving Equation 2.1 requires a starting value that is a 'best guess' of θ_1. Given available information, the MLE ($\hat{\theta}_1$) can be used as the starting value given by

$$\tilde{\theta}_{1,0} = \hat{\theta}_1. \tag{2.3}$$

The starting value is then substituted into the following iterative equation (Whitehead, 1986a),

$$\tilde{\theta}_{1,i} = \tilde{\theta}_{1,i-1} + \frac{\left(\hat{\theta}_1 - \tilde{\theta}_{1,i-1}\right) - b(\tilde{\theta}_{1,i-1})}{1 + b'\left(\tilde{\theta}_{1,i-1}\right)}, \tag{2.4}$$

where

$b'(\cdot)$ is the first derivative of the bias,

The bias in Equation 2.2 and its first derivative are estimated from the sequential distribution of the test statistic using recursive numerical integration (Armitage *et al.*, 1969) or via simulation.

The iteration is repeated replacing the starting value with the estimate of $\tilde{\theta}_1$ from the previous step until convergence is reached. This provides a BAMLE of the primary outcome that accounts for the adaptive nature of the trial.

2.5.2 Adjusted Point Estimate for Secondary Outcomes

Clinical trials often collect information about a number of key secondary outcomes that will also require adjustment following a group sequential design. Whitehead, 1986b extend the BAMLE method to calculate adjusted point estimates of secondary outcomes.

Let

θ_2 be the treatment effect parameter for the secondary outcome,

$\hat{\theta}_2$ be the MLE and,

$\tilde{\theta}_2$ be the BAMLE.

The BAMLE is calculated using (Yan *et al.*, 2009; Jennison *et al.*, 2015),

$$\tilde{\theta}_2 = \hat{\theta}_2 - \rho\frac{\sigma_2}{\sigma_1}\Big(\hat{\theta}_1 - \tilde{\theta}_1\Big), \tag{2.5}$$

where

ρ is the pooled correlation coefficient between primary and secondary outcomes across arms,

σ_1 is the pooled standard deviation for the primary outcome across arms,

σ_2 is the pooled standard deviation for the secondary outcome across arms.

In practice, the MLEs $\hat{\theta}_1, \hat{\theta}_2$ are calculated from the trial data as they would be for a fixed sample size design. The BAMLE for the primary outcome ($\tilde{\theta}_1$) is estimated using the steps described in Section 2.5.1. The standard deviations and correlation can also be estimated from the trial data, as they would be for a fixed sample size design.

The correlation ρ can be a Pearson correlation coefficient or a Spearman rank correlation. The Pearson correlation coefficient relies on an assumption of Normality and is not invariant to non-linear transformations such as the log-transformation (Briggs *et al.*, 2006). As discussed in Section 2.7, health economic data can be non-Normal hence violating the Normality assumption, therefore, throughout this thesis, the Spearman rank correlation is used.

2.5.3 Adjusted Point Estimate for Absolute Values

In addition to the primary and secondary outcomes, it may be of interest to calculate the marginal or absolute effects in one of the trial arms. For example, a health economist may be interested in the marginal effect of the primary outcome in the intervention arm only rather than the difference between the two arms for their health economic model. Let

$\nu_{1,I}$ be the marginal effect of the primary outcome in the intervention arm,

$\nu_{1,C}$ be the marginal effect of the primary outcome in the control arm.

Using the approach of Whitehead, 1997 adjusted estimates are given by

$$\tilde{\nu}_{1,C} = \bar{\nu}'_1 + \frac{n_I}{n}\tilde{\theta}_1, \tag{2.6}$$

$$\tilde{\nu}_{1,I} = \bar{\nu}'_1 - \frac{n_C}{n}\tilde{\theta}_1, \tag{2.7}$$

where

$\bar{\nu}'_1$ is the sample mean outcome across both treatment arms,

n_I is the number of participants in the intervention arm,

n_C is the number of participants in the control arm,

n is the total number of participants.

Marginal effects for secondary outcomes are denoted by $\tilde{\nu}_{2,I}$ and $\tilde{\nu}_{2,C}$ calculated by replacing terms in Equation 2.6 with the relevant secondary outcome equivalents.

2.5.4 Adjusted Confidence Intervals for Primary Outcomes

As discussed in Section 2.4.2, the confidence interval for the primary outcome is also affected by the adaptive nature of a group sequential design. Adjustments are required to ensure the interval has the desired properties (Jennison *et al.*, 2000).

In the frequentist framework a p-value is the probability, under the null hypothesis, of observing a test statistic as or more extreme than that observed in the actual trial (Chang, 2014). In a fixed sample size design trial, the calculation of the p-value and confidence interval is straightforward as the set of all possible results from the trial, known as the sample space, is one dimensional. The sample space can, therefore, be ordered based on the real numbers (Wassmer *et al.*, 2016). Let

$$Z_A, Z_B \text{ denote two test statistics,}$$

for two datasets A and B. If $|Z_A| > |Z_B|$ this implies that Z_A is more extreme than Z_B. This allows a p-value function to be defined and used to calculate p-values and confidence intervals following the fixed sample size design.

2.5.4.1 The p-Value Function

The p-value function for the primary outcome (θ_1) is defined to be

$$P(\theta_1) = P(X \geq_e x; \theta_1), \tag{2.8}$$

where

X is a random variable representing possible alternative datasets for the trial,

x is an observed value of X,

$\geq_e$ denotes that x provides evidence as or stronger than X.

The p-value function can be defined for a range of θ_1 values and used to calculate a two-sided p-value with

$$p = 2\text{min}\big\{P(0), 1 - P(0)\big\}, \tag{2.9}$$

where $P(0)$ is the p-value function evaluated at $\theta_1 = 0$.

A $100(1 - \alpha)\%$ confidence interval can be calculated from the percentiles of this function. Let

$\theta_{1,L}$ be the lower bound of the confidence interval for the primary outcome,

$\theta_{1,U}$ be the upper bound of the confidence interval for the primary outcome.

These values are calculated from the p-value function using

$$P(\theta_{1,L}) = \frac{\alpha}{2}, \tag{2.10}$$

$$P(\theta_{1,U}) = 1 - \frac{\alpha}{2}, \tag{2.11}$$

where α is the desired significance level.

When a group sequential design is used the sample space is two-dimensional depending on both (Z, M) where

Z is the test statistic,

M is the interim analysis number.

To calculate the p-value function it is required to understand when a sequential test statistic is as or more extreme, which is now more complicated. To determine whether observed statistics (Z_A, M_A) are more extreme than (Z_B, M_B) an ordering of the sample space needs to be defined. Several alternative suggestions have been proposed. To date there is no best or uniformly optimal choice (Emerson *et al.*, 1990; Emerson, 2012). In this thesis, the SMO approach is used, as recommended by Emerson *et al.*, 1990.

2.5.4.2 Sample Mean Ordering of the Sample Space

The SMO provides an ordering of the multidimensional sample space following a group sequential design. The SMO approach considers the observed result of the group sequential design to be as or more extreme than another result if the estimate of the treatment effect (its MLE) is more extreme, that is,

$$(Z_A, M_A) > (Z_B, M_B) \text{ iff } \hat{\theta}_A > \hat{\theta}_B. \tag{2.12}$$

This ordering is illustrated in Figure 2.5, where the dashed lines represent values of the sample mean that would be considered stronger evidence than the sample mean observed, marked with a cross. This ordering is used as it was shown by Emerson *et al.*, 1990 to perform best when compared to the ordering suggested by Tsiatis *et al.*, 1984. The Tsiatis ordering (an alternative to the SMO considered here) states that data points crossing the boundaries at earlier interim analyses than that observed and data more extreme than the point estimate observed at the same interim analysis that the trial stops at are classed as more extreme. The SMO also performs better than the ordering proposed by Chang *et al.*, 1986 that defines the ordering based on the likelihood principle.

Once an ordering has been specified the p-value function can be calculated using Equation 2.8 and 2.9 using recursive numerical integration to approximate the sequential distribution.

2.5.5 Adjusted Confidence Intervals for Secondary Outcomes

The SMO approach can also be used to calculate adjusted confidence intervals for secondary outcomes following a group sequential design. The p-value function for secondary outcomes can be constructed using the simulation based approach outlined by Skalland, 2015. It is necessary to understand the distribution of the sample mean of the secondary outcome (θ_2). This can be estimated under a grid of potential values denoted by θ_2^*. Using an estimate of the primary outcome $\hat{\theta}_1$ and the correlation between the primary and secondary outcome $\hat{\rho}$, calculated from the trial data. The distribution can then be simulated over the chosen grid values.

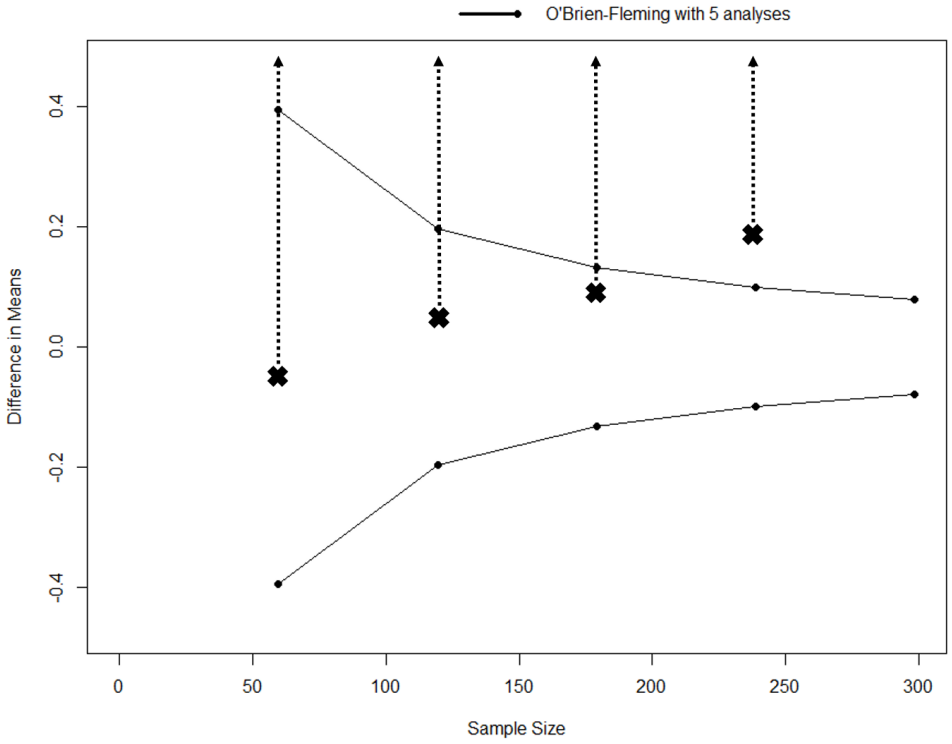


FIGURE 2.5: Illustration of the how the sample mean ordering defines an ordering of stronger evidence when the sample mean is more extreme

The observed sample mean for the secondary outcome ($\hat{\theta}_2$) is compared against the grid using the SMO to give a one-sided p-value, denoted by $p^{(1)}$, for each grid value,

$$p^{(1)}(\theta_2^*) = P(\theta_2 \geq \hat{\theta}_2; \theta_2). \tag{2.13}$$

This is the probability, when the actual treatment effect is θ_2, of obtaining evidence as strong or stronger than $\hat{\theta}_2$ from a clinical trial of the same design. The two-sided p-value is calculated using Equation 2.9. The confidence interval is then given by taking the lowest and highest grid values where the two-sided p-value is greater than α.

2.6 `RCTdesign` Package to Design and Analysis a Group Sequential Trial

To implement the theory described in the previous sections the `R` package `RCTdesign` is used. The `RCTdesign` package (`RCTdesign.org`) was originally created by Professor Scott Emerson for use in `SPlus` but has since been transferred and developed in `R`. This package has a

range of functions that allow the design (`seqDesign`), monitoring and inference of a group sequential design using appropriate adjustment (`seqMonitor`). As this is freely available software it is possible to generate outputs and code from the thesis that can be implemented easily by other researchers.

The package is used in this thesis to calculate the stopping rules considered in the simulation study of Chapter 7 and the exploration of health economics in the design of an adaptive trial in Chapter 8. The package is used to calculate adjusted point estimates for the primary outcome using the BAMLE approach and adjusted confidence intervals for the primary outcome using the SMO approach.

At present, the package does not include adjusted point estimates or confidence intervals for secondary outcomes. I have written additional `R` code to incorporate these adjustments for this research. The point estimate of secondary outcomes has been coded using Equation 2.5 and adjusted confidence intervals using code adapted from Dr Timothy Skalland using the approach outlined in Section 2.5.5.

2.7 Health Economics Analysis

To facilitate the transparent and fair allocation of scarce resources to fund health technologies and health research the methods of health economics namely; economic evaluation (establishing the cost-effectiveness of an intervention) and value of information analysis (establishing the cost-effectiveness of the research) have been developed with the aim of maximising the health gained for the money spent (Drummond *et al.*, 2015).

Clinical trials are commonly used as the vehicles for collecting information for a health economic analysis. The following sections outline existing theory and approaches for conducting a health economic analysis in the health technology assessment setting.

2.7.1 Economic Evaluation

An economic evaluation is described by Drummond *et al.*, 2015 as the 'comparative analysis of alternative courses of action in terms of both their costs and consequences'. These methods have been developed to determine which allocation of resources maximises the health gains

from limited research budgets. The costs and consequences of all relevant technologies must be identified, measured and valued.

The most common approach for the economic evaluation of health technologies is a cost-effectiveness analysis (Jones, 2006). The costs of a health technology are compared to a single outcome, for example the cost per unit of effect (Drummond *et al.*, 2015). Cost-utility analysis is a variant of cost-effectiveness analysis where the unit of effect is a measure of health, known as utility. Cost-utility analysis is the preferred approach for a number of healthcare decision making bodies including NICE in the UK, Canada and Australia (Pharmaceutical Benefits Advisory Committee, 2008; National Institute for Health and Care Excellence, 2013a; Canadian Agency for Drugs and Technologies in Health, 2017).

2.7.2 Calculating Costs of a Healthcare Intervention

The key steps in any cost analysis are identification, measurement and valuation (Drummond *et al.*, 1993). Exactly what should be included will depend on the perspective of the analysis. Different perspectives include; societal, patient, employer and healthcare provider. In the UK, NICE stipulate that economic evaluations should use an NHS and personal and social services perspective (National Institute for Health and Care Excellence, 2013a). This should include all costs incurred by the NHS in the delivery of the technology under evaluation.

Once the perspective has been defined all relevant costs should be measured, using a microcosting study, case report forms completed during a clinical trial or existing data . Each resource is then valued using sources such as the Personal Social Services Research Unit (PSSRU) document of up-to date unit costs for a range of health and social care services in England (Curtis *et al.*, 2015) and the British National Formulary (BNF) for drug costs (Joint Formulary Committee, 2016). NICE require costs to be discounted by 3.5% so they reflect the present value of any costs and benefits to be accrued, known as discounting (National Institute for Health and Care Excellence, 2013a).

2.7.3 Calculating Benefits of a Healthcare Intervention

In a cost-utility analysis, benefits are commonly measured using the quality adjusted life year (QALY) (National Institute for Health and Care Excellence, 2013a). A QALY is a measure of both the quantity and quality of life gained (Drummond *et al.*, 2015). The QALY is a 'common

currency' that allows decision makers to compare the benefits of a variety of interventions, across a range of disease areas (Whitehead *et al.*, 2010). The QALY is calculated by multiplying the amount of time spent in a health state by its utility. The utility quantifies the quality of life experienced in a particular health state and is anchored between zero and one, where one represents full health and zero represents death. It is possible for utility values to fall below zero representing health states thought to be worse than death (Patrick *et al.*, 1994).

In the UK, NICE prefer utilities to be measured using the EQ-5D, a measure of health that does not depend on a specific illness, condition or patient population. The EQ-5D consists of five domains; mobility, self-care, usual activities, pain/discomfort and anxiety/depression (Herdman *et al.*, 2011). Each domain has five levels ranging from no problems (1) to extreme problems (5). During a clinical trial, participants are often asked to complete the EQ-5D at a number of time points during the study to describe their quality of life at that given time. Their responses give a five level descriptive health state, for example 11111 describes someone in perfect health with no problems in each of the five domains. This descriptive health state is then converted to a utility score based on existing algorithms for the population of interest. Currently, NICE recommend the five level health states are converted to a utility score based on a previous version of the EQ-5D with three levels for each domain (Van Hout *et al.*, 2012; National Institute for Health and Care Excellence, 2013b). A QALY is then calculated for each participant over the duration of the trial using linear interpolation.

2.7.4 Incremental Analysis

In a health economic analysis, two or more competing interventions are compared in an incremental analysis. Two key summary statistics include the incremental cost-effectiveness ratio (ICER) and the incremental net benefit (INB). The ICER divides the incremental costs by the incremental benefits of two competing health technologies (Drummond *et al.*, 2015). As the ICER is a ratio it has undesirable statistical properties that make the calculation of a confidence interval challenging (Briggs *et al.*, 2002). This makes it difficult to represent the uncertainty of the ICER (Fenwick *et al.*, 2001). For example, when a negative ICER is reported there is no way to distinguish between a scenario where the comparator is cheaper and provides more benefit or where it provides less benefit and is more costly (Drummond *et al.*, 2015).

In light of the limitations of the ICER, this thesis considers the INB of the health technologies. The net benefit converts the benefits (here assumed to be measured by QALYs), onto a monetary scale. Let

$\mu_{2,I}$ be the QALY for a new intervention,

$\mu_{2,C}$ be the QALY for a control.

The true difference in mean QALY between the two treatment arms is given by

$$\theta_2 = \mu_{2I} - \mu_{2C}. \tag{2.14}$$

Let

$\mu_{3,I}$ be the cost for a new intervention,

$\mu_{3,C}$ be the cost for a control.

The true difference in mean costs between the two treatment arms is given by

$$\theta_3 = \mu_{3I} - \mu_{3C}. \tag{2.15}$$

The net benefit for each treatment is calculated using,

$$\mu_{4,I} = \lambda\mu_{2,I,j} - \mu_{3,I}, \tag{2.16}$$

$$\mu_{4,C} = \lambda\mu_{2,C,j} - \mu_{3,C}, \tag{2.17}$$

where λ is the willingness to pay threshold described in Section 2.7.5. The true INB between the two treatment arms is then given by

$$\theta_4 = \mu_{4,I} - \mu_{4,C}. \tag{2.18}$$

2.7.5 Healthcare Decision Making

For the net benefit approach, a willingness to pay threshold value (λ) is required to aid decision making. The threshold value represents the willingness of the decision maker to pay for a health technology (Drummond *et al.*, 2015). It is the additional amount they are prepared to pay for one more unit of benefit, here an additional QALY. In the UK, the threshold value used by NICE is considered to be £20,000-£30,000 per QALY (National Institute for Health and Care Excellence, 2013a). If the INB is greater than zero the intervention is deemed cost-effective and if it is less than zero it is not deemed cost-effective.

2.7.6 Within Trial Economic Evaluation

It is increasingly common for economic evaluations to be 'piggybacked' onto a randomised controlled trial (O'Brien, 1996; Ramsey *et al.*, 2015). Also known as an economic evaluation alongside a clinical trial (EEACT), information is collected prospectively during the trial giving individual level patient data to assess cost-effectiveness (Drummond *et al.*, 2015). Data relating to the costs and quality of life may be collected during the trial. These data are likely to be correlated with the clinical primary outcome. For example, a drug that lowers blood pressure is likely to have a positive impact on a participant's quality of life as they have fewer problems completing day-to-day tasks. Additionally, they may incur fewer costs as they need to take less medication to control their symptoms and make fewer visits to the hospital or GP. The difference in mean costs and difference in mean QALY can be calculated from the trial data to calculate the INB. This will estimate the cost-effectiveness of the interventions over the duration of the trial.

EEACTs have a number of advantages such as greater control of bias and greater consistency as the cost and quality of life data are collected from the same sample of patients as the clinical outcome assessment (Drummond *et al.*, 2015). They, however, present a number of challenges for a comprehensive assessment of cost-effectiveness including a short term follow-up period and a trial setting which may not truly reflect how the intervention will be delivered in practice (Ramsey *et al.*, 2015). Health economic modelling provides an alternative approach that can be used to supplement EEACTs when evaluating cost-effectiveness in a clinical trial setting.

2.7.7 Health Economic Modelling

A health economic model is used to describe the relationship between a number of inputs including the clinical effectiveness of the intervention, resource use, quality of life data and unit costs. Unlike the within trial analysis, information can be synthesised from a range of sources external to the trial (Drummond *et al.*, 2015). The model can be used to supplement a within trial analysis overcoming some of the limitations, described in Section 2.7.6, such as extrapolating beyond the end of the clinical trial (Buxton *et al.*, 1997). Modelling can also be used when a trial is not conducted to compare health technologies of interest using evidence synthesis (Briggs *et al.*, 1998; Brennan *et al.*, 2000; Welton *et al.*, 2012).

Health economic modelling has a grounding in decision analysis; a systematic approach to decision making under uncertainty to aid decision makers by identifying the optimal health technology (Raiffa *et al.*, 2000). Sculpher *et al.*, 2006 suggest that a modelling approach is the only way to meet all the requirements of a decision maker during an economic evaluation, while Buxton *et al.*, 1997 suggests modelling is 'unavoidable'. A number of approaches can be used in a model-based health economic analysis as summarised by Brennan *et al.*, 2006. This thesis focusses on Markov models, as this is the approach used by the case study described in Chapter 5.

2.7.7.1 Markov Model

A Markov model consists of Markov states that represent each of the states a participant can enter during the disease process under consideration, for example well, unwell and dead health states. These states are mutually exclusive; a participant cannot be in more than one state at a given time (Briggs *et al.*, 1998).

Transition probabilities represent the chance of a participant moving from one state to the next (Briggs *et al.*, 1998). A transition takes place in every cycle of the model. For example, the cycle of a model may be defined to be a month. At the end of each month (cycle), a participant has a given probability of moving, say, from the well state to the unwell state. The choice of transition probability might be chosen based on evidence of the treatment effect of an intervention from a clinical trial (Drummond *et al.*, 2015).

Each of the health states in the model has associated costs and benefits incurred by each participant while they are in that health state for example; in the well state participants may incur a monthly cost of £100 and have a utility of 0.6 per month. The model is then run for a sample of participants. The expected cost (benefits) per cycle is the sum of the costs (or benefits) in each state weighted by the number of participants in the state for that cycle. The overall expected cost and benefits is the sum accumulated over each cycle. This process is repeated assigning different probabilities, costs and utilities to each state of each cycle.

An incremental analysis is performed to establish cost-effectiveness calculating the ICER or the INB as described in Section 2.7.4. This gives an estimate of the long-term cost-effectiveness of the interventions.

2.7.8 Uncertainty Analysis

Once the point estimate of cost-effectiveness is calculated from either a within trial or model-based analysis it is important to understand the uncertainty in the results. Uncertainty exists in the results of a cost-effectiveness analysis, as decisions are made based on imperfect information about the structure of a health economic model, its parameters and the data used to inform the analysis (National Institute for Health and Care Excellence, 2013a). Failure to adequately consider the uncertainty in both within trial and model-based analyses could result in incorrect decisions being made regarding which health technologies should be funded (Welton *et al.*, 2008). This has serious consequences for the public who may not be able to receive the treatment they need as more costly alternatives are being funded instead. Once implemented in practice it can be costly to reverse such decisions (Claxton, 2008).

2.7.8.1 Probabilistic Sensitivity Analysis

Probabilistic sensitivity analysis (PSA) methods facilitate a thorough assessment of the uncertainty in an economic evaluation (Drummond *et al.*, 2015). Model parameters might be informed by trial data or external sources. For example, the distribution for intervention costs might be informed by the cost data collected as part of the trial.

A parametric approach to PSA uses the following steps (Saltelli *et al.*, 2000; Briggs *et al.*, 2006):

1. Assign a distribution to each of the parameters in the health economic model.

2. Randomly sample N_{PSA} values from the assigned distribution. Repeat for all parameters in the model to create a PSA sample.

3. Evaluate the output (usually the INB) for each of the N_{PSA} set of parameters.

4. Use the sample of outputs to summarise its distribution for example the mean INB,

$$\frac{1}{N_{PSA}} \sum_{i=1}^{N_{PSA}} \theta_4^{(i)}. \tag{2.19}$$

Assigning a distribution to each of the input parameters is often based on prior knowledge of the behaviour of the parameter. Briggs *et al.*, 2006 outline some common distributions. All distribution choices should be documented and justified (Drummond *et al.*, 2015).

Alternatively, bootstrapping can be used to create a PSA sample where it is not possible to define the structure of the data using parametric distributions (Efron *et al.*, 1986). The steps for creating a bootstrap PSA sample are:

1. Sample the data with replacement in each arm of the trial N_{boot} times.

2. Evaluate the output (usually the INB) for each of the N_{boot} set of parameters.

3. Use the sample of outputs to summarise its distribution for example the mean INB,

$$\frac{1}{N_{boot}} \sum_{i=1}^{N_{boot}} \theta_4^{(i)}. \tag{2.20}$$

The results of a PSA can be summarised graphically to represent the uncertainty in the economic evaluation by plotting the incremental costs against the incremental QALY in a cost-effectiveness plane. A cost-effectiveness acceptability curve (Van Hout *et al.*, 1994) can also be used to summarise the probability of cost-effectiveness for a range of willingness to pay thresholds (Fenwick *et al.*, 2001).

2.8 Value of Information Analysis

As discussed in Chapter 1, healthcare decision makers have to decide which health technologies to adopt. Often there is uncertainty when making this decision, and a risk that better

outcomes could have been achieved if an alternative decision had been made (Fenwick *et al.*, 2020). When making resource allocation decisions, acquiring additional information is valuable as it can reduce the decision uncertainty (Claxton *et al.*, 1996; Chilcott *et al.*, 2003; Claxton *et al.*, 2006). However, this additional information often comes at a cost, for example randomising participants in a clinical trial incurs a cost, but can help researchers to learn more about an intervention. The methods of value of information analysis (VOIA) provide a framework for quantifying the value of learning more information balancing the benefits with the costs (Raiffa *et al.*, 2000; Fenwick *et al.*, 2020).

VOIA is underpinned by Bayesian decision theory. Hee *et al.*, 2016 describe Bayesian decision theory as a statistical technique that can be used to formally assess decision making under uncertainty. Using this approach, it is possible to choose the optimal decision from a range of potential actions. This choice is informed by considering the consequences of each action under different scenarios (Raiffa *et al.*, 2000). Decision theory can be used to determine whether it is worthwhile conducting further research and where it is, to inform the optimal design of the future research (Chen *et al.*, 2014). In this thesis, VOIA is considered in the context of clinical trials; however, the future research could take other forms such as an observational study.

Once a clinical trial has ended, it is necessary to choose from a set of possible actions denoted by

$$a = \{a_1, a_2, \dots\}. \tag{2.21}$$

Actions might include choosing a new intervention or the existing intervention tested in the trial. The consequences of taking each of these actions can be expressed using a utility function also known as a loss function or gain function. Let

$$U_a \text{ be the utility function.}$$

This function should quantify the decision-maker's preference for each action (Raiffa *et al.*, 2000). This might be a monetary loss or reward if a particular action is taken. This function

depends on some unknown parameters θ and so let

$$U_a(\theta) \text{ be the utility function for each action.}$$

In health economics, the utility function inputs are assigned a monetary value and are primarily based on costs and benefits, as discussed in Section 2.7. The INB defined in Section 2.7.4 is commonly used (Rothery *et al.*, 2020). When choosing whether to conduct further research prior information is synthesised. In the context of clinical trials, this might include data from a pilot trial, an observational study or expert opinion. Adopting the notation of Strong *et al.*, 2015, this information is then used to inform a health economic model, where

θ are the input parameters to the model,

$d = 1, \ldots, D$ are the interventions under consideration,

$NB(d, \theta)$ is the net benefit of each intervention under consideration.

2.8.1 Expected Value of Perfect Information

When there is perfect information (there is no decision uncertainty) the optimal action maximises utility (Raiffa *et al.*, 2000). As highlighted by Rothery *et al.*, 2020, this assumes that the decision maker is risk-neutral. This assumption is made throughout this thesis; however, a decision maker may be more risk averse, preferring to adopt a decision that has a smaller, guaranteed utility over a larger utility with greater uncertainty. The expected value of perfect information (EVPI) considers the scenario where further research would eliminate all decision uncertainty, representing the most that can be gained from further research (Ades *et al.*, 2004; Welton *et al.*, 2008; Griffin *et al.*, 2010). The EVPI can only take positive values (Briggs *et al.*, 2002). It is potentially worthwhile conducting further research if the associated costs are less than the EVPI (Boyd *et al.*, 2010). The EVPI is calculated as follows (Brennan *et al.*, 2007):

1. Choose the optimal intervention based on information currently available,

$$\max_d \left[\mathbb{E}_\theta[\mathrm{NB}(d, \theta)]\right]. \tag{2.22}$$

2. Choose the intervention if perfect information was available (there was no uncertainty),

$$\max_d \left[\mathrm{NB}(d,\theta)\right]. \tag{2.23}$$

3. As the true values of the parameters (θ) in the economic evaluation are unknown it is necessary to average over a sample of their possible values,

$$\mathbb{E}_\theta \left[\max_d \left[\mathrm{NB}(d,\theta)\right]\right]. \tag{2.24}$$

4. The EVPI is the difference between the expected net benefit given perfect information and the expected net benefit given current information

$$\mathrm{EVPI} = \mathbb{E}_\theta \left[\max_d \left[\mathrm{NB}(d,\theta)\right]\right] - \max_d \left[\mathbb{E}_\theta \left[\mathrm{NB}(d,\theta)\right]\right]. \tag{2.25}$$

Multiplying the EVPI by the future population expected to benefit from the information gained by the further research gives the population EVPI (Ades *et al.*, 2004; Griffin *et al.*, 2010).

2.8.2 Expected Value of Sample Information

EVPI estimates the value of eliminating all uncertainty in a decision problem. It may, however, be more feasible to consider the value of reducing some uncertainty (Strong *et al.*, 2015), for example by conducting another clinical trial. Using a process of backwards induction; specifying the actions at the end of the trial and considering the optimal action, it is possible to obtain the optimal trial design for a future clinical trial (Berry, 1995; Raiffa *et al.*, 2000). The expected value of sample information (EVSI) estimates the value of a specific research design that will be used to inform a decision (Strong *et al.*, 2015). For a fixed sample size design, to determine the optimal sample size (n) for a future trial, the possible actions at the end of the proposed trial are defined as,

$$a_1 = \text{choose new intervention},$$

$$a_2 = \text{choose existing intervention}.$$

For each action, an expected utility is calculated. These expected utilities are compared to give the optimal action a given data X.

The EVSI can be calculated as follows (Ades *et al.*, 2004; Strong *et al.*, 2015):

1. Choose the optimal intervention based on the information currently available,

$$\max_d \left[\mathbb{E}_\theta \left[\mathrm{NB}(d,\theta)\right]\right]. \tag{2.26}$$

2. Choose the intervention if data X is collected that is useful for some of the parameters denoted by θ_u,

$$\max_d \left[\mathbb{E}_{\theta|X} \left[\mathrm{NB}(d,\theta)\right]\right]. \tag{2.27}$$

3. As the value of the parameters θ_u given the data (X) is unknown it is necessary to average over a sample of their possible values,

$$\mathbb{E}_X \left[\max_d \left[\mathbb{E}_{\theta|X} \left[\mathrm{NB}(d,\theta)\right]\right]\right]. \tag{2.28}$$

4. The EVSI is the difference between the expected benefit given sample information and the expected net benefit given current information,

$$\mathrm{EVSI} = \mathbb{E}_X \left[\max_d \left[\mathbb{E}_{\theta|X} \left[\mathrm{NB}(d,\theta)\right]\right]\right] - \max_d \left[\mathbb{E}_\theta \left[\mathrm{NB}(d,\theta)\right]\right]. \tag{2.29}$$

Using the tower property,

$$\mathbb{E}_X \left[\mathbb{E}(Y|X)\right] = \mathbb{E}(Y), \tag{2.30}$$

the optimal treatment given current information can (Equation 2.26) be written as

$$\max_d \left[\mathbb{E}_X \left[\mathbb{E}_{\theta|X} \left[\mathrm{NB}(d,\theta)\right]\right]\right], \tag{2.31}$$

and so the EVSI equation can be written as

$$\text{EVSI} = \mathbb{E}_X \left[\max_d \left[\mathbb{E}_{\theta|X} \left[\text{NB}(d, \theta) \right] \right] \right] - \max_d \left[\mathbb{E}_X \left[\mathbb{E}_{\theta|X} \left[\text{NB}(d, \theta) \right] \right] \right] . \tag{2.32}$$

The population EVSI is calculated by multiplying the EVSI by annual prevalence for the population under consideration and the time horizon of the decision problem (Welton *et al.*, 2013).

Approaches to calculate the EVSI include the importance sampling method (Menzies, 2016); the Gaussian approximation method (Jalal *et al.*, 2015; Jalal *et al.*, 2018) and the moment matching method (Heath *et al.*, 2018). This thesis focusses on the non-parametric regression approach described by Strong *et al.*, 2015. This method is chosen as it is a flexible and computationally efficient approach that does not require the existence of conjugate distributions or parametric assumptions to be made (Ades *et al.*, 2004; Brennan *et al.*, 2007; Strong *et al.*, 2015). This approach is described in detail in Chapter 8 where it is used to calculate the EVSI to guide the design of adaptive clinical trials.

2.8.3 Value of Information Analysis in Health Technology Assessments

The use of VOIA has been slowly increasing in health technology assessments in the UK, however they are not routinely used (Mohiuddin *et al.*, 2014; Welton *et al.*, 2015). Steuten *et al.*, 2013 highlight that there are many methodological articles describing the methods with few 'real-time' applications. Barriers to their routine use include their complexity, high computational burden and areas of methodological contention around the effective population and time horizon used in calculations.

To overcome the perceived barriers to their use the Collaborative Network for Value of Information (ConVOI) have been developing resources to aid researchers in their use of these methods. This includes a summary of existing methods for the computation of EVSI and their application in three case studies (Heath *et al.*, 2019; Kunst *et al.*, 2019). An ISPOR task force has also been convened with the aim of developing good practice guidance for using methods of VOIA for use in health technology assessments and to inform the prioritisation of research (ISPOR, 2020; Fenwick *et al.*, 2020).

Additional resources include the RANE - Rapid Assessment of Need for Evidence tool that helps researchers to calculate the value of their research proposals in a timely manner (Glynn *et al.*, 2020). This `R Shiny` application requires inputs from the researcher relating to their primary outcome, interventions and proposed research. The SAVI - Sheffield Accelerated Value of Information `R Shiny` application also aids the calculation of value of information. Researchers can upload their PSA output and the EVPI can be calculated (Strong *et al.*, 2020). SAVI uses the non-parametric regression techniques discussed in Chapter 8.

2.9 Literature that Considers Health Economics in Adaptive Trials

The previous sections introduce the methods of adaptive clinical trials, focussing on group sequential designs and the methods of health economics including economic evaluations and VOIA. The following sections summarise literature that considers how health economics can be used in the context of adaptive clinical trials.

2.9.1 Bayesian Decision Theory in Group Sequential Trials

The Bayesian decision theoretic approach described in Section 2.8 can be applied in the context of adaptive clinical trials. For a group sequential design, at each interim analysis the optimal action depends on the actions taken at all future analyses (Hee *et al.*, 2016). This differs from the fixed sample size design approach where the optimal action at the end of the trial does not depend on any future analyses, as the trial will terminate at this point. The conditionality on future interim analyses in the adaptive design setting adds a number of steps to the backwards induction process previously described.

At the end of the trial, a decision must be made and one of the first two actions in Equation 2.33 taken, as in the fixed sample size design. The optimal action can be found by choosing the action with the maximum expected utility. Then, working backwards, the optimal action at the penultimate analysis is considered. There is now a third action to consider

$$a_1 = \text{stop the trial and choose new intervention}, \tag{2.33}$$

$$a_2 = \text{stop the trial and choose existing intervention}, \tag{2.34}$$

$$a_3 = \text{continue to the next analysis}. \tag{2.35}$$

Again, it is necessary to calculate the expected utility for each of the possible actions. The expected utility for continuing the trial is the average of the future utilities (assuming that the optimal action is taken at each of these analyses) (Lewis *et al.*, 2007). Comparing the expected utilities for the three actions will determine whether the trial should stop or continue at this point. This process is repeated until the start of the trial is reached to determine the optimal design (Berry *et al.*, 1994). The sequential nature of this process is computationally intensive, increasing vastly as the number of interim analyses increases (Berry *et al.*, 1988).

2.9.2 Stopping Rules based on Health Economic Considerations

Berry *et al.*, 1988 apply the sequential Bayesian decision theoretic approach to calculate stopping boundaries for a one-sided sequential clinical trial, where consequences of possible decisions are considered explicitly on a monetary scale. Their approach allows the trial to stop early if there is sufficient evidence to suggest it is futile to continue, but continues as long as the accumulating data show a benefit. The authors take an industry perspective and their utility function includes costs to the company if they attempt to get the drug approved for market. This approach is extended by Cressie *et al.*, 1993; Cressie *et al.*, 1994 to also consider optimisation of the sample size for the trial rather than pre-specifying. Carlin *et al.*, 1998; Kadane *et al.*, 2002; Brockwell *et al.*, 2003; Wathen *et al.*, 2006; Orawo *et al.*, 2009 provide alternative approaches to the backwards induction algorithm to speed up the computation and reduce the computational burden as the number of interim analyses increases.

Willan *et al.*, 2005 calculate the optimal sample size for a trial using VOIA. Willan *et al.*, 2008 extend this work to the multi-stage trial. The utility function is the expected net gain given by the EVSI (based on the INB) minus the total cost, as described in Section 2.7.4. The authors take a societal perspective, though this approach is extended to the industry perspective by Chen *et al.*, 2014. The authors provide solutions to a two-stage design, and note that extending to more interim analyses has a high computational burden.

Mehta *et al.*, 2006 argue that in an industry-based trial, traditional economic and frequentist considerations may not be well aligned when choosing the sample size of a trial. The authors propose a hybrid approach. A frequentist approach is used to determine whether the trial should stop or continue but the magnitude of the sample size in the remainder of the trial, if the trial continues, is based on maximising the expected net present value.

Pertile *et al.*, 2014 use a Bayesian sequential economic evaluation model to approximate an optimal stopping rule based on cost-effectiveness (the costs of the trial as well as the wider societal costs and benefits). The rule considers the cost of carrying out further research against the value gained from having a more accurate estimate of cost-effectiveness. The authors build on the work of Chernoff *et al.*, 1981 and consider the sequential problem as 'real options' where the decision maker has the right but not the obligation to stop the trial early if there is sufficient evidence to suggest one treatment is superior to the alternative. The problem is solved using backwards induction and can be applied to a number of interim analyses overcoming the key limitations of other approaches.

Chick *et al.*, 2017 extend the Pertile *et al.*, 2014 method to allow a delay in observing the primary outcome, as is common in many clinical trials. The approach of Chick *et al.*, 2017 can also be used to determine whether it is worthwhile conducting further research and to guide whether the trial should be a fixed sample size design or sequential design. Forster *et al.*, 2019 apply the Chick *et al.*, 2017 method retrospectively to the ProFHER randomised trial that investigated the use of surgery versus non-surgical intervention for patients with a displaced proximal humeral fracture (Handoll *et al.*, 2015). The authors found that the trial's expected sample size could have been reduced by up to 40%, saving approximately 15% of the trial budget if cost-effectiveness had been considered at interim analyses throughout the trial.

Kouvelis *et al.*, 2017 adopt a similar approach to Pertile *et al.*, 2014; Chick *et al.*, 2017 but from an industry perspective to determine when and how many centres should be opened during a trial with a non-linear rate of recruitment. They use a utility function based on net present value of a drug, as in the approach by Mehta *et al.*, 2006. The authors use the frequentist stopping rules to determine when the trial should stop at an interim analysis, however they use their Bayesian framework to determine the number of centres and to adjust the recruitment rate at each interim.

Thijssen *et al.*, 2017 use a sequential hypothesis testing, rather than sequential estimation used by Pertile *et al.*, 2014; Chick *et al.*, 2017, to derive an optimal stopping rule in a health technology assessment.

2.9.3 Impact of an Adaptive Design on a Health Economic Analysis

Marschner *et al.*, 2019 discuss the underestimation of treatment effects in sequential clinical trials when they do not stop early for benefit. The authors discuss the importance of an unbiased estimate of the treatment effect for cost-effectiveness analyses. They highlight, using a reanalysis of the GUSTO study (Gusto Investigators, 1993; Mark *et al.*, 1995) that the treatment effect may have been underestimated and the experimental therapy appeared less cost-effective than it actually was. There is no discussion in this paper about the bias in secondary outcomes, and the assessment of potential bias in the cost-effectiveness analysis is a simplistic calculation that does not consider a health economic model.

Rothery *et al.*, 2020 have published guidance for methodologists on conducting value of information calculations. In this article, the authors highlight that EVSI calculations for adaptive designs have received little attention to date in the literature. This adds further impetus for the extension of existing methods to the adaptive design setting. The authors also suggest that when simulating datasets as part of the EVSI calculation methods it is important consider any potential biases in the data. Whilst not specifically mentioning the potential bias from adaptive designs, this acknowledges the importance of understanding the impact of the designs being considered on the data. Both of these issues are explored in Chapter 8.

In summary, the Bayesian decision theoretic approach has been extended to the adaptive design setting by these authors. The approach has been considered from both the societal and industry perspectives and with methods accounting for the preference for frequentist approach in clinical trials currently. However, many of the identified methods discuss that they are limited by their high computational burden and there is little discussion of how the methods can be applied in the real-world setting. This makes it challenging for researchers to apply these approaches in their clinical trials.

2.10 Chapter Summary

Decision makers rely on high quality, accurate information about the clinical and cost-effectiveness of a health technology. Central to the evidence base for healthcare decision making are clinical trials. Increasingly adaptive clinical trial designs are being used to collect information about

clinical and health economic outcomes. This information is potentially being presented to decision makers with little consideration of the impact on the health economic analysis. Furthermore, there is potential to increase the efficiency of an adaptive design by maximising the use of health economic data both prior to the start and at an interim analysis of a trial. While the methods of adaptive trials and health economics are well developed and ultimately have the same aim of increasing the efficiency of health research there are important issues that might arise if appropriate care is not given when applying the methods in health technology assessment.

Chapter 3 explores how these methods have been considered together in practice. Chapter 4 will discuss the barriers of using the methods of adaptive designs and health economics outlined in this chapter in practice with key stakeholders. Chapter 5 illustrates the key health economic concepts discussed in the context of the CACTUS case study. Chapter 6 extends the existing theory for adjustment of analyses following an adaptive design to the context of a health economic analysis following an adaptive design. Chapter 7 operationalises these adjustments in a simulation study and Chapter 8 extends the theory of EVSI to guide the design of adaptive clinical trials.

Chapter 3

Review of the Current Use of Health Economics in Adaptive Trials

3.1 Introduction

Chapter 2 introduces the key concepts of adaptive clinical trials and health economic analysis in the context of efficient healthcare decision making. A number of reviews have highlighted how the current use of adaptive clinical trial designs is increasing in practice (Hatfield *et al.*, 2016; Bothwell *et al.*, 2018). Adaptive designs have also received a lot of attention in the literature (Morgan *et al.*, 2014a), however, research to date has focused on using clinical outcomes to design the trial and when analysing interim data. The use of VOIA has been slowly increasing in health technology assessments in the UK, however they are not routinely used (Mohiuddin *et al.*, 2014; Welton *et al.*, 2015).

The available research literature on the use of health economics in the design and analysis of adaptive clinical trials is less common. As summarised in Chapter 2, the idea of using health economics in the design and analysis of adaptive clinical trials has been discussed by some authors, however, this has largely been in a theoretical context. It is unclear whether any of these methods have been used in the real-world setting.

3.2 Chapter Aims

This chapter summarises the current use of both health economics and adaptive designs together in the design, analysis and reporting of 'real-life' clinical trials . This builds on the literature summarised in Chapter 2 to achieve the first research aim of this thesis, reviewing the current use of health economics in the design and analysis of adaptive clinical trials in practice and the research literature. These findings are used to inform the work in Chapters 4 to 8, with recommendations made for exploration in the thesis and more broadly for the research community in Chapter 9.

This review of current practice has been published in *Value in Health* (Flight *et al.*, 2019a) and was produced in collaboration with three undergraduate medical students working under my supervision as part of their second year six-week research attachment.

3.3 Review Objectives

The use of health economics in the design and analysis of adaptive trials has the potential to increase the efficiency of healthcare research and delivery. Chapter 2 summarises methodological articles on this topic dating back to 1988. Before extending this work further it is important to understand how the methods are currently used together (if at all) in practice. This will identify potential barriers to the use of the current methods in the real-world setting and identify areas of importance for further development.

The primary aim of this review is to establish how health economic outcomes are utilised in adaptive trials in the

- design - such as secondary outcomes or informing sample size using VOIA (described in Section 2.8).

- analysis – such as whether adjustments are used to account for the adaptive nature of the trial (described in Section 2.5).

- reporting - by applying elements of established reporting guidelines.

To supplement these findings, guidance documents for the conduct of health economic analyses and adaptive designs are reviewed to identify any guidance for researchers relating to the use of these methods together in practice.

3.4 Methods

3.4.1 Data Sources and Search Strategy

To identify a diverse and representative sample of adaptive clinical trials six sources were used:

1. `clinicaltrials.gov` – to identify trials with an adaptive design registered from 2011 onwards (accessed on 19.08.2016) (clinicaltrials.gov, 2017).

2. Peer reviewed journals via MEDLINE, EMBASE, Cochrane Library and Web of Science. Any articles reporting a clinical trial with an adaptive design were included in the current review.

3. A review by Hatfield *et al.*, 2016 of 158 registered clinical trials obtained from `clinicaltrials.gov` 2000 to 2014, the National Institute for Health Research (NIHR) register and contacted experts for known adaptive designs.

4. A review by Stevely *et al.*, 2015 of the reporting of 68 published clinical trials using a group sequential design identified on MEDLINE for years 2001 to 2014.

5. Health Technology Assessment (HTA) Journal - this journal publishes research on the effectiveness, costs and broader impact of healthcare technologies used on the NHS including the results of clinical trials funded by the NIHR HTA Programme.

6. Known adaptive designs - identified by contacting experts in statistics and health economics – known to the research team and via emails sent to the online forums AllStat (28.09.2016) and HealthEconAll (29.09.2016).

To identify articles on clinicaltrials.gov and in the HTA journal the search strategy implemented by Hatfield *et al.*, 2016 (see Appendix A) was adapted. The strategy aimed to identify trials with an adaptive design using common words such as 'adaptive', 'sequential' and 'interim'. Development and validation of the strategy is reported in Hatfield *et al.*, 2016.

Country specific guidance documents for the economic evaluation of clinical trials were identified using the ISPOR webpages in addition to known guidance documents for health economic analyses and adaptive designs.

3.4.2 Inclusion Criteria

Five reviewers Laura Flight, Kian Patel, Fahid Arshad, Rachel Barnsley, Steven Julious identified articles which met criteria adapted from Hatfield *et al.*, 2016:

1. Trial documentation available in English.
2. Phase III clinical trial.
3. Trial investigating an intervention(s) on humans with a comparator.
4. Registered or published before 01.08.2016.
5. Multiple treatment arms.
6. An adaptive design clinical trial - defined to be a trial with any pre-planned early examination of the data, including any monitoring of the data by a data monitoring and ethics committee (DMEC) where it is clear there had been or there is a planned formal analysis of the data.
7. A planned health economic analysis.

The chief investigators for clinical trials with a pre-planned adaptive design but with no clear health economic analysis were contacted via email to ask whether any health economic analyses were carried out.

Guidance documents were required to outline guidance for the use of health economic methods in the design or analysis of phase III adaptive trials.

3.4.3 Data Extraction

For trials that met the inclusion criteria, information was extracted relating to their characteristics, design, analysis and reporting. A data extraction sheet was developed using items from five key checklists/quality assessment processes in the areas of clinical trials, HTA and cost-effectiveness. These included the CONSORT (Consolidated Standards of Reporting Trials) statement and the CHEERS (Consolidated Health Economic Evaluation Reporting Standards) checklist (Moher *et al.*, 2001; Shemilt *et al.*, 2006; SAACTD Workshop Committee, 2009; Detry *et al.*, 2012; Husereau *et al.*, 2013).

The CHEERS checklist was developed by the ISPOR Health Economic Evaluation Publication guidelines - CHEERS Good Reporting Practices Task Force. The main aim of the guideline is to consolidate and update existing guidelines for the optimal reporting of health economic evaluations (Husereau *et al.*, 2013). The CONSORT guidelines were first developed in 1996 (Begg *et al.*, 1996) and revised in 2001 (Moher *et al.*, 2001) and 2010 (Schulz *et al.*, 2010). Their main aim is to address the lack of adequate reporting of the results of two-armed, parallel group, individually randomised clinical trials.

Trial documentation was identified from trial registries, protocols and journal publications (identified via MEDLINE and Google Scholar). When a large number of results were returned by a database, filters were used to identify the most relevant publications. The main trial paper or the HTA monograph was used to assess the reporting of the trial results.

Inclusion criteria were applied and two reviewers extracted data independently. Any discrepancies were resolved by a group discussion. Some of the information extracted, such as the level of detail reported about health economics in the protocol and reporting questions required a subjective decision. For these questions, I reviewed data extracted by the team and made a final decision to maintain consistency. Any subsequent changes or additional information were documented.

Guideline documents were searched for key adaptive designs terms ('adaptive' and 'interim'). If there was discussion of adaptive designs the article was included in the full review.

3.4.4 Outline of Analysis

A descriptive analysis was undertaken to provide an overview of how health economics was used in the comprehensive sample of adaptive designs and relevant guidance documents. Continuous variables were summarised using their mean and standard deviation (SD). Categorical variables were summarised using counts and percentages.

3.5 Results

A total of 553 articles about trials and 48 guidance documents were identified (see Figure 3.1). Of the trial articles, 278 were identified on `clinicaltrials.gov` and 159 were registered before 2011. I applied the inclusion criteria to a scoping sample of 79/159 `clinicaltrials.gov` articles from 2010 or earlier. In this subsample, only one trial was found to meet the inclusion criteria. It was decided to omit the 159 `clinicaltrials.gov` trials registered before 2011 from the review. This decision is justified by the work of Hatfield *et al.*, 2016 who found adaptive designs were increasingly used between 2012/2013. Given the small number of articles identified in the subsample, it is unlikely that many trials were missed.

Thirty-seven trials met the inclusion criteria and were subject to full data extraction.

Forty-eight guidance documents were identified that outlined guidance for health economic evaluation or the use of adaptive designs around the world. Forty-four were identified from the ISPOR Pharmacoeconomics Guidelines Around the World Webpages (ISPOR, 2016), two were identified from relevant ISPOR good practice research documents (Briggs *et al.*, 2012; Ramsey *et al.*, 2015) and two were known regulatory guidance documents relating to adaptive design clinical trials (European Medicines Agency, 2007; U.S. Food and Drug Administration, 2019). Four guidance documents were included in the review.

3.5.1 Trial Characteristics

All trial characteristics are summarised in Table 3.1. The types of adaptive designs identified are summarised in Table 3.2. One trial did not provide sufficient information to assess the methods used. It was common for multiple adaptations to be implemented in a single trial. Table 3.2 includes all the adaptations.

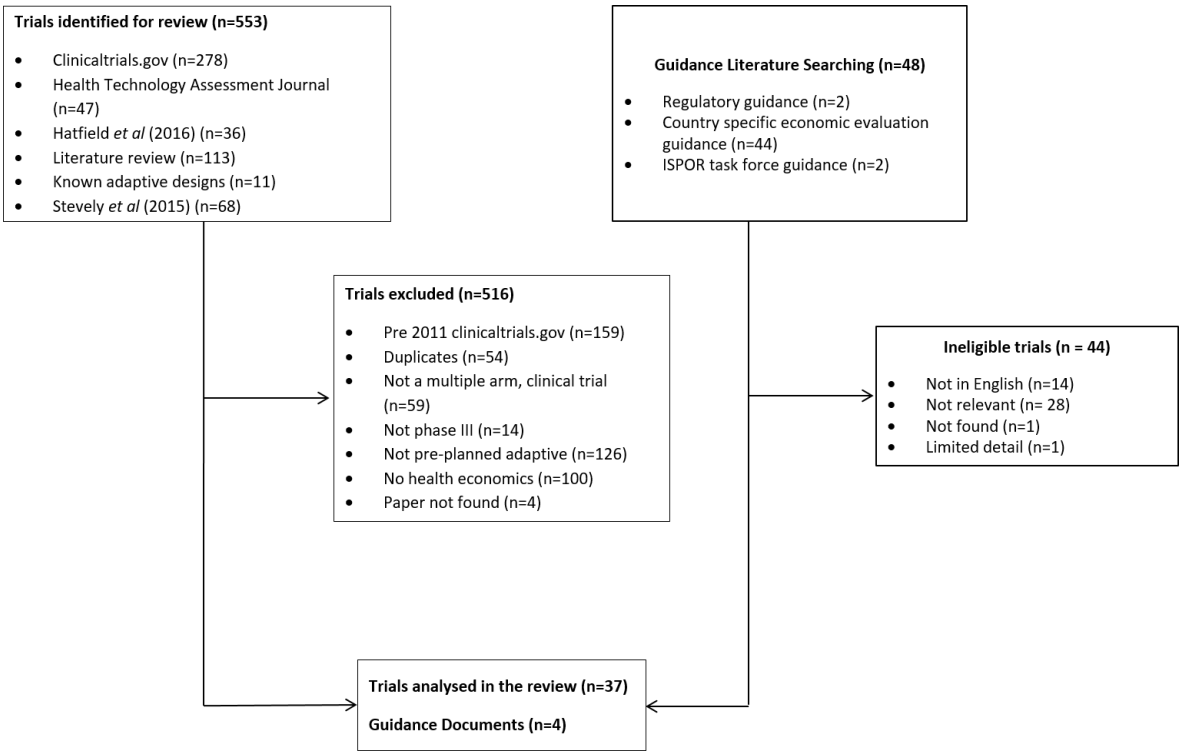


FIGURE 3.1: Summary of articles identified in the review of clinical trials with an adaptive design and health economic analysis and guidance documents for health economic analyses or using adaptive designs

The rationale for choosing an adaptive design was clear in 38% ($^{14}/_{37}$) of trials. The most common rationale was to check the uncertain assumptions made at the design stage of the trial. For example, the EVIDENCE study (North *et al.*, 2011) identified that there was a lack of information to inform their sample size. They pre-planned an interim analysis to re-estimate the required number of patients to achieve sufficient statistical power.

3.5.2 Health Economics in the Design of Adaptive Trials

A trial protocol was identified in 73% ($^{27}/_{37}$) of the trials. The pre-specification of health economic analyses was limited with 41% ($^{11}/_{27}$) of trials not providing any detail and 59% ($^{16}/_{27}$) providing only limited detail in the trial protocol, such as a paragraph outlining that a Markov model would be used for the health economic analysis but little further elaboration. Fifteen percent ($^{4}/_{27}$) of trials included a full analysis plan for their proposed statistical analyses in their protocol and 85% ($^{23}/_{27}$) provided limited detail. Fifty-two percent ($^{14}/_{27}$) of the trials reported limited detail relating to both their health economic and statistical analyses. Some trials may have reported a full statistical analysis or health economic analysis plan in a separate document not appended to the protocol, which has not captured here.

Characteristic		n	%
State	Ongoing	8	22
	Recruiting	7	19
	Completed	20	54
	Not Clear	2	5
Country of Chief Investigator	Canada and USA	9	24
	China	1	3
	Europe (not including UK)	9	24
	UK	17	46
Funder	Private	6	16
	Public	25	68
	Private and Public	2	5
	Not Clear	4	11
Experimental Treatment	Medicinal	17	46
	Device	3	8
	Educational	2	5
	Psychological	0	0
	Complex Intervention	0	0
	Other	15	41
Comparator	Active	35	95
	Placebo	2	5
Therapeutic Area	Oncology	11	30
	Cardiology	5	14
	Vascular and Haematology	4	11
	Spinal	2	5
	Other	15	41

TABLE 3.1: Summary of the characteristics of trials with an adaptive design and health economic analysis that met the inclusion criteria ($n = 37$)

Type of Adaptation	n	%
Adaptive Randomisation (Section 2.3.3.2)	1	2
Drop the Loser (Section 2.3.3.3)	2	3
Efficacy (Section 2.4)	3	5
Efficacy and Futility (Section 2.4)	4	7
Efficacy and Safety (Section 2.4)	9	15
Futility (Section 2.4)	5	8
Futility and Non-Inferiority (Section 2.4)	1	2
Futility and Safety (Section 2.4)	1	2
Futility and Safety and Efficacy (Section 2.4)	5	8
Interim*	4	7
Internal Pilot	12	20
Sample Size Re-estimation (Section 2.3.3.1)	7	12
NA	1	2

TABLE 3.2: Summary of the adaptations used in a trial. All adaptations discussed in a particular trial are included. Therefore, percentages are expressed in terms of the total number of adaptations. *Interim denotes a trial where an interim examination of the data was mentioned but it was not possible to ascertain the motivation or methods used.

Characteristic		n	%
Were health economic outcomes considered in the design?	Yes	3	8
	No	32	86
	Not Clear	2	5
Were any health economic outcomes a primary outcome?	Yes	1	3
	No	34	92
	Not Clear	2	5
Were health economic outcomes considered in the sample size calculation?	Yes	0	0
	No	34	92
	Not Clear	3	8
Was value of information analysis considered in the design?	Yes	2	5
	No	32	86
	Not Clear	3	8

TABLE 3.3: Summary of how health economic outcomes were considered in the design of the adaptive design clinical trials in the review ($n = 37$). Two trials were conference abstracts and the third did not have a protocol available

The role of health economic outcomes in the design of the adaptive trials was limited, as summarised in Table 3.3. Three trials, namely 'GDHT' (Bartha *et al.*, 2012; Bartha *et al.*, 2013a; Bartha *et al.*, 2013b), 'PRESSURE-2' (Brown *et al.*, 2016) and 'OPTIMA' (Hall *et al.*, 2012; Hall *et al.*, 2015; Stein *et al.*, 2016; Stein, 2016) were considered to have used health economics in their design. The OPTIMA trial listed a health economic outcome as a primary outcome however; this was not considered in the sample size calculation. The Persephone study listed costs and quality of life outcomes in relation to their study design but it was not clear what role these outcomes took (Earl, 2009).

OPTIMA considered VOIA to inform their design (Stein, 2016) and PRESSURE-2 planned to include an EVSI analysis at an interim analysis (Brown *et al.*, 2016). This is discussed in more detail in Section 3.6.

3.5.3 Health Economics in the Analysis Conducted

Information about the analysis of the adaptive designs was extracted from trials with results available ($51\%(19/37)$). The remaining 18 trials did not have any results at the time of data extraction and so were not included.

Of the 19 trials, those thought to require adjustment to their analysis to allow for the adaptive nature of the design, specifically trials using a group sequential design, were identified. The reporting of the methods was not always explicitly clear and a judgement was made about the need for adjustment. Where a trial simply stated that there was an interim analysis (four trials), it was assumed that this was a group sequential design, as this is the most common type of adaptation used in practice as identified by Hatfield *et al.*, 2016 where 78% of phase III trials used a group sequential design. Two of the 19 trials were not thought to have used sequential methods or required adjustment; one used adaptive randomisation and the second a sample size re-estimation.

In the remaining $^{17}/_{19}$ trials, where an adjustment of the point estimate for the primary and any correlated secondary outcomes was thought appropriate, there was no clear indication that the primary outcome was adjusted. Additionally, none of the 17 trials indicated whether they used adjusted primary or secondary outcomes in their health economic analysis.

3.5.4 Health Economic Analysis Using Interim Data

Tessitore *et al.*, 2014 reported the interim analysis of a randomised trial on the elective repair of subclinical stenosis (ISRTC69115386). The authors calculated the cost-effectiveness of the intervention using the interim data. They concluded that given a large clinical benefit of the intervention and little difference in cost it was unethical to continue the trial. There was no indication as to whether the treatment effect estimate was adjusted for the interim analysis.

3.5.5 Health Economics in the Reporting of Adaptive Trials

Of the 19 trials with results available one trial only provided results on `clinicaltrials.gov` and a second had information in a short conference abstract; therefore, it was not possible to assess their reporting. Table 3.4 summarises how well the remaining trials reported their results.

There was little consideration for how the adaptive design might affect the clinical and health economic analyses. One trial discussed how stopping at an interim analysis resulted in less data and therefore greater uncertainty in the health economic analysis. The authors also acknowledged that this would likely impact on the generalisability of their results. It was only

clear for one trial that prior interim results were available and were discussed at a conference and disseminated to trial participants.

3.5.6 Guidance Documents

There is no discussion of health economics or the evaluation of cost-effectiveness in the Food and Drug Administration (FDA) (U.S. Food and Drug Administration, 2019) or European Medicines Agency (EMA) (European Medicines Agency, 2007) documents on adaptive designs. However, these documents outline that care is needed when choosing an adaptive design given their potential to impact on secondary outcomes, such as biased estimation and smaller sample sizes. These points are relevant to health economic outcomes collected as part of an adaptive design, but as health economics is not explicitly mentioned this may be overlooked by researchers.

Four of the country specific guidelines for economic evaluation referenced adaptive designs (Academy of Managed Care Pharmacy, 2016; Australian Government, Department of Health, 2016; Canadian Agency for Drugs and Technologies in Health, 2017; Scottish Medicines Consortium, 2019). The relevant sections are summarised in Table 3.5. Each of these guidance documents emphasises the importance of understanding the type of trial design used to generate the data for the health economic analysis and the potential for this to impact on the results. However, little guidance is provided on the exact issues that might arise or solutions to overcome any biases. This highlights the need for more specific guidance on the use of data from adaptive clinical trial designs in health economic analyses.

3.6 Exemplars of Health Economics used in Adaptive Trials

The following trials highlight the use of adaptive designs and health economics as part of the clinical trial process. None of the exemplars utilise health economics and adaptive designs to their full potential or consider the impact that using data from an adaptive design might have on their analysis. However, given the limited research and awareness in this area these trials illustrate the potential use of the methods.

Characteristic		n	**%**
Was the Trial Identified as an Adaptive Design in the Title?	Yes	1	3
	No	16	43
	NA	20	54
Was the Economic Evaluation or More Specific Identified in the Title?	Yes	5	14
	No	12	32
	NA	20	54
Was the Economic Evaluation or More Specific Identified in the Abstract?	Yes	8	22
	No	9	24
	NA	20	54
Were Health Economic Outcomes Discussed on the Main Trial Paper?	Yes	9	24
	No	8	22
	NA	20	54
Discussion of How the Adaptive Design Might Have Impacted on the Health Economic Analysis	Yes	1	3
	No	14	38
	Trial stopped before AD	2	5
	NA	20	54
Was the Potential for Bias in the Results Discussed?	Yes	0	0
	No	15	41
	Trial stopped before AD	2	5
	NA	20	54
Was the Generalisability of the Findings from the Adaptive Design Discussed?	Yes	1	3
	No	14	38
	Trial stopped before AD	2	5
	NA	20	54
Were Lessons Learnt from Using the Adaptive Design Discussed?	Yes	0	0
	No	15	41
	Trial stopped before AD	2	5
	NA	20	54
Were Prior Interim Results Provided or Discussed?	Yes	1	3
	No	11	30
	Trial stopped before AD	2	5
	Trial stopped at first interim	3	8
	NA	20	54

TABLE 3.4: Summary of the level of reporting in the adaptive design (AD) clinical trials in the review. ($n = 37$)

Country	**Guideline Document**	**Extract**
Australia	Guidelines for preparing a submission to Committee Version 5.0	alternative trial designs may be acceptable (eg [...] trials with a randomised adaptive design). [...] Where a submission is based on such a trial, risk of bias can be addressed as for randomised trials. (page 29). Justify and discuss any early stopping of a trial or reliance on interim analysis in the interpretation of the results. (page 40)
Scotland	Guidance to Manufacturers for Completion of New Product Assessment Form (NPAF)	Where data for a single study have been taken from more than one source this should be made clear. Examples of this include: [...] additional analyses (e.g. interim or post-hoc). (page 15)
United States of America	A format for submission of clinical and economic evidence for pharmaceuticals in support of formulary consideration.	...manufacturers should include studies that generate evidence about clinical outcomes which may include, for example, randomized controlled trials (Phase 2, 3, 4),[...] and studies that use adaptive trial designs. See FDA guidance for adaptive trial designs. (page 29)
Canada	Guidelines for the Economic Evaluation of Health Technologies: Canada 4th Edition	Adaptive trial designs are the most commonly used novel approaches. [...] the designs of clinical trials should be assessed to identify any features that may affect the use or interpretation of the data (considering fitness for purpose, credibility and consistency). Given the evolving nature of such study designs, it is recommended that where other data sources exist, these should be used to inform parameter estimates in the reference case and the impact of information from novel designs considered in scenario analyses.

TABLE 3.5: Summary of the discussion of adaptive design clinical trials in worldwide health economic guidance documents

3.6.1 Exemplar 1: OPTIMA trial

The optimal personalised treatment of breast cancer using multi-parameter analysis (OPTIMA) trial illustrates how VOIA can be used to inform the design and conduct of an adaptive design (Hall *et al.*, 2012; Hall *et al.*, 2015; Stein *et al.*, 2016; Stein, 2016; Hall *et al.*, 2017). The OPTIMA trial is designed to explore the personalised treatment of breast cancer by using laboratory tests to determine who should receive chemotherapy. This trial is an adaptive design with two interim analyses assessing futility and non-inferiority.

The OPTIMA trial was preceded by a cost-utility analysis (Hall *et al.*, 2012) using health economic modelling to determine the cost-effectiveness of genomic test-directed chemotherapy and chemotherapy for all patients. A VOIA was used to inform the value of conducting further research and highlight areas that required further work. This showed substantial uncertainty in the cost-effectiveness results and hence the OPTIMA prelim study was planned.

The OPTIMA prelim trial (ISRCTN42400492) was used to assess the feasibility of a larger trial. One of the main objectives was to evaluate the performance and cost-effectiveness of different laboratory tests to determine what would be evaluated in the main trial. The analysis highlighted considerable uncertainty in the cost-effectiveness of all the tests (Hall *et al.*, 2017). A VOIA suggested that there was high value in conducting further research. The EVSI calculation for a large trial with 2,500 patients per arm, comparing chemotherapy for all patients with chemotherapy directed by one of the tests under consideration was £8,397,961 for the 10-year incident population, suggesting value in carrying out this research design. However, this calculation does not appear to have informed the sample size calculation for the OPTIMA trial (Stein *et al.*, 2016).

3.6.2 Exemplar 2: GDHT trial

The Goal Directed Haemodynamic Therapy (GDHT) study for patients with proximal femoral fracture included an interim analysis on efficacy and safety after 100 (from a planned 460) patients had been recruited. At this point it was decided to continue with the trial, however after a further 50 patients were enrolled over the following 12 months the decision was made to stop the study (Bartha *et al.*, 2012; Bartha *et al.*, 2013a; Bartha *et al.*, 2013b).

Prior to the trial, the authors developed a probabilistic decision analytic cost-effectiveness model. The pre-trial modelling highlighted that postoperative complications heavily influence the cost-effectiveness of GDHT and so the trial was designed to assess the risk of postoperative complications and their influence on quality of life. The pre-trial model was then used for a VOIA using interim data. While this analysis was not pre-planned the authors highlight the potential for these methods to be used in this way. This study is a useful case study for using VOIA during an adaptive design to inform whether it is cost-effective to continue with a trial based health economic grounds.

3.6.3 Exemplar 3: PRESSURE 2 trial

The PRESSURE-2 trial aims to determine the clinical and cost-effectiveness of high specification foam and alternating pressure mattresses for the prevention of pressure ulcers (Brown *et al.*, 2016). As at the time of data extraction, there were no results available only the use of health economics in the design of the trial was considered. PRESSURE-2 considers the cost-effectiveness of the research using interim data as part of an EVSI calculation. This analysis used the interim data to determine whether it is cost-effective to continue with the trial from an NHS decision makers' perspective. This illustrates how health economic outcomes can be included as part of the interim analysis of an adaptive design.

3.7 Discussion

3.7.1 Summary of Key Findings from this Chapter

In a review of 37 clinical trials with an adaptive design and health economic analysis, only three trials utilised health economic outcomes in the design and none of the trials seemed to appropriately adjust the health economic outcomes to account for biases introduced by the adaptive study design. One study utilised health economic outcomes at the interim analysis in the 19 trials with results. The reporting of health economic results was suboptimal for all trials. While VOIA was considered by two trials, none discussed using value of information to inform their adaptive design, such as the optimal number of interim analyses. This highlights a missed opportunity to potentially increase the efficiency of adaptive designs by using health economic outcomes to identify the most cost-effective design.

Four out of the 48 country specific guidelines for economic evaluation referenced adaptive designs. Although these documents emphasise the importance of understanding the type of trial design used to generate data for the health economic analysis and the potential for this to impact on the results, little guidance is provided on the exact issues that might arise or propose adjustments.

3.7.2 How this fits with Existing Literature

In this review, the majority of trials were UK based and 68% ($^{25}/_{37}$) of trials were publicly funded. A previous review by Hatfield *et al.*, 2016 noted that adaptive designs were predominantly conducted in the USA and Canada. Additionally, Hatfield *et al.*, 2016 and Stevely *et al.*, 2015 found that industry funded trials were more common ($^{101}/_{143}$) and ($^{35}/_{68}$) respectively. This contrast could reflect the important role of health economic analyses in healthcare decision making in the public, UK setting (National Institute for Health and Care Excellence, 2013a).

Since completing the review, the results of the PRESSURE-2 trial have been published. Nixon *et al.*, 2019 reported that the first planned interim analysis was due to be conducted after 300 events, noting that this would give the minimum number of events required for the economic analysis. However, after approximately 12-months the observed recruitment rate was slower than anticipated. An unplanned VOIA was conducted at the request of the funder (NIHR HTA Programme) to inform whether the trial should continue. The trial was continued with a second VOIA conducted approximately 28-months into the trial. The use of VOIA to inform the continuation of the trial highlights the potential for the cost-effectiveness of the research to be used during adaptive clinical trials. The authors provide a detailed description of the methods and results of each of these analyses in the trial report allowing other researchers to understand and critique their approach (Nixon *et al.*, 2019).

As described in Chapter 2, Bayesian decision theory has been extended to the design of sequential trials by a number of authors (Berry *et al.*, 1988; Mehta *et al.*, 2006; Willan, 2008; Pertile *et al.*, 2014; Chick *et al.*, 2017). However, the review of current practice demonstrates these methods are not being used in practice. This could be explained by the computational complexity of the approach, especially as the number of interim analyses increases. Only one extended

case study was identified (Forster *et al.*, 2019) highlighting a lack of practical resources for researchers interested in applying these methods. More recent methods may not have had chance to be implemented in real-world trials, hence why there were not identified in the review of current practice.

3.7.3 Considerations for Practice

Proposed health economic analyses should be outlined in a detailed Health Economic and Decision Modelling Analysis Plan (HEDMAP) before the start of an adaptive design. This will be crucial in maintaining the validity and integrity of adaptive designs that use health economics, with analysis plans including a description of the monitoring and adaptation plan, as well as pre-specification of methods used at interim analyses (Thorn *et al.*, 2017; U.S. Food and Drug Administration, 2019)

To improve the reporting of the health economic analysis of adaptive designs the CHEERS checklist should consider the specific issues relating to adaptive designs. It is recommended that these guidelines should be extended to this setting to improve the reporting of clinical trials with an adaptive design and health economic analysis.

To ensure that health economic analyses following an adaptive design are not compromised it is important that guidance documents for the economic evaluation of clinical trials, such as the NICE Guide to Technology appraisals (National Institute for Health and Care Excellence, 2013a) are updated to highlight the potential impact of the design. Further support could be provided to research teams though technical support documents developed by the Decision Support Unit (National Institute for Health and Care Excellence, 2020).

3.7.4 Strengths and Limitations

The sample of adaptive designs reviewed were identified from a range of sources; however, this will not include every adaptive design. Trials identified using `clinicaltrials.gov` prior to 2011 were excluded. Given the limited use of adaptive designs before this time it is not felt that this exclusion will affect the representativeness of the sample. The level of detail provided on `clinicaltrials.gov` can vary which meant many trials were excluded, as there was insufficient information to determine whether it was an adaptive design. This issue was also faced by Hatfield *et al.*, 2016.

The trials where a health economic analysis was planned after the adaptive design was proposed were not captured. There are likely to be consequences for these types of analyses if they use data from adaptive trials too and this should be considered when planning such analyses. A number of the main trial reports did not give details about health economic analyses conducted and so it was difficult to ascertain the methods used and whether any adjustments were made for the adaptive nature of the trial.

3.7.5 Considerations for the Thesis

The review of clinical trials found many authors were not adjusting their analysis to allow for the adaptive nature of the trial. A similar finding was reported by Stevely *et al.*, 2015 who found that the bias correction (for clinical effectiveness outcomes) for early stopping was only reported in 7% (3/46) of group sequential designs. There was also a lack of methods literature on how existing adjustment methods can be applied in the health economic context. In Chapter 6 the existing theory of Chapter 2 is extended to fill this research gap. A simulation study in Chapter 7 then considers how this theory can be operationalised in practice.

It is clear that health economics is a secondary consideration in many trials and is rarely used to inform the design or adaptive decision making at an interim analysis. In Chapter 8 the use of EVSI, calculated using non-parametric regression, to guide the design of adaptive trials is explored.

3.8 Chapter Summary

A review of clinical trials with an adaptive design and a health economic analysis has shown only a small proportion of adaptive trials are considering the use of health economics in practice and none of the trials are giving appropriate consideration to the potential for an adaptive design to affect a health economic analysis. This chapter, in conjunction with Chapter 2, answers the first research aim of the thesis; to review the current use of health economics in the design and analysis of adaptive clinical trials in the research literature and in practice

While there is work in the methods literature around the use of decision theory to design sequential trials the review in this chapter has shown that this work has not yet translated into practice. It is unclear what the barriers to the routine use of these approaches might be. In

Chapter 4 researchers, decision makers and members of the public are interviewed to ask their views on the advantages and barriers to using health economics in the design and analysis of adaptive designs to inform the direction of methods development in this area.

Chapter 4

Understanding Stakeholder Views towards the Use of Health Economics in the Design and Analysis of Adaptive Clinical Trials

4.1 Introduction

As highlighted in Chapter 3, the use of health economics in the design and analysis of adaptive clinical trials has been limited in practice. There exists an opportunity to bring together these two areas to increase the efficiency of HTAs, however, for these methods to be useful and applied in practice it is important to think beyond the methodological issues and to understand stakeholder views towards the use of health economics in the design and analysis of adaptive clinical trials.

4.2 Chapter Aims

This chapter reports findings from a qualitative study that aimed to establish the views of stakeholders (researchers, decision makers and the public) in the HTA process on the use of health economics in the design and analysis of adaptive clinical trials to answer the second research aim. While quantitative methods can be used to develop the methodology to address statistical issues, qualitative methods explore the ethical and practical considerations too. This

chapter first describes the methods used in the study, summarises the results and uses the findings to inform recommendations for the wider research community and further methods development in the thesis.

The research in this chapter was supported by a qualitative expert (Professor Daniel Hind (DH)), advising on the planning, providing scientific review and co-coding six transcripts in the development of the final theoretical framework. The findings of this chapter are published in the journal *Trials* (Flight *et al.*, 2020) and are reported in line with the consolidated criteria for reporting qualitative research (COREQ) (Tong *et al.*, 2007).

4.3 Methods

Unlike purely quantitative research, it is common when conducting qualitative research to outline how the researcher views the world as justification for the choice of approach taken and the impact this might have on the results. Creswell, 2013 define a world-view as a set of beliefs that influence the actions and choices that a researcher makes when carrying out a research project. In this thesis, a pragmatic world-view has been adopted, as pragmatists look for the most appropriate way to answer their research question (Creswell, 2013).

4.3.1 Development of the Thematic Framework for Design and Analysis of Qualitative Study

A thematic framework is a structured summary of key ideas or topics (themes) relating to the research question. An initial thematic framework was developed based on the research aims, literature and discussions with the public advisory group. Initial themes identified included Pre-specification of Health Economic Analysis; Private Sector Attitude; Aims of a Clinical Trial and Advantages and Limitations of Adaptive Designs. This initial framework was used to develop the set of questions asked during data collection, known as the topic guide.

Once data collection was under way, the framework was amended based on inferences about participant perspectives with no obvious place in the initial framework. In consultation with

Section	Subsection	Description
Ethical	Planning	Ethical issues that should be considered when planning trials (before they begin) using adaptive designs and health economics in HTAs
	Implementation and Conduct	Ethical issues that should be considered when conducting or implementing trials using adaptive designs and health economics in the real-world (during the trial)
	Documenting and Reporting	Ethical issues that should be considered when writing documentation for trials using adaptive designs and health economics (not just reporting after the trial has ended)
Methodological	Development	Important statistical issues to consider when developing the use of adaptive designs and health economics in HTA
	Current Use	How are adaptive designs, economic evaluations and VOIA currently used in practice and what are their advantages and disadvantages
	Experience and Knowledge	What experience and knowledge do the participants have of the these methods and where has this come from
Practical	Planning	Practical issues to consider when planning a trial/before conducting a trial with an adaptive design and health economics
	Implementation and Conduct	Practical issues to consider when conducting or implementing a trial with an adaptive design and health economic analysis
	Documenting and Reporting	Practical issues to consider when reporting and writing documentation for a trial with an adaptive design and health economic analysis (not just reporting when trial has ended)

TABLE 4.1: Final thematic framework applied to the data collected during the qualitative study

DH the final framework condensed the themes into three main headings reflecting the key areas of interest; ethical, methodological and practical considerations. The framework is summarised in Table 4.1 and was informed by existing literature (O'Brien *et al.*, 1994; Briggs *et al.*, 1998; Briggs, 2000; Cook *et al.*, 2004; Ashby *et al.*, 2005; Koerkamp *et al.*, 2008; Trotta *et al.*, 2008; Koerkamp *et al.*, 2010; Hollingworth *et al.*, 2013; Mullins *et al.*, 2014; Tuffaha *et al.*, 2014; Legocki *et al.*, 2015; Meurer *et al.*, 2016; Husbands, 2016; Dimairo *et al.*, 2018), discussions with the public advisory group and early data collected.

4.3.2 Participant Selection

Participants were identified based on their characteristics and experiences of adaptive clinical trials, health economics and HTAs. This approach was adopted to identify the views of a range of key stakeholders in the HTA process. The desired stakeholders were categorised into two groups.

1. Public - Any member of the public regardless of their experiences as a patient, user of the NHS or participation in clinical trials.

2. Researchers and Decision Makers - Any researchers or decision makers involved in HTA, especially those who will be using the methods developed. No exclusions were made based on sector, experience or location.

The views of participants in each group were not thought to be mutually exclusive. All researchers and decision makers are also members of the public and their opinions cannot be separated.

Members of the public were first contacted through a public involvement co-ordinator at a clinical research unit at a local hospital, a departmental public involvement lead and the public advisory panel supporting the research project. After the initial data collection, to extend the reach beyond members of the public with research experience, a message was posted to an online forum (Sheffield Forum).

Researchers and decision makers were identified by contacting researchers known to be working in relevant areas such as prominent health economists, statisticians and clinicians and by utilising known contacts in industry.

4.3.3 Data Collection and Sampling

Semi-structured interviews and focus groups were used to collect data. This choice was made after considering their benefits and limitations as summarised by Pope *et al.*, 2000 and following discussions with the public advisory group.

Data collection with members of the public was conducted face to face at the University of Sheffield where possible; telephone interviews were available to avoid excluding members of

the public who were unable to travel. Telephone interviews were the preferred approach for researchers and decision makers to reduce the burden on participants as previous research reported difficulties in engaging with health economists (Dimairo *et al.*, 2015).

Separate study documentation and topic guides were produced for the public and researcher groups reflecting the technical language and information provided to the respective participants. Prior to data collection, the researcher guide was piloted with a colleague. Recommendations from the public advisory group on how to modify the topic guide for members of the public were implemented.

Prior to data collection, all participants were sent a short video introducing the key concepts of adaptive designs and health economic analysis in plain English to facilitate participation in the study. This video was created with the support of the public advisory group so as not to include any of my preconceived ideas as this might influence the subsequent responses of the participants. (https://www.sheffield.ac.uk/scharr/research/centres/medical-statistics/trials. Last Accessed: 24.07.2020)

I conducted all the interviews and focus groups and they lasted between 30 to 60 minutes and 60 to 90 minutes respectively. The interviews and focus groups were audio recorded. There were no repeat data collection. I transcribed all the interviews and any identifying information was removed. Demographic data were collected using an online Google Form.

4.3.4 Sample Size

Literature on the required sample size for qualitative research suggests that the sample should provide sufficient data to develop an understanding of the experience under scrutiny whilst not producing more material than is manageable (Sandelowski, 1995; Braun *et al.*, 2013; Fugard *et al.*, 2015). Fugard *et al.*, 2015 and Braun *et al.*, 2013 suggest between six to 12 participants for interviews and two to four focus groups with group size ranging from four to 12 (Pope *et al.*, 2000; Guest *et al.*, 2006; Creswell *et al.*, 2007; Carlsen *et al.*, 2011).

Based on this literature, advice from the public advisory group, and keeping time constraints in mind a target of a maximum of 20 researchers and 20 members of the public was set. Focus groups were planned to have a maximum of seven participants. At each stage the demographics of the sample were reviewed and subsequent participants selected to give a balanced and

representative sample where possible, similar to the approach recommended by Francis *et al.*, 2010.

4.3.5 Analysis and Findings

Once transcribed and anonymised, the interviews and focus groups were imported into NVivo (QSR International v11). A framework analysis approach was adopted for the analysis (Ritchie *et al.*, 1993). A framework analysis identifies similarities and differences in the responses of participants to identify descriptive themes that summarise key points in the data (Gale *et al.*, 2013). This systematic approach to analysis had the following stages:

1. **Familiarisation** - the interview transcript and any field notes were used to familiarise myself with the data.

2. **Coding** - the transcripts were read line by line assigning a code (a word or phrase that describes what is discussed in the section of the transcript) from the *a priori* thematic framework or a new code.

3. **Thematic framework** - the thematic framework was updated with any new codes that emerge from the transcripts.

4. **Indexing** - transcripts were then re-read assigning the codes from the thematic framework to appropriate sections.

5. **Charting and Mapping** - all the data relating to each code in the thematic framework, from all of the transcripts were summarised into a table.

6. **Interpretation** - the results were interpreted in the context of the research aims.

Early results were discussed with the public advisory panel, to provide a relevant and accurate interpretation of the findings . Additionally, a table of recommendations was sent to study participants to comment and suggest changes. These were incorporated where appropriate.

Ethics approval was granted by the University of Sheffield, School of Health and Related Research Ethics Committee (ref:009699). The public advisory group reviewed the study documentation for the public data collection. All participants were issued with an information sheet and were required to sign a consent form or give verbal consent, agreeing to be part of the research

before data were collected. Participants were allowed to withdraw from the study at any time. All data were anonymised to protect participants' confidentiality. The meaning and integrity of the data was maintained.

4.4 Results

This section firstly describes the participants that took part in the study and their experiences, reflecting the current use of health economics in adaptive clinical trials. The three key themes emerging from the results are then discussed in turn; ethical, methodological and practical considerations. Data have been selected to illustrate these points based on quotations of interest. Quotes are presented with a description of the participant (for example a health economist) and a participant number (for example P1) so it is possible to identify where multiple quotes are provided by the same participant and to give some context to the quote being reported.

4.4.1 Description of Participants

Twenty-nine participants took part in the research between October 2017 and March 2018. At this point, it was felt that few new themes were emerging and a range of participants had been included. There were 18 one-to-one interviews and two focus groups. Participant characteristics are summarised in Table 4.2. One researcher did not complete the demographic survey. There were $^{2}/_{13}$ female researchers and $^{8}/_{15}$ female members of the public. Researchers predominantly came from statistical or health economic areas of expertise from both the private and public sectors. Two of the 13 researchers were based in the United States. Ten of the fifteen members of the public had participated in research through public involvement groups and $^{6}/_{15}$ had participated in a clinical trial.

The health economists interviewed had limited experience of adaptive designs and reported that they are frequently not consulted in the design of clinical trials. Participants working in the private sector suggested that they were increasingly involved throughout the whole trial as cost-effectiveness is a key requirement of HTA agencies such as NICE. However, the experience of health economists was also largely influenced by the clinical teams they work with and how important they perceive health economics to be in both the public and private sectors.

Question	**Response**	**Public** ($n = 15$)	**Researcher** ($n = 13$)
Gender	Female	8	2
	Male	7	11
Ethnicity (Free Text)	White, European	1	0
	White	3	3
	White British	8	7
	English	1	0
	Asian	1	1
	British	1	1
	Caucasian	0	1
Age (years)	30 or younger	1	1
	31-35	0	3
	36-40	5	2
	41-45	0	2
	46-50	1	2
	Older than 50	8	3
Highest Academic Qualification	Doctorate	1	8
	Post-Graduate Degree (Masters)	2	5
	Registered nurse & registered midwife	1	0
	Undergraduate Degree	6	0
	Undergraduate Degree, Teaching certificate	1	0
	GCSE	1	0
	NA	3	0
Have you ever participated in a clinical trial?	No	9	-
	Yes	6	-
Have you ever been a member of a public involvement group?	No	5	-
	Yes	10	-
Experience in clinical trials research (years)	11-15 years	-	5
	16-20 years	-	1
	5-10 years	-	1
	Less than 5 years	-	2
	More than 20 years	-	3
	NA	-	1
Current Employment Sector	Private	-	6
	Public	-	7
Location	United Kingdom	-	11
	United States of America	-	2
Area of Expertise (can choose multiple)	Clinician	-	1
	Health Economics	-	8
	Health Services Research	-	1
	Value of Information Analysis	-	2
	Adaptive Designs	-	3
	Clinical Trial Management	-	1
	Statistics	-	5

TABLE 4.2: Demographics of participants who took part in the qualitative study. One researcher did not complete the demographics questionnaire

Research participants suggested that health economists and statisticians frequently work independently, with their key point of contact being the chief investigator. However, study participants did not feel this negatively affected their work.

The knowledge and experience of members of the public who took part in the study came predominately from three main sources: the media, involvement in research as public advisors, and participating in a clinical trial. Participants acknowledged that the media was likely to give biased views on healthcare decision making for the public and there was perhaps a lack of trustworthy resources available.

> *. . . Jeremy Vine is talking about the cost per QALY and no one knows what that means. And why would they?* P39 Health Economist

4.4.2 Participant Experience of Health Economics in Adaptive Trials

Increasing the prominence of cost-effectiveness; improving the quality of health economic analyses and preventing adaptive designs stopping before there is sufficient evidence for an accurate health economic analysis were suggested by study participants as some of the advantages of using health economics in adaptive clinical trials. A study participant suggested that while formal VOIA calculations are rare, informal 'rules of thumb', based on the same ideas are commonplace. Developing decision rules for cost-effectiveness of the research similar to the rule for clinical effectiveness ($p<0.05$) and cost-effectiveness of the intervention (willingness to pay thresholds) would help the practical interpretation and implementation of this approach.

Few participants had seen the use of health economics in the design and analysis of adaptive clinical trials in practice. It was felt that further work is needed to extend existing theory to the real-world context and concerns were raised that combining these two complex methods might be deemed too challenging and hinder their uptake. It was also suggested that improvements to the methods of analysis are required to provide accepted standards such as the time horizon over which value of information is calculated. The use of adaptive designs and VOIA in trial design and research prioritisation would need to be more commonplace before adding further complexity by combining the two approaches in a trial.

> *. . . putting both together is going to be doubly complicated and time consuming and costly and so on . . .* P35 Health Economist

4.4.3 Ethical Considerations for the Use of Health Economics in Adaptive Trials

Participants in the qualitative study agreed that clinical effectiveness should be the aim of a clinical trial whether it is adaptive or a fixed sample size design. Both researcher and public participants could appreciate the importance of demonstrating cost-effectiveness in a trial; however, they felt it was still important to answer the clinical question first.

> *It feels to me like it might be unfortunate to move into a world where, [...] clinical research, in general per se, was shaped too much, designed too much with economic considerations ...* P19 Member of the Public.

On the other hand, some participants argued that cost-effectiveness decisions provide unseen benefits in healthcare, for example, if a trial stopped early on cost-effectiveness grounds this would allow resources to be spent elsewhere in a more cost-effective way. This would confer benefits to participants outside of the clinical trial. One researcher suggested incorporating health economics into the adaptive clinical trial changes the focus of the trial from the individual to a population level. However, they felt this might be deemed unacceptable to some stakeholders in HTAs. Study participants suggested that trial information sheets could include more information about the wider societal benefits of research, but this would need to be handled carefully as potential participants may be vulnerable and facing difficult clinical decisions.

Participants felt that the acceptability of the methods and the changing role of cost-effectiveness to members of the public might vary depending on their personal circumstances. Members of the public who are fit and healthy are more likely to accept the increased role of cost-effectiveness in an adaptive design that uses health economics. Other patients may be more concerned about whether the intervention can improve their clinical outcomes regardless of cost.

> *I mean the cost-effectiveness is very important if you don't need it [...] but if you're the person who needs the drug then the drug is important even if it's very expensive.* FG1 Member of the Public

A number of the participants agreed that a health economist should be included on a data monitoring and ethics committee (DMEC) if health economics is used as part of the design and

analysis of an adaptive trial. On the other hand, one researcher argued that the funder should make funding decisions and it would not be appropriate to delegate this responsibility to the DMEC.

> *...the DMEC have a very specific role looking at safety [...]. But once you get into value of information that's the funder's decision and I don't think the funder should delegate that to the DMEC.* P1 Funding Panel Member

4.4.4 Methodological Considerations for the Use of Health Economics in Adaptive Trials

A number of participants were concerned about how the volatile nature of cost data would be handled if interim adaptations were made based on cost-effectiveness. This adds complexity when deciding whether to modify (or even stop) a trial on a given day when an interim analysis takes place, as this may be the wrong decision if, for example a drug price decreases in the future.

> *There's still this question of whether it is cost economic now as per tomorrow, so you might miss an opportunity.* P6 Statistician

Study participants also highlighted that global multi-centre trials may encounter difficulties when using cost-effectiveness to inform interim adaptations. An interim analysis in one centre might demonstrate sufficient evidence to stop the trial early based on the cost-effectiveness rules in their jurisdiction, but this might not be reflected in other centres around the world. Additionally, the extent to which adaptive trials incorporate health economics in the private sector might depend on whether their focus is on the US market or other countries where cost-effectiveness plays are more prominent role, such as the UK.

> *I mean no pharma company is going to have a trial that's based on cost-effectiveness because the biggest market is the US* P37 Health Economist

It was unappealing to many participants that an adaptive design could stop early based on the cost-effectiveness of the intervention or the research itself. Alternative uses for the interim data suggested by participants included:

1. A hierarchy of stopping rules where the clinical outcome is the first consideration in the adaptive decision making and then, dependent on this result, cost-effectiveness may also be used to inform modifications to the trial.

2. The trial could be modified based on health economic grounds rather than the more extreme case of stopping a trial.

4.4.5 Practical Considerations for the Use of Health Economics in Adaptive Trials

Study participants felt that successfully implementing this approach would require sufficient resourcing, careful planning, and building a study team of researchers and clinicians who have sufficient training and are supportive of the approach. Adding the health economic layer to adaptive clinical trials raised a resourcing problem for many participants as more time and resources will be required before a project has been funded. In the private sector, the impact of this was thought to vary depending on the size of the company.

Funding panels and HTA committees are comprised of non-experts. They will require sufficient training in the methods to allow them to assess their appropriateness in answering the research question and to interpret and assess the quality of the evidence provided. Study participants highlighted the value of developing software to help make methodology accessible and increase its use. Case studies were suggested as a way of demonstrating the value of using health economics in adaptive clinical trials. This could increase the understanding of the methods and appease concerns of many participants regarding the complexity of this approach. The methodology would need to be presented in an accessible way for non-experts.

> *. . . the crucial thing is that the analysis has to be very understandable because [. . .] in practice it will be unhelpful if a value of information analysis was updated and appeared to everyone involved in the trial to be a black box analysis.* P1 Funding Panel Member

Study participants suggested focussing attention initially on clinical areas where adaptive designs are more commonplace such as oncology or innovative areas of research such as medical devices. However, a health economist highlighted that health economic models in oncology

are often complex. Additionally, clinicians and patients might challenge the use of health economics in this way as their focus is on the individual and clinical decisions rather than the population health decisions.

It will be increasingly important for the health economic analyses to be pre-specified if they are used to inform adaptive decision making during a trial. There would need to be a greater level of scrutiny of the health economic analyses. However, concerns were also raised by participants, about the potential challenge of pre-specifying everything for a health economic analysis prior to any trial results. Participants working in industry were concerned about the impact of early examinations of the data on cost-effectiveness grounds and the requirements to publicly report all analyses.

> *. . . it's sort of . . . you've got chicken and egg here. [. . .] Your model is developed as you look at the data but I think it would be quite difficult to do that because, yeah they sort of evolve alongside each other* P39 Health Economist

4.5 Discussion

4.5.1 Summary of Key Findings from this Chapter

Members of the public, researchers and decision makers identified ethical, methodological and practical issues associated with the use of health economics in adaptive clinical trials. Recommendations and proposed actions are drawn from the themes identified in the data, reflecting the differing views of stakeholders in HTAs and experiences of the research team and public advisory group. Recommendations are made under the three key themes ethical, methodological and practical and are summarised in Table 4.3.

4.5.2 Considerations for Current Practice and fit with Existing Literature for the Use of Health Economics in Adaptive Trials

4.5.2.1 Ethical Considerations

As shown in Chapter 3, the use of health economics in the design and analysis of adaptive and non-adaptive clinical trials is largely secondary to clinical outcomes. Study participants felt that for health economics to play a greater role there needs to be a shift of emphasis from

Issue	Recommendation	Action
Clinical effectiveness is the main focus when thinking about the aims of a clinical trial.	The importance of clinical effectiveness should be reflected in the development of methods for using health economics in adaptive trials. Possible approaches include • Using early examinations of the trial to check all health economic data are being collected as required. • Use early trial data to update the health economic model. • Using a hierarchy of interim decision rules where any decisions made based on cost-effectiveness depend on decisions made about clinical outcomes. • Only considering health economic outcomes at later examinations of the data. • Using health economic data to make modifications to the trial such as increasing the sample size but not major changes such as to stop the trial early.	Explore how existing methods for the use of health economic based stopping rules would work in the real-world setting, by applying the methods to a diverse range of case studies.
Study participants appreciate the importance of cost-effectiveness to decision makers but they considered this secondary to clinical effectiveness.	There needs to be a change in the mentality of the research community towards the role of health economics and cost-effectiveness in healthcare decision-making.	Development of materials to educate on the importance and ethical motivations for thinking about cost-effectiveness. This could be written with members of the public and developed into workshops, online materials and leaflets.
Stakeholders may not be familiar with the methods of adaptive clinical trials or VOIA and their potential advantages and limitations.	Develop software and tutorial style case studies for researchers to help them understand the methods and allow them to interpret the results of trials using this approach or use these methods in their own research. Development of plain English summaries and case studies highlighting the impact of the methods on patients and the public.	A Practical Adaptive and Novel Designs Toolkit (PANDA) is under development that aims to provide researchers with training materials on adaptive design clinical trials (Dimairo *et al.*, 2019b). This could include materials aimed at members of the public. Organise workshops (such as at conferences) for researchers highlighting the potential for the methods to be used together and issues to consider.
If health economics is to inform decisions made using early examinations of data there may be a need for this specialist knowledge on trial committees.	Include health economists on DMECs where health economics is used as part of the design and analysis of adaptive trials.	Existing resources that help research teams identify DMEC statisticians, such as StatLink (NIHR Statistics Group, 2018), could be extended to identify health economists. All DMEC members could be paid for their contribution and time. Use mock DMECs to allow members to review the health economic and clinical data and see where issues with using the health economic data arise.

Issue	Recommendation	Action
Using health economics in the design and analysis of adaptive trials will require more work before the trial is funded and so researchers are not paid for this work.	Funding bodies should provide alternative funding options that allow resources to develop new designs. Researchers could include time at the start of a study to fully develop an adaptive trial design that uses health economics. Researchers should look for methodology grants to fund the development of designs.	Groups representing statisticians and health economists (such as the MRC Adaptive Designs Working Group and ISPOR) should work together to persuade funders and regulators on the need for alternative ways to fund adaptive clinical trials and the benefits this will have to health research and maximising limited research budgets.
There is currently limited interaction between health economists and statisticians working on trials and more generally between the two research communities.	Encourage statisticians and health economists to work together and increase communication to facilitate the implementation of health economics in adaptive trials by sharing expertise.	Locally, health economists and statisticians working on the same clinical trial should have regular meetings throughout the study. Nationally, joint events between groups such as the NIHR Statistics Group and the Health Economic Study Group discussing common issues and encourage training in statistics for health economists and health economics for statisticians.
If health economics is to be used in adaptive clinical trials, the methods need to be outlined in advance so the results of the trial are still valid and robust.	Before the trial begins researchers should outline how health economics is going to be used in the trial and how the early examinations of the data will be used to calculate cost-effectiveness and inform decision making.	Extend current work developing guidance for health economic analysis plans to think about specific issues that might arise in adaptive clinical trials (Thorn *et al.*, 2017). This will likely require health economists to have some experience of working on an adaptive trial.
Calculating the costs of conducting an adaptive clinical trial can be complicated. For example, providing justifications of costs and cost projections in a grant application.	It is important to understand the costs of conducting an adaptive trial such as the costs of finding patients, training staff and analysing data so that you can compare adaptive and non-adaptive trial designs and inform stopping rules based on health economics.	Develop a standardised approach for calculating the costs of an adaptive clinical trial, illustrated using a case study.

TABLE 4.3: Summary of key recommendations and proposed actions from the qualitative study

the ethical perspective of beneficence – the duty to allocate resources to benefit the individual (Beauchamp *et al.*, 2001) - to the broader perspective of distributive justice - allocating resources to benefit the whole population (Rawls, 1971; Daniels, 1985). To change the status quo requires the public and policy makers to find the compelling ethical arguments that cost-effectiveness decisions provide an unseen benefit (Williams, 1992). Practical steps towards achieving this could include developing unbiased and simple materials on adaptive clinical trials and health economics. Public advisors should be involved in the development of these materials and consent forms. These documents should explain the methods of adaptive designs and health economics in sufficient detail for potential trial participants to make an informed decision to participate, as well as highlighting the broader implications of taking part in the research.

The acceptability of the role of cost-effectiveness to stakeholders is crucial when considering the direction of methods development. It is fruitless to develop statistical methods that incorporate health economics into an adaptive clinical trial if this is never implemented in practice due to concerns over the role of cost-effectiveness. For example, patients may be unwilling to be randomised as they do not agree with decision making centred on cost-effectiveness. Based on discussions with study participants and the public advisory group, I recommend exploring alternative approaches including:

- Interim analyses used to check that all health economic data are being collected as required.
- Interim data are used to update the health economic model.
- A hierarchy of interim decision rules are used where the health economic decisions are dependent on the outcome of the clinical effectiveness analysis.
- Health economics decision rules that are only considered at later interim analyses, giving the advantage of more mature data and more weight to clinical outcomes at the start of the trial.
- Health economics used to make less extreme modifications such as increasing the sample size based on an EVSI calculation (described in Section 2.8.2).

Summarised in Section 2.9, the methods of Mehta *et al.*, 2006 and Kouvelis *et al.*, 2017 maintain the importance of clinical effectiveness in their application of Bayesian decision theory in sequential clinical trials. This could be considered by clinical teams who want to keep the focus of their trial on clinical effectiveness but acknowledge the importance of cost-effectiveness in healthcare decision making. The choice of approach could be made in collaboration with a public advisor so that the identified approach is acceptable to patients.

4.5.2.2 Methodological Considerations

The advantages of using health economics in adaptive designs include saving resources and preventing patients from being needlessly randomised (Chow *et al.*, 2012b; U.S. Food and Drug Administration, 2019). Such an approach will formalise the informal rules of thumb described by a study participant as commonly used by decision makers and funding panels, so that they

can be applied in an objective and consistent manner. In agreement with Bindels *et al.*, 2016 I suggest developing decision rules for cost-effectiveness of the research similar to the rule for clinical effectiveness (p-values) and cost-effectiveness of the intervention (willingness to pay thresholds) would help the practical interpretation and implementation of this approach. Public advisors could be used to ensure these rules are appropriate and reflect the views of patients and the public.

It is also important to consider the limitations of both methodologies; including the potential for adaptive designs to introduce bias in the estimation of point estimates and confidence intervals, as discussed in Section 2.5 and illustrated in Chapter 7. It is also important to maintain appropriate levels of blinding during interim analyses so as not to influence the conduct of the trial (Whitehead, 1997; Pallmann *et al.*, 2018).

The methods of adaptive designs (Hatfield *et al.*, 2016; Bothwell *et al.*, 2018) and VOIA (Mohiuddin *et al.*, 2014; Steuten *et al.*, 2013) are not common place in HTAs. It will be vital to address the limitations of these methods to successfully implement them together in clinical trials. As identified by the study participants, these methods could be seen as more complex, potentially acting as a barrier to their use in practice. This reflects the findings of Dimairo *et al.*, 2015 who found that adaptive designs were perceived as more complex than traditional fixed sample size designs. Additionally, Bindels *et al.*, 2016 found that participants who worked in pharmaceutical companies thought that VOIA may be complex and that policy makers have limited knowledge to interpret the results.

Work in the field of adaptive clinical trials to promote their use and understanding amongst the health research community includes,

- The Adaptive Designs Working Group, part of the Medical Research Council (MRC) Hubs for Trials Methodology Research that aims to increase implementation of adaptive designs through tutorial papers in mainstream medical journals such as (Pallmann *et al.*, 2018), development of software, presentations and lectures.

- The Adaptive Designs CONSORT Extension (ACE) project that aims to enhance transparency, credibility, reproducibility, and replicability of adaptive trials by developing a

consensus driven extension to the CONSORT statement specific to adaptive designs (Dimairo *et al.*, 2018).

Measures for increasing the use of VOIA include

- The ISPOR VOIA for Research Decisions task force that aims to develop good practice guidance for using methods of VOIA to inform both technology reimbursement decisions and research prioritization decisions (ISPOR, 2020; Fenwick *et al.*, 2020; Rothery *et al.*, 2020).

- Collaborative Network for Value of Information (ConVOI) group are an international network working to improve the calculation, adoption and application of VOIA in clinical and public health research Heath *et al.*, 2019; Kunst *et al.*, 2019.

Similar initiatives could be used to promote the use of health economics in the design and analysis of adaptive designs.

4.5.2.3 Practical Considerations

If the use of health economics becomes common in the analysis of adaptive clinical trials, then guidance on the composition of DMECs (DAMOCLES Study Group, 2005) should be updated and lists which help identify DMEC members – such as STAT link (NIHR Statistics Group, 2018) – could be extended to include health economists .

Such an approach will have cost consequences for clinical trials with the involvement of health economics throughout. Dimairo *et al.*, 2015 discuss the need for more work when choosing an adaptive design prior to the trial being funded. When health economics is used as well, more work will be required upfront by health economists, such as to develop a health economic model. The development of a health economic model and cost-effectiveness analysis based on systematically reviewing evidence is an essential step prior to undertaking any trial regardless of its design (Glasziou *et al.*, 2015). In the public sector, where there are few resources to bridge grant funding, researchers would need to include the costs of trial design within grant applications. Groups representing statisticians and health economists will need to persuade funders and regulators of alternative ways to fund adaptive clinical trials and VOIAs.

If health economics is to be successfully implemented as part of an adaptive clinical trial, I recommend that trial sponsors ensure integrated working between health economists and statisticians throughout the trial process. More generally across the research community links between statisticians and health economists could be encouraged through networking events and joint workshops to identify common ground and explore ways of working together more efficiently.

Pre-specification of health economic analyses in a Health Economic and Decision Modelling Analysis Plan (HEDMAP) will be crucial in maintaining the validity and integrity of adaptive designs that use health economics, with analysis plans including a description of the monitoring and adaptation plan, as well as pre-specification of methods used at interim analyses (Thorn *et al.*, 2017; U.S. Food and Drug Administration, 2019). The timing of interim analyses must be realistic given their complexity and their blinding must be carefully considered. Cross-disciplinary training materials will be required, supplemented by case studies on the use of health economics in sequential trials, such as Forster *et al.*, 2019, to raise awareness and advance methodological development by highlighting practical problems with their application.

4.5.3 Strengths and Limitations

This is the first qualitative study to explore the ethical, methodological and practical issues of using health economics in the design and analysis of adaptive clinical trials. Participants came from different backgrounds and with experiences across the healthcare decision making process. This study included a number of health economists, a group previously found to be hard to reach in adaptive designs research (Dimairo *et al.*, 2015).

Members of the public were an important stakeholder group in this research as the ultimate beneficiaries of clinical trials and health care decision making that uses an adaptive design and health economic analysis. Additionally, members of the public are potential participants in trials that use these methods. Failing to reflect their views in the design and analysis of clinical trials could make it difficult, for example, to recruit to trials that use unpopular designs based on cost-effectiveness.

Few female researchers took part in the study; however, it is not felt that the gender of the participants is likely to influence their responses. Instead, participants' knowledge and experiences are more likely to inform their comments. Further work could consider a larger sample size and a broader range of experience of participants, such as programme managers of NIHR funding streams.

A public advisory group played an intrinsic part in the design and analysis of the qualitative study. They provided feedback on the information sheet, consent form and topic guide for members of the public to ensure the questions and proposed plan were suitable. The group developed the script for the short video sent to all study participants. After analysis was complete the group met again to discuss the results and check the interpretation of findings, where appropriate their interpretation has been embedded in the thesis.

A limitation of the research is that the members of the public and non-expert researchers and decision makers were given a top-level understanding of the topic of VOIA so their responses are considering a broad view of cost-effectiveness in clinical trials. However, the views of participants are still important in understanding the priorities of stakeholder groups around the role of cost-effectiveness in adaptive clinical trials. Further work could consider a more detailed explanation of the methods of VOIA providing scenarios of potential roles for the methods in adaptive clinical trials and using these to explore the views of stakeholders with a deeper understanding of the method.

4.5.4 Considerations for the Thesis

In Chapter 2 the literature on health economics and adaptive designs was reviewed. Only one article was identified that discussed the potential for bias in the health economic analysis of an adaptive design (Marschner *et al.*, 2019). This finding was also reflected in the experience of the qualitative study participants and in Chapter 3 where none of the trials with an adaptive design and health economic analysis identified adjusted their analysis to account for the adaptive nature of the trial. This is despite existing literature on the potential impact and adjustments available, as summarised in Chapter 2.

Understanding the impact these designs have on a health economic analysis and how existing adjustments can be extended to this context is coming more urgent, as the use of adaptive

designs increases. In Chapter 6, focussing on the most common adaptive design, the theory for adjusting an analysis following a group sequential design is extended to the health economic context. This is then operationalised in a simulation study in Chapter 7 to explore the potential impact the designs have on the health economic analysis and how well existing adjustment methods control for this.

The qualitative study participants were clear that the aim of a clinical trial should be to establish the clinical effectiveness of an intervention. The public advisory group supported these views. This may explain why the methods summarised in the literature review (Section 2.9) have not yet been used in practice (shown in Chapter 3). Many of these methods establish stopping criteria for the trial based on cost or cost-effectiveness considerations rather than the clinical effectiveness. On this basis, it was decided that the focus of the methodological work in the thesis should build on current practice and these views by keeping health economic outcomes secondary to the demonstrating clinical effectiveness.

In Chapter 8 cost-effectiveness considerations will be used to guide the design of a group sequential clinical trial with a clinical effectiveness stopping rule. This acts as a compromise, acknowledging the importance of clinical effectiveness as the aim of the trial but not neglecting cost-effectiveness considerations in the allocation of scarce research budgets. The extension of these methods will consider the impact of the adaptive nature of the trial designs on the methods and explore whether any adjustments are required, building on the work of Chapters 6 and 7.

4.6 Chapter Summary

This chapter summarises the views of stakeholders in the HTAs process on the use of health economics in the design and analysis of adaptive clinical trials, answering the second research aim outlined in Chapter 1.

Noting the suggestions from the qualitative study to aid the implementation of health economics in adaptive designs, Chapter 5 describes the case study used to illustrate potential issues, and recommendations are made around the appropriate adjustments and considerations when planning and conducting a health economic analysis using data from an adaptive design.

Chapter 5

Introduction of the CACTUS Case Study

5.1 Introduction

In Chapter 3 it was shown that there is a lack of clinical trials with an adaptive design using health economics in their design and analysis. The qualitative study in Chapter 4 highlighted potential barriers to the use of these methods. The work in these chapters highlights a clear gap in the research and areas for further methodological development regarding the use of health economics in adaptive clinical trials. Firstly, the need to understand the impact of an adaptive design on a health economic analysis, specifically establishing the adjustments needed for analysis following a group sequential design. Secondly, there is an opportunity for EVSI methods, described in Section 2.8.2, to be extended to guide the design of group sequential trials. These issues are explored further in Chapters 6 and 8 respectively.

A case study is used as the basis for these investigations, which needs to meet three key criteria relating to its characteristics, feasibility and generalisability.

5.2 Chapter Aims

In this chapter each of the criteria (characteristics, feasibility and generalisability) are discussed, outlining the requirements for a suitable case study. The CACTUS pilot trial and Big CACTUS trial are described highlighting how they meet the key criteria and the health economic

methods used in the original analysis are described in detail. The original methods have been replicated in R for use in subsequent chapters of this thesis.

5.3 Criteria for a Suitable Case Study

5.3.1 Characteristics

A suitable case study for exploration of the use of health economics in the design and analysis of adaptive clinical trials needed to be a randomised controlled trial. The trial needed to have a group sequential design or the potential to have used this design. A trial with a fixed sample size design has the potential to use an adaptive design when the time for participants to reach the primary outcome is short relative to the length of recruitment. Sully *et al.*, 2014 suggest that where the time to the primary outcome relative to the length of recruitment is less than 0.25 there is potential for the trial to use an adaptive design. In this instance, there is sufficient time for the first patients to reach the outcome of interest, for an interim analysis to take place and modifications to be made to the trial while it is still recruiting.

The case study was also required to have a body of pilot work that could be used to inform the design of a full-scale trial. This might include a pilot or feasibility trial or a summary of existing evidence in a meta-analysis or evidence synthesis. This pilot work needed to have included a health economic analysis, including a health economic model so that a new health economic model was not required, given the time constraints of this thesis.

5.3.2 Feasibility

The chosen case study needed to be feasible for the work in this thesis. The data for at least the pilot work needed to be available from 2016 and it was necessary for the health economic model to be reproducible in R so that it could be incorporated into analyses using existing R packages for the design and analysis of group sequential trials. It was also desirable for the study team (chief investigator, study manager and health economist) to be available and willing to provide advice and clarification on the methods used and analysis undertaken.

5.3.3 Generalisability

For the findings of the thesis to be generalisable beyond the chosen trial, it was important for the chosen case study to reflect current practice in as many ways as possible. It is acknowledged that some findings will be specific to the chosen trial and setting, however lessons learnt and process based findings are likely to be transferable beyond the chosen trial if the right case study was identified. On this basis, the chosen case study was required to include:

- A within trial and model-based health economic analysis;
- A health economic model informed by trial data;
- Health economic analysis that meets NICE guidance for technology appraisals;
- EQ-5D utilities collected during the trial (required by NICE);
- Resource use data collected as part of the trial.

5.4 CACTUS Pilot Trial

Aphasia is a language disorder that affects a person's understanding, talking, reading and writing, commonly occurring in patients who have survived a stroke (National Health Service, 2018). An estimated 50 per 100,000 of the population will have aphasia 6-months following a stroke (Royal College of Speech and Language Therapists, 2009). The Cost-effectiveness of Aphasia Computer Treatment Compared to Usual Stimulation (CACTUS) pilot clinical trial aimed to assess the feasibility of conducting a large scale clinical trial into the effectiveness of self-managed computer treatment for people with long-standing aphasia post stroke (Palmer *et al.*, 2012). Participants were randomised to either receive a computer-based intervention Step-by-Step (Steps Consultancy Ltd, 2018) designed to improve word finding ability through language exercises for people with aphasia or usual care acting as the control.

The CACTUS pilot was a single blind parallel group, stratified, pilot randomised controlled trial. A 5-month intervention period was followed by a 3-month period without the intervention to determine whether the effect of the treatment was maintained (Palmer *et al.*, 2012). The CACTUS pilot recruited 34 participants to each arm, with 28 participants followed-up at

5-months (15 computer-based intervention, 13 usual care). The difference between the intervention and control arm in improvement in percentage word naming ability from baseline to follow-up was 19.8% (95%CI : 4.4 to $35.2; p = 0.014$) in favour of the computer based intervention (Palmer *et al.*, 2012).

As reported by Latimer *et al.*, 2013, the pilot health economic analysis concluded that the computer-based intervention had an ICER of £3,058 per QALY (INB=£2,636.13) compared with usual care . The probability of cost-effectiveness was 75.8% for a £20,000 per QALY threshold. Latimer *et al.*, 2013 calculated the per-patient EVPI to be £143.68. Extrapolating this to the population level gave an EVPI of approximately £37.0 million based on an average of 27,616 patients being treated over a ten-year period. There was no EVSI calculation. Table 5.3 summarises results from the pilot health economic analysis.

The pilot trial concluded that the computer-based intervention was feasible and that it would be possible to recruit participants for a full scale randomised controlled trial (Palmer *et al.*, 2012). The pilot trial was therefore followed by the Big CACTUS trial.

5.5 Big CACTUS Trial

The Big CACTUS trial, was a three arm trial in participants with aphasia at least four months after having a stroke (Palmer *et al.*, 2015). The trial compared the computer-based intervention and usual care arms included in the pilot trial (Palmer *et al.*, 2012). The full-scale trial included an additional attention 'standard care' arm to control for the potential impact of elements of the intervention which do not provide or require specific speech and language intervention. Participants in this arm were given books of standard puzzles to carry out on a daily basis (Palmer *et al.*, 2015).

Co-primary outcomes of change in the number of words named correctly and improved functional communication were measured at baseline, 6, 9 and 12 months follow-up. The study recruited 278 participants between October 2014 and August 2016. The primary outcome results are summarised in Table 5.1 as reported by Palmer *et al.*, 2019. The trial concluded that the computer-based intervention resulted in a clinically significant improvement in word naming ability but did not improve functional conversation.

	CSLT vs usual care		CSLT vs attention control	
	Adjusted mean difference in change (95% CI)	***p*-value**	**Adjusted mean difference in change (95% CI)**	***p*-value**
Change in word finding (%)	16·2 (12·7 to 19·6)	< 00001	14·4 (10·8 to 18·1)	< 00001
Change in functional communication	-0·03 (-0·21 to 0·14)	0·709	-0·01 (-0·20 to 0·18)	0·915

TABLE 5.1: Summary of the primary outcome results from the Big CACTUS clinical trial (Palmer *et al.*, 2019), CI: confidence interval, CSLT; computer speech and language therapy

The health economic model from the pilot trial was adapted for Big CACTUS. The cost-utility analysis concluded that the computer-based intervention is unlikely to be cost-effective at the £20,000 per QALY threshold when compared to usual care. The incremental cost was £733 (95% credible interval 674 to 798) and incremental QALY was 0·017 (–0·05 to 0·10) giving an incremental cost per QALY gained of £42,686. In sub-group analyses the ICER was £22,371 per QALY gained for participants with a mild word finding difficulty indicating that it may be more cost-effective in participants who have mild word naming difficulties and £28,819 for participants with a moderate word finding difficulty (Palmer *et al.*, 2019). For the computer-based intervention compared with attention control, the ICER was £40,164 per QALY gained when evaluated across all participants.

5.6 CACTUS Pilot Health Economic Analysis

A health economic analysis using data from the CACTUS pilot aimed to provide an early analysis of the likely long-term cost-effectiveness of self-managed computer therapy for people with aphasia and the value of conducting further research (Latimer *et al.*, 2013). Using a model based cost-utility analysis outcomes were estimated using QALY to calculate an incremental cost per QALY, as discussed in Chapter 2. An NHS and personal social service perspective was taken and costs and benefits were discounted at a rate of 3.5% as per the NICE guidance (National Institute for Health and Care Excellence, 2013a). The analysis was based on available data for each outcome. There was no imputation for missing data.

5.6.1 Health Economic Model

A Markov model, as described in Section 2.7.7, with three states, Response, Aphasia and Dead states, was developed to assess the cost-effectiveness of the computer-based intervention. The

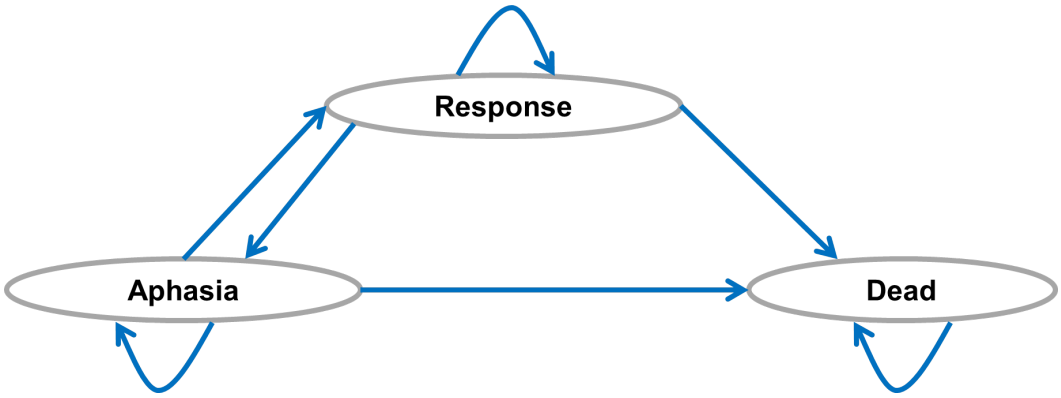


FIGURE 5.1: Illustration of the CACTUS Pilot health economic model based on the work of Latimer *et al.*, 2013

model is illustrated in Figure 5.1. In the model, all participants start in the Aphasia state. Participants can then move from the Aphasia state to the Response state, where they experience an improvement in their word naming ability and any costs and benefits associated with this. Participants move to the Dead state when they die.

Movement of participants in the model is based on month long cycles. This means after one month participants are able to move from the initial Aphasia state to the other states. Once a participant enters the Dead state, they cannot leave. The model imagines 1,000 participants, with the same characteristics as the trial population, move through the model over their lifetime. The costs and benefits they accrue by spending time in each health state are recorded and used to estimate the long-term cost-effectiveness of the computer-based intervention compared to control.

5.6.2 Transition Probabilities

To determine when a participant moves from one state to the next, transition, probabilities are defined, as described in Section 2.7.7. In the CACTUS pilot health economic analysis, transition probabilities between the Aphasia and Response states were calculated from the primary outcome pilot trial data. The proportion of participants who responded to the intervention was calculated as the number of participants in the intervention arm thought to have had a 'good' response to treatment divided by all the participants in the intervention arm. A 'good' response was defined to be an improvement of 17% or more in the percentage of words named correctly, from baseline to follow-up. It was assumed that if a participant demonstrated a good response

at month five of the trial this response occurred in month one of the model. This was estimated to be 53.3% from the CACTUS pilot data.

A relapse rate was defined to be the probability of a good response at 5-months in the intervention arm minus the probability of a good response at 8-months in the intervention arm and was incorporated into the model after 5-months allowing participants in the Response state to move back to the Aphasia state. In the CACTUS pilot $^{9}/_{17}$ participants in the intervention arm had a good response to treatment at 5-months. At 8-months this was $^{6}/_{12}$ giving a relapse rate of 0.8% per month.

None of the control arm participants experienced a 'good' response to treatment essentially giving a transition probability of zero between the Aphasia and Response states. Participants in the control arm are assumed to only move between the Aphasia and the Dead state. This assumption is maintained in the analysis of Chapter 7 and Chapter 8.

Transitions to the Dead state were based on the available literature on long-term survival following stroke (Brønnum-Hansen *et al.*, 2001) and Office for National Statistics Lifetables (Office for National Statistics, 2018). Mortality rates for participants who had experienced a stroke one or more years previously were applied to both the Response and Aphasia states after the first five years of the model.

5.6.3 Health Utilities

The benefits of being in a particular health state were measured using utilities, as described in Section 2.7.3, calculated from the pilot trial data using the EQ-5D. These values were reduced over time according to multipliers given by Ara *et al.*, 2010, to reflect how ageing decreases quality of life. It was assumed that utility scores were the same for non-responders in each treatment group. An incremental increase in utility score for responders was calculated as the difference between the utility of improvement of responders compared to non-responders from baseline to 5-month follow-up. An incremental increase in utility of 0.07 was found between those who did and did not respond to treatment (irrespective of their treatment group) in the pilot data.

5.6.4 Healthcare Costs

The cost of the computer-based intervention included; the cost of computers for the 65% of participants in the CACTUS pilot without their own (£495.99), the cost of the Step-by-Step software (£250), cost of microphones (£7.50) and the cost of support and training from a speech and language therapist (SLT). SLT time was converted into a cost using national unit costs (Curtis *et al.*, 2011). In the CACTUS trial, this incurred a cost of £190.83 per participant. The total intervention cost was £769.25 per participant. It is noted that the intervention cost was reported by Latimer *et al.*, 2013 as £801.60 as this included costs for one participant associated with care received from a volunteer. On discussion with the authors, this cost was not included in the analysis presented by Latimer *et al.*, 2013 and so and intervention cost of £769.25 is used throughout this thesis.

Other costs, including GP visits and hospital admissions, were calculated based on participant diaries. These were £203.08 in the intervention arm and £270.97 in the control arm for the first 5-months. After 5-months, it was assumed the intervention arm participants incurred the same cost as the control arm participants.

5.6.5 Deterministic Analysis

A cost-utility analysis was used to assess the long-term cost-effectiveness of the computer-based intervention compared to usual care. QALYs were calculated from EQ-5D utility scores collected at baseline, 5-month and 8-month follow-up using linear interpolation for each participant. Total costs were the sum of the intervention and other resource use costs incurred by a participant during the trial. The ICER was estimated as the difference in the mean total costs divided by the difference in mean QALYs. The ICER was then compared to the NICE cost-effectiveness threshold of £20,000 per QALY.

5.6.6 Probabilistic Sensitivity Analysis

A probabilistic sensitivity analysis (see Section 2.7.8) was conducted to assess the uncertainty in the results. Latimer *et al.*, 2013 assigned a distribution to the model parameters, as summarised in Table 5.2. Values were sampled from these distributions 500 times and the health economic

Model Parameter	**Distribution**
Probability of good response in intervention arm	$Be(8.00, 7.00)$
Relapse rate	$Be(0.20, 23.80)$
Utility decrement in Aphasia health state	$logN(-0.80, 0.19)$
Utility improvement in Response state	$N(0.07, 0.11)$
Percent who required computer	$Be(11.00, 6.00)$
Mean SLT face-to-face time	$Ga(8.59, 0.56)$
Mean SLT non face-to-face time	$Ga(1.24, 0.09)$
Other health care resource use control arm) per month	$Ga(14.86, 18.24)$
Other health care costs (intervention arm Aphasia state) per month	$Ga(3.10, 65.57)$
Other health care costs (intervention arm Response state) per month	$Ga(3.10, 65.57)$

TABLE 5.2: Summary of parameters in the CACTUS pilot probabilistic sensitivity analysis and their distributions reported by Latimer *et al.*, 2013 SLT; speech and language therapist

model recalculated. It was assumed that there was no correlation between parameters. A cost-effectiveness threshold of £20,000 per QALY over a ten-year period, with a discount rate of 3.5% was assumed for an EVPI calculation run for 500 inner loops and 100 outer loops.

5.7 Rebuilding the CACTUS Health Economic Model

5.7.1 `heemod` R Package

The health economic model, described in Section 5.6.1, for the CACTUS pilot trial was built in `Excel` by Latimer *et al.*, 2013, referred to as the `Excel` model. This model has been reproduced for this thesis in `R`, referred to as the `R` model. As discussed in Section 2.6, `R` is used to conduct the group sequential analysis for this thesis. Building the CACTUS model `R` allows the exploration of the impact of a group sequential design on the health economic analysis in Chapter 7 and incorporates existing `R` code developed as part of the SAVI platform, discussed in Section 2.8.3, to explore the extension of EVSI methods to the adaptive design setting in Chapter 8.

To facilitate the development of the CACTUS health economic model in `R`, the `heemod` package (Markov Models for Health Economic Evaluations), developed by Filipović-Pierucci, Zarca and Durand-Zaleski is used (Filipović-Pierucci *et al.*, 2017). The package is used to define the model

parameters summarised in Table 5.2, transition probabilities described in Section 5.6.2 and costs and benefits associated with each health states as summarised in Section 5.6.3.

There are a small number of differences and simplifications between the original `Excel` model and the `R` model developed.

1. The `R` model does not breakdown the different components of the costs as in Table 5.2. Instead, a single cost was used to include all components of the research costs for each arm and a single parameter for the cost of the intervention.

2. Bootstrapping is used to generate a PSA sample rather than the parametric approach used in the Latimer *et al.*, 2013 analysis as this maintains the correlation structure of the variables which is important in Chapters 7 and 8.

3. The `R` model assumes that participants move states at the end of a cycle, whereas in the `Excel` model the half cycle correction is used that assumes movement between states takes place during the cycle.

For a comparison of the results from the `R` model and the `Excel` model the CACTUS pilot health economic analysis is conducted using the `R` model. As summarised in Table 5.3, there are minimal differences between the `Excel` and `R` model results for the deterministic analysis. The INB is equal to £2,363.13 and £2,380.44 for the `Excel` model and `R` model respectively. There were greater differences between the results of the two probabilistic analyses. The probability of cost-effectiveness in the `R` and `Excel` model both approximately 75% and a per patient EVPI of £143.68 and £167.68 for the `Excel` and `R` models, based on 500 and 1,000 samples respectively.

As there are number of differences in the composition and estimation methods between the two models, such as the bootstrapping and parametric probabilistic sensitivity analyses it is not expected that these results would be the same. However, these results show that the `R` model provides a similar replication of the health economic analysis conducted in the original CACTUS pilot trial that can be used in the remainder of the thesis to conduct a health economic analysis representative of the original Latimer *et al.*, 2013 analysis.

	Per person treated		Incremental					
	Costs	**QALYs**	**Costs**	**QALYs**	**ICER**	**INB**	**Prob CE**	**EVPI (pp)**
Probabilistic `Excel` Model								
Control	18608	3.08	-	-	-	-	-	-
Intervention	19020	3.32	411.40	0.24	1703.45	4388.60	75.8%	143.68
Probabilistic `R` Model								
Control	18698	3.08	-	-	-	-	-	-
Intervention	19137	3.26	437.94	0.18	2414.00	3190.42	75.6%	167.68
Deterministic `Excel` Model								
Control	18687	3.07	-	-	-	-	-	-
Intervention	19124	3.22	436.87	0.14	3058.21	2363.13	-	-
Deterministic `R` Model								
Control	18689	3.07	-	-	-	-	-	-
Intervention	19126	3.21	437.28	0.14	3103.76	2380.44	-	-

TABLE 5.3: Comparison of Excel Cactus model and R Cactus model for a willingness to pay threshold of £20,000 per QALY. PSA; probabilistic sensitivity analysis, QALY; quality adjusted life year, ICER; incremental cost-effectiveness ratio, INB; incremental net benefit, ProbCE; probability of cost-effectiveness, EVPI(pp); per person expected value of perfect information

5.8 CACTUS and the Case Study Criteria

The CACTUS pilot trial, Big CACTUS and the pilot health economic model (referred to collectively as the CACTUS case study) were selected as a suitable case study for the exploration of the use of health economics in the design and analysis of adaptive clinical trials. The following sections outline how the CACTUS case study met the three key criteria; characteristics, feasibility and generalisability.

5.8.1 Characteristics

In the Big CACTUS trial, it took participants 6-months to reach the primary outcome. It was planned for participants to be recruited over an 18-month period (Palmer *et al.*, 2015). While slightly larger than the guidance of Sully *et al.*, 2014 it was still deemed possible for the Big CACTUS trial to have used a group sequential design.

The Big CACTUS study was preceded by the CACTUS pilot trial and hence provided a body of pilot work, with a health economic analysis, that could be used to consider alternative designs for a future clinical trial (on the basis that Big CACTUS had not taken place).

5.8.2 Feasibility

As the Big CACTUS trial was ongoing during the thesis the CACTUS team (including Dr Rebecca Palmer, Dr Nicholas Latimer and the study manager Ms Liz Cross) were available for discussions throughout. Permission was granted for use of the CACTUS pilot data and health economic model by the chief investigator Dr Rebecca Palmer and lead health economist Dr Nicholas Latimer. A data sharing agreement was put in place and ethics approval for the use of the data obtained (ScHARR ethics committee ref:014510).

The Big CACTUS trial data were not available until late 2018 and so did not meet the criteria for use in the thesis. Instead, the pilot data was used to simulate trial data representative of a full-scale trial. Further work could explore the methods discussed in Chapter 6 applied to the Big CACTUS trial data.

The health economic analysis for the CACTUS pilot trial (described in more detail in Section 5.6) was developed by Dr Latimer using `Excel`. This model was simple in its structure and it was therefore reproducible in `R`.

5.8.3 Generalisability

The CACTUS pilot health economic analysis included both within trial and model based analyses, allowing for exploration of how both analyses are affected by the adaptive trials. The analyses were also conducted following NICE guidance for technology appraisals (National Institute for Health and Care Excellence, 2013a). This is a standard set of procedures recommended by NICE when a health technology is to be assessed and recommendations made about whether the technology should be made available on the NHS. Consequently, most UK based trials in both the public and private sectors will follow these procedures.

5.9 Chapter Summary

The CACTUS pilot trial, health economic model and Big CACTUS trial form the CACTUS case study for use in this thesis. The CACTUS case study met the key criteria of characteristics, feasibility and generalisability. The original `Excel` model has been re-created using `R` to allow greater flexibility in the methods development than is offered in `Excel`. The `Excel` and `R`

model give similar results when applied to the CACTUS pilot data, hence providing a valid `R` model for use in subsequent chapters.

The CACTUS case study is used in Chapter 6 to extend the existing theory for the adjustment of a group sequential design to the health economic context. The case study is then used to inform a simulation study in Chapter 7 that operationalises the extended theory. The case study is also used in Chapter 8 to explore how EVSI methods can be used to guide the design of a group sequential trial.

Chapter 6

Adjustment of a Health Economic Analysis following a Group Sequential Trial

6.1 Introduction

It is not appropriate to use the same analysis methods as a fixed sample size design when analysing data from an adaptive clinical trial, as discussed in Chapter 2. Adjustment methods are available that account for the potential bias introduced by the adaptive design, so that reliable inferences can be made following the trial. In this thesis, as described in Section 2.5.1, Bias Adjusted Maximum Likelihood Estimates (BAMLE) are considered for primary and secondary outcomes and the Sample Mean Ordering (SMO) approach is used for adjusted confidence intervals (described in Section 2.5.4.2).

Stevely *et al.*, 2015 found that only 7% ($^3/_{46}$) of the group sequential trials which they identified reported appropriately adjusted primary outcomes. The review in Chapter 3 found that no trials with an adaptive design and a health economic element discussed adjusting their analysis of primary or health economic outcomes for the adaptive nature of the trial. Marschner *et al.*, 2019 discuss how bias may be introduced into a health economic analysis following an adaptive design. There is, however, no discussion of the potential for bias in secondary outcomes from the trial data and a limited assessment of potential bias in the cost-effectiveness analysis.

Failing to account for the adaptive nature of a group sequential design in a subsequent health economic analysis may introduce bias into healthcare decision making. This could penalise patients who cannot receive the treatment they need as something that is not cost-effective is being funded instead. On this basis, it is important to investigate further the impact a group sequential design might have on the health economic analysis following the trial.

6.2 Chapter Aims

This chapter considers how bias adjustments can be operationalised in a health economic analysis following a group sequential design, to achieve the third research aim of this thesis; to investigate the impact a group sequential design might have on the health economic analysis following a trial. Existing theory introduced in Chapter 2 is extended to consider how adjustments can be made to the point estimates and confidence intervals of a within trial and model based health economic analysis. The CACTUS health economic model from Chapter 5 is used as a case study.

6.3 Existing Bias Adjustment Methods

Table 6.1 summarises the BAMLE and SMO approaches for calculating adjusted point estimates and confidence intervals for primary and secondary outcomes from an adaptive clinical trial, as described in Chapter 2. Existing theory also provides adjustments for the estimate of the absolute primary and secondary outcomes building on work by Whitehead, 1997 and Skalland, 2015. As identified in Chapter 3, this theory has not been considered in the context of a health economic analysis following a group sequential design. The following sections discuss how these methods can be applied and extended to a within trial and a model based analysis that requires some model parameters to be estimated from trial data.

6.4 Adjustments for a Within Trial Health Economic Analysis

A within trial health economic analysis is considered (described in Section 2.7.6) that uses the quality adjusted life year (QALY), as a measure of benefit of an intervention, and cost data from a clinical trial to estimate the short term cost-effectiveness of an intervention. Cost-effectiveness is estimated using the incremental net benefit (INB).

Estimate	**Method**	**Notes**
Primary point estimate	$\tilde{\theta}_1 = \hat{\theta}_1 - b(\theta_1)$	BAMLE (Equation 2.1)
Primary percentiles (95% CI)	$(\tilde{\theta}_{1,L}, \tilde{\theta}_{1,U})$	SMO of the sample space
Secondary point estimate	$\hat{\theta}_2 - \rho\frac{\sigma_2}{\sigma_1}(\hat{\theta}_1 - \tilde{\theta}_1)$	BAMLE ρ and σ are pooled estimates from the individual level data (Equation 2.5)
Secondary outcome percentiles (95% CI)	$(\tilde{\theta}_{2,L}, \tilde{\theta}_{2,U})$	SMO of the sample space
Absolute primary point estimate	$\tilde{\nu}_{1,C} = \bar{\nu}'_1 - \frac{n_I}{n}\tilde{\theta}_1$ and $\tilde{\nu}_{1,I} = \bar{\nu}'_1 + \frac{n_C}{n}\tilde{\theta}_1$	BAMLE $\bar{\nu}'_1$ is the mean for the primary outcome regardless of treatment arm (Equation 2.6)
Absolute primary percentiles (95% CI)	$\tilde{\pi}_{1,C} = \bar{\pi}'_1 - \frac{n_I}{n}\tilde{\theta}_1$ and $\tilde{\pi}_{1,I} = \bar{\pi}'_1 + \frac{n_C}{n}\tilde{\theta}_1$	$\bar{\pi}'_1$ is the chosen percentile of the population distribution for the primary outcome regardless of treatment arm
Absolute secondary point estimate	$\tilde{\nu}_{2,C} = \bar{\nu}'_2 - \frac{n_I}{n}\tilde{\theta}_2$ and $\tilde{\nu}_{2,I} = \bar{\nu}'_2 + \frac{n_C}{n}\tilde{\theta}_2$	$\bar{\nu}'_2$ is the mean for the secondary outcome regardless of treatment arm (Equation 2.6)
Absolute secondary percentiles (95% CI)	$\tilde{\pi}_{2,C} = \bar{\pi}'_2 - \frac{n_I}{n}\tilde{\theta}_2$ and $\tilde{\pi}_{1,I} = \bar{\pi}'_2 + \frac{n_C}{n}\tilde{\theta}_2$	$\bar{\pi}'_2$ is the chosen percentile of the population distribution for the primary outcome regardless of treatment arm

TABLE 6.1: Summary of existing adjustment methods for analysis of a trial following a group sequential design where θ_1 and θ_2 are primary and secondary outcomes respectively with standard deviations σ_1, σ_2 and correlation ρ, n_I and n_C are sample sizes in the intervention and control arm of a trial and n is the total sample size. CI: confidence interval

The key parameters for the within trial analysis can be estimated using the secondary outcome point estimate and confidence interval approaches, summarised in Table 6.1, by applying the existing theory to the health economic context. The adjustments to the within trial health economic analysis are generalisable to all trials collecting cost and benefit data. The following sections define the health economic parameters and then discuss how adjusted and unadjusted point estimates and confidence intervals can be estimated.

6.4.1 Health Economic Parameters

Let $x_{1I,j}$ and $x_{1C,j}$ represent the observed primary outcome for participant j in the intervention and control arm respectively, with underlying true means $\mu_{1,I}$ and $\mu_{1,C}$ and standard deviation σ_1 in each arm. The true difference in mean response in the primary outcome between the two treatment arms is given by

$$\theta_1 = \mu_{1I} - \mu_{1C}. \tag{6.1}$$

Let $x_{2I,j}$ and $x_{2C,j}$ represent the observed QALY for participant j in the intervention and control arm respectively, with means $\mu_{2,I}$ and $\mu_{2,C}$ and standard deviation σ_2 in each arm. The true difference in mean QALY between the two treatment arms is given by

$$\theta_2 = \mu_{2I} - \mu_{2C}. \tag{6.2}$$

Let $x_{3I,j}$ and $x_{3C,j}$ represent the observed total costs for participant j in the intervention and control arm respectively, with means $\mu_{3,I}$ and $\mu_{3,C}$ and standard deviation σ_3 in each arm. The true difference in mean costs between the two treatment arms is given by

$$\theta_3 = \mu_{3I} - \mu_{3C}. \tag{6.3}$$

Let $x_{4I,j}$ and $x_{4C,j}$ be the observed net benefit for participant j in the intervention and control arm respectively. This is calculated using

$$x_{4I,j} = \lambda x_{2I,j} - x_{3I,j}, \tag{6.4}$$

$$x_{4C,j} = \lambda x_{2C,j} - x_{3C,j}, \tag{6.5}$$

where λ is the willingness to pay threshold. Here a threshold of £20,000 per QALY is used as per NICE guidance (National Institute for Health and Care Excellence, 2013b). The underlying true mean net benefits are $\mu_{4,I}$ and $\mu_{4,C}$ and standard deviation σ_4 in each arm. The true INB between the two treatment arms is given by

$$\theta_4 = \mu_{4,I} - \mu_{4,C}. \tag{6.6}$$

6.4.2 Unadjusted Within Trial Analysis

In this section, the methods for conducting a within trial health economic analysis following a group sequential design that does not account for the adaptive nature of the trial are outlined. This approach is commonly used in practice, as shown in Chapter 3, where none of the identified trials appeared to adjusted their health economic analysis when it was deemed necessary. This is referred to as the unadjusted analysis. The same methods are used as if the trial has used a fixed sample size design, essentially ignoring the potential impact of the adaptive nature of the trial.

6.4.2.1 Unadjusted Point Estimates

An unadjusted estimate of $\hat{\theta}_1$, the mean difference in the primary outcome, is calculated using the MLE and is given by

$$\hat{\theta}_1 = \sum_{j=1}^{n_I} \frac{x_{1I,j}}{n_I} - \sum_{j=1}^{n_C} \frac{x_{1C,j}}{n_C}, \tag{6.7}$$

where n_I and n_C are the number of participants in the intervention and control arms respectively and $x_{1I,j}$, $x_{1C,j}$ are the absolute primary outcome for participant j in the intervention arm and control arm respectively.

An unadjusted estimate of $\hat{\theta}_2$, the difference in QALY, is calculated using the MLE and is given by

$$\hat{\theta}_2 = \sum_{j=1}^{n_I} \frac{x_{2I,j}}{n_I} - \sum_{j=1}^{n_C} \frac{x_{2C,j}}{n_C}. \tag{6.8}$$

where $x_{2I,j}$, $x_{2C,j}$ are the absolute QALYs for participant j in the intervention arm and control arm respectively. As discussed in Section 2.7.3, QALYs are typically calculated using EQ-5D scores measured at multiple time points during the trial and linear interpolation.

An unadjusted estimate of $\hat{\theta}_3$, the difference in costs, is calculated using the MLE and is given by

$$\hat{\theta}_3 = \sum_{j=1}^{n_I} \frac{x_{3I,j}}{n_I} - \sum_{j=1}^{n_C} \frac{x_{3C,j}}{n_C}, \tag{6.9}$$

where $x_{3I,j}$, $x_{3C,j}$ are the absolute costs for participant j in the intervention arm and control arm respectively.

An unadjusted estimate of $\hat{\theta}_4$, the INB, is calculated using the MLE and is given by

$$\hat{\theta}_4 = \sum_{j=1}^{n_I} \frac{x_{4I,j}}{n_I} - \sum_{j=1}^{n_C} \frac{x_{4C,j}}{n_C}, \tag{6.10}$$

where $x_{4I,j}$, $x_{4C,j}$ are the absolute net benefits for participant j in the intervention arm and control arm respectively.

6.4.2.2 Unadjusted Confidence Intervals

An unadjusted 95% confidence interval is calculated for each parameter using the Normality assumption. This assumption is valid for large sample sizes by the central limit theorem (Swinscow, 1997). The confidence interval formula for the primary outcome (θ_1) is given by

$$\hat{\theta}_{1,L} = \hat{\theta}_1 - \Phi^{-1}\left(1 - \frac{\alpha}{2}\right)\left(\sqrt{\frac{1}{n_I} + \frac{1}{n_c}}\right)\hat{\sigma}_1, \tag{6.11}$$

$$\hat{\theta}_{1,U} = \hat{\theta}_1 + \Phi^{-1}\left(1 - \frac{\alpha}{2}\right)\left(\sqrt{\frac{1}{n_I} + \frac{1}{n_c}}\right)\hat{\sigma}_1. \tag{6.12}$$

where

$\hat{\theta}_1$ is the MLE,

$\hat{\sigma}_1$ is the standard deviation estimated directly from the trial data,

n_I is the sample size in the intervention arm,

n_C is the sample size in the control arm,

Φ^{-1} is the inverse of the standard Normal distribution,

α is the significance level.

The pooled standard deviation $\hat{\sigma}_1$ is using

$$\hat{\sigma}_1 = \sqrt{\frac{\hat{\sigma}^2_{1,I} + \hat{\sigma}^2_{1,C}}{2}}. \tag{6.13}$$

Equation 6.11 and Equation 6.12 can be used to calculate unadjusted confidence intervals for the other health economic parameters by replacing $\hat{\theta}_1$ and $\hat{\sigma}_1$ with the trial estimates of the mean and standard deviation for the given parameter.

6.4.3 Adjusted Within Trial Analysis

In this section, the methods for conducting a within trial health economic analysis following a group sequential design that account for the adaptive nature of the trial are outlined. This approach has not been considered in practice to date as shown in Chapter 3, however, the existing theory introduced in Chapter 2 can be extended to the within trial health economic analysis. This analysis is referred to as the adjusted within trial health economic analysis. The adjusted approach aims to reduce the bias in the point estimates and give confidence intervals with desirable properties such as coverage close to the nominal value (for example coverage close to 0.95 for a 95% confidence interval) and narrow width.

6.4.3.1 Adjusted Point Estimates

An adjusted estimate of the mean difference in the primary outcome is calculated using the BAMLE described in Chapter 2.5.1 (Equation 2.1) and is given by

$$\tilde{\theta}_1 = \hat{\theta}_1 - b(\theta_1), \tag{6.14}$$

where

$\hat{\theta}_1$ is the unadjusted MLE from Equation 6.7 ,

$b(\theta_1)$ is an estimate of the bias.

An adjusted estimate of the mean difference in QALY is calculated using the BAMLE and is given by

$$\tilde{\theta}_2 = \hat{\theta}_2 - \hat{\rho}_{12}\frac{\hat{\sigma}_2}{\hat{\sigma}_1}(\hat{\theta}_1 - \tilde{\theta}_1), \tag{6.15}$$

where

$\hat{\rho}_{12}$ is the estimated pooled Spearman rank correlation between the primary outcome and QALY,

$\hat{\sigma}_1$ is the estimated pooled standard deviation for the primary outcome,

$\hat{\sigma}_2$ is the estimated pooled standard deviation for QALY.

In this thesis, the pooled correlation $\hat{\rho}_{1,2}$ is approximated using

$$\hat{\rho}_{1,2} = \frac{1}{n_I + n_C}\left(n_I\rho_{1,2,I} + n_C\rho_{1,2,C}\right). \tag{6.16}$$

Alternatively, the pooled correlation could be calculated using the Fisher transformation to account for the skew in the distribution of correlation coefficients (Alexander, 1990).

As discussed in Chapter 2, Spearman rank correlations are used throughout the thesis. When outcomes are not Normally distributed, such as the logNormal distribution, the Normality assumption required for calculating a Pearson correlation coefficient is violated (Briggs *et al.*, 2006). The Spearman correlation coefficient however, is a rank correlation and not affected by

non-Normality. The Spearman rank correlation is, therefore, used health economic data is often non-Normal (Drummond *et al.*, 2015).

An adjusted estimate of the mean difference in costs is calculated using the BAMLE and is given by

$$\tilde{\theta}_3 = \hat{\theta}_3 - \hat{\rho}_{13}\frac{\hat{\sigma}_3}{\hat{\sigma}_1}(\hat{\theta}_1 - \tilde{\theta}_1), \tag{6.17}$$

where

$\hat{\rho}_{13}$ is the estimated pooled Spearman rank correlation between the primary outcome and cost,

$\hat{\sigma}_1$ is the estimated pooled standard deviation for the primay outcome,

$\hat{\sigma}_3$ is the estimated pooled standard deviation for cost.

An adjusted estimate of the INB is calculated using the BAMLE and is given by

$$\tilde{\theta}_4 = \hat{\theta}_4 - \hat{\rho}_{14}\frac{\hat{\sigma}_4}{\hat{\sigma}_1}(\hat{\theta}_1 - \tilde{\theta}_1), \tag{6.18}$$

where,

$\hat{\rho}_{14}$ is the estimated pooled Spearman rank correlation between the primary outcome and net benefit,

$\hat{\sigma}_1$ is the estimated pooled standard deviation for the primary outcome,

$\hat{\sigma}_4$ is the estimated pooled standard deviation for net benefit.

6.4.3.2 Adjusted Confidence Intervals

An adjusted 95% confidence interval can be calculated for the primary outcome (θ_1) using the SMO approach described in Section 2.5.4.2 and denoted by $(\tilde{\theta}_{1,L}, \tilde{\theta}_{1,U})$. Adjusted confidence intervals can be calculated for the health economic parameters using the approach of Skalland, 2015, described in Section 2.5.4.2.

Practically, $\hat{\theta}_1$ (the unadjusted estimate of the primary outcome), $\hat{\sigma}_1$ (an estimate pooled standard deviation for the primary outcome), $\hat{\sigma}_2$ (an estimate of the pooled standard deviation for

the secondary outcome), $\hat{\rho}_{1,2}$ (an estimate of the pooled correlation between the primary and secondary outcome) and n_k (the number of participants recruited between each analysis) are extracted from the terminal analysis dataset of the completed trial. The mean of the secondary outcome $\hat{\theta}_2$ is set to zero.

Simulations are then used to generate estimates of the mean primary and secondary outcome for the group of participants recruited between each analysis of the trial. The following bivariate Normal distribution is used, assuming an equal allocation of participants between treatment arms

$$BVN\left(\begin{pmatrix}\hat{\theta}_1\\ \hat{\theta}_2\end{pmatrix}, \left(\frac{1}{\frac{n_k}{2}}+\frac{1}{\frac{n_k}{2}}\right)\begin{pmatrix}\hat{\sigma}_1^2 & \hat{\rho}_{1,2}\hat{\sigma}_1\hat{\sigma}_2\\ \hat{\rho}_{1,2}\hat{\sigma}_1\hat{\sigma}_2 & \sigma_2^2\end{pmatrix}\right). \tag{6.19}$$

For a trial with five analyses of the data this gives

$$\text{baseline to analysis 1 } (\theta_{1,k_1}, \theta_{2,k_1}), \tag{6.20}$$

$$\text{analysis 1 to analysis 2 } (\theta_{1,k_2}, \theta_{2,k_2}), \tag{6.21}$$

$$\text{analysis 2 to analysis 3 } (\theta_{1,k_3}, \theta_{2,k_3}), \tag{6.22}$$

$$\text{analysis 3 to analysis 4 } (\theta_{1,k_4}, \theta_{2,k_4}), \tag{6.23}$$

$$\text{analysis 4 to analysis 5 } (\theta_{1,k_5}, \theta_{2,k_5}), \tag{6.24}$$

where θ_{1,k_1} is the sample mean for the primary outcome for participants recruited from baseline to analysis one and θ_{1,k_5} is the sample mean for participants recruited from analysis four to five.

Based on these simulated means, the cumulative mean primary and secondary outcome at each analysis is calculated. For the five analysis design this is calculated using

$$\text{analysis 1 } (\theta_{1,k_1}, \theta_{2,k_1}), \tag{6.25}$$

$$\text{analysis 2 } \left(\frac{(\theta_{1,k_1}+\theta_{1,k_2})}{2}, \frac{(\theta_{2,k_1}+\theta_{2,k_2})}{2}\right), \tag{6.26}$$

$$\vdots \tag{6.27}$$

$$\text{analysis 5 } \left(\frac{(\theta_{1,k_1}+\theta_{1,k_2}+\theta_{1,k_3}+\theta_{1,k_4}+\theta_{1,k_5})}{5}, \frac{(\theta_{2,k_1}+\theta_{2,k_2}+\theta_{2,k_3}+\theta_{2,k_4}+\theta_{2,k_5})}{5}\right). \tag{6.28}$$

The cumulative primary outcome mean at each analysis is then compared to the stopping boundary used in the original trial, noting the analysis where the mean crosses the boundary. The cumulative mean for the secondary outcome at this same analysis is recorded and this process repeated N times to give a matrix of N secondary outcome point estimates.

The aim of this process is to understand the distribution of the secondary outcome mean under a range of true values. A vector of true values, denoted by θ_2^*, is selected. This vector should cover a wide range of plausible values at small increasing intervals. Each of the N secondary outcome means is then added to the grid values. For example if $\theta_2^* = (-2.00, -1.99, \ldots, 1.99, 2.00)$ and the secondary outcome mean for the simulation $N = 1$ is equal to 0.05 the grid vector for this simulation becomes $(-1.95, -1.94, -1.93 \ldots, 2.03, 2.04, 2.05)$. This is repeated for each of the N grid vectors giving the distribution of the sample mean of the secondary outcome where the true mean is equal to θ_2^*

A one-sided p-value is calculated by comparing the $\hat{\theta}_2$ observed in the terminal analysis set of the trial to the grid vectors and calculating the probability that the grid vector value is greater than that observed. A two-sided p-value is calculated using

$$p = 2\text{min}\big\{P(0), 1 - P(0)\big\}. \tag{6.29}$$

The 97.5^{th} percentile is the maximum value of the grid values where the two-sided p-value is greater than 0.05. The 2.5^{th} percentile is the minimum value of the grid values where the two-sided p-value is greater than 0.05.

These practical steps for conducting the Skalland approach are summarised in Figure 6.1. `R` code provided by Dr Skalland has been adapted for use in the simulation study to allow estimation of adjusted confidence intervals for the designs under consideration.

6.5 Adjustments for a Model Based Health Economic Analysis

As outlined in Section 2.7.7, a model-based health economic analysis can be used to assess the long-term cost-effectiveness of an intervention. In the CACTUS case study (Chapter 5), a three state Markov model is used to assess the long-term cost-effectiveness of the computer-based intervention compared to usual care. The following sections discuss how the bias adjustment

FIGURE 6.1: Summary of the steps required for the calculation of adjusted confidence intervals for health economic outcomes using the Skalland, 2015 approach

1. From the terminal analysis dataset from the completed trial extract
   (a) $\hat{\theta}_1$ - MLE of the primary outcome
   (b) $\hat{\sigma}_1$ - estimate pooled standard deviation for the primary outcome
   (c) $\hat{\sigma}_2$ - estimate of the pooled standard deviation for the secondary outcome
   (d) $\hat{\rho}_{1,2}$ - estimate of the pooled correlation between the primary and secondary outcome
   (e) n_k - the number of participants recruited between each analyses

   and set the mean of the secondary outcome $\hat{\theta}_2$ to be zero.

2. Using a bivariate Normal distribution sample a mean primary and secondary outcome for the group of participants recruited between each analysis of the trial.

3. Calculate the cumulative mean primary and secondary outcome at each interim analysis.

4. Compare the cumulative primary outcome mean at each interim analysis to the stopping boundary used in the original trial. Record the primary outcome and secondary outcome cumulative means at the analysis that the trial crosses the boundary or the final analysis is the boundary is not crossed at an earlier analysis.

5. Repeat steps 2 to 4 N times to give a matrix of N primary and secondary point estimates.

6. Calculate a vector of true values for the secondary outcome denoted by θ_2^*. This vector should cover a wide range of plausible values at small increasing values.

7. Each of the N secondary outcome means is then added to the grid values. Each giving the distribution of the sample mean of the secondary outcome where the true mean is equal to θ_2^*.

8. A one-sided p-value is calculated by comparing $\hat{\theta}_2$ observed in the terminal analysis set of the trial (step 1) to the grid vectors and calculating the probability that the grid vector value is greater than that observed.

9. A two-sided p-value is calculated using Equation 2.9

$$p = 2\text{min}\big\{P(0), 1 - P(0)\big\}.$$

10. The 97.5^{th} percentile is the maximum value of the grid values where the two-sided p-value is greater than 0.05.

11. The 2.5^{th} percentile is the minimum value of the grid values where the two-sided p-value is greater than 0.05.

methods described in Section 6.3 can be extended to health economic model parameters. The CACTUS case study is used to motivate the development of this theory, outlining the adjustments of the model parameters required for the model-based analysis. However, the extension of the bias adjustment theory is generalisable to other contexts where similar summaries of trial data are required for a health economic model.

The CACTUS model-based analyses used five parameters estimated from the trial as well as data from external sources, as summarised in Table 5.2. Each parameter is discussed in turn, recapping its definition and discussing how the adjusted parameter estimate can be calculated. Proposed approaches for adjusted confidence intervals are then discussed.

Adjusted and unadjusted estimates of the long-term cost-effectiveness of the intervention are estimated by evaluating the health economic model. This is carried out using adjusted and unadjusted parameter estimates respectively to give an adjusted and unadjusted deterministic INB.

6.5.1 Probability of a Good Response

As discussed in Chapter 5, the probability of good response at 6-months is calculated from the trial data. This is defined to be the proportion of participants in the intervention arm who have a 'good' response to the computer-based intervention. As discussed in Section 5.6.2, a 'good' response was defined to be an improvement from baseline to follow-up in the percentage words named correctly of 17% or more in the CACTUS pilot health economic analysis.

6.5.1.1 Point Estimate

Let

$\nu_{6,I}$ be the true probability of good response at 6-months,

$\hat{\nu}_{6,I}$ be the unadjusted estimate,

$\tilde{\nu}_{6,I}$ be the bias adjusted estimate.

The unadjusted estimate is calculated from the trial data by taking all participants in the intervention arm and counting the number with an improvement in percentage words named correctly between baseline to 6-months of 17% or more and dividing by the total number of

participants in the intervention arm. This is calculated using

$$\hat{\nu}_{6,I} = \frac{\sum_{j=1}^{n_I} \mathbb{1}(x_{1I,j} \geq 0.17)}{n_I}, \tag{6.30}$$

where

n_I is the number of participants in the intervention arm,

$\mathbb{1}$ is the indicator function,

$x_{1I,j}$ is the improvement in percentage words named correctly between baseline and 6-months in the intervention arm.

It is not possible to use the adjustment methods given in Table 6.1. There are two issues with these existing methods

1. The model parameter is a marginal probability, however, the primary outcome is based on a difference in mean values.

2. The model parameter is a marginal probability for responders rather than a marginal probability for the intervention or control arms.

Using Equation 2.6, it is possible to estimate an adjusted mean percentage improvement in words named correctly for participants in the intervention or control arm using the adjusted estimate of the primary outcome $\tilde{\theta}_1$ (difference in percentage improvement between the intervention and control arms). If the primary outcome had been the log odds ratio for a response in each arm it would be possible to adapt Equation 2.6 to give the marginal probablity of response in each arm as described by Whitehead, 1997. However, Whitehead, 1997 do not suggest a method for estimating the probability of response in each arm when the primary outcome is based on a difference in means. Additionally, there is no solution when the marginal estimates are required for responders and non-responders across the whole trial and not by treatment arm.

In this thesis, I have developed the following approach to calculate an approximate adjusted estimate. This works by sampling from the distribution of the improvement in percentage words named correctly between baseline and 6-months in the intervention arm. However, the

distribution is affected by the adaptive nature of the trial and this needs to be accounted for in the calculation. Therefore adjusted percentiles are calculated to describe the distribution. Values are then sampled from the distribution and the proportion of sampled values greater than or equal to 17% are used to estimate an adjusted probability of a good response at 6-months in the intervention arm.

Adjusted percentiles can be calculated using the following steps:

1. Use bootstrapping to calculate the percentiles of the population distribution for the improvement in percentage words named correctly at 6-months across all participants regardless of treatment arm by,

   (a) Repeatedly sample the data with replacement $b = 1, \ldots, B$ times.

   (b) Calculate the percentiles for each sample $(\pi'_{0.01,b}, \ldots, \pi'_{0.99,b})$.

   (c) Calculate the mean for each percentile $(\bar{\pi}'_{0.01}, \ldots, \bar{\pi}'_{0.99})$.

2. Calculate the percentiles in the intervention arm that are adjusted for the adaptive nature of the trial using the methods described in Table 6.1.

$$(\bar{\pi}_{I,0.01} = \bar{\pi}'_{0.01} + \frac{n_C}{n}\tilde{\theta}_1, \ldots, \bar{\pi}_{I,0.99} = \bar{\pi}'_{0.99} + \frac{n_C}{n}\tilde{\theta}_1), \tag{6.31}$$

   where $\tilde{\theta}_1$ is the BAMLE of the primary outcome, n_C is the number of participants in the intervention arm and $\bar{\pi}'_{0.01}$ is the 1^{st} percentile from the population distribution from step 1.

3. Sample from the population distribution for the improvement in percentage words named correctly at 6-months

   (a) Create a 'look-up' table of percentiles and the estimated parameter value at the percentile.

   (b) Randomly sample values from a Uniform distribution between zero and one $(U(0, 1))$.

(c) Choose the estimated parameter value associated with the percentile closest to the sampled Uniform value. For example, if the rounded sampled uniform value is 0.5 then the sampled parameter value will correspond to the 50^{th} percentile.

(d) Sampled parameter values then form a sampled dataset of values from the chosen distribution.

4. Calculate the proportion of times the treatment effect is greater than or equal to 17% in the sampled values.

This process is repeated to give the probability of a good response at 9-months by using the number of participants in the intervention arm with an improvement in percentage words of greater than or equal to 17% between baseline and 9-months. In this case, let

$\nu_{9,I}$ be the true probability of a good response at 9-months,

$\hat{\nu}_{9,I}$ be the unadjusted estimate,

$\tilde{\nu}_{9,I}$ be the bias adjusted estimate.

This approach can be adapted and applied to the estimation of health economic parameters that do not meet the requirements for the existing adjustment methods in other trial settings.

6.5.1.2 Confidence Interval

An adjusted 95% confidence interval for the probability of a good response can be approximated using bootstrapping as follows:

1. Sample trial data with replacement for each arm independently;

2. Apply the trial analysis for the chosen design to see when the trial would have stopped early;

3. Calculate an adjusted probability of a good response using the steps outlined in Section 6.5.1;

4. Repeat 1000 times;

5. Rank the adjusted probabilities of good response;

6. Take the 25^{th} and 975^{th} values as the lower and upper adjusted confidence intervals.

These values are denoted by

$$(\tilde{\nu}_{6,I,L}, \tilde{\nu}_{6,I,U}) \text{ for 6-months follow-up,}$$
$$(\tilde{\nu}_{9,I,L}, \tilde{\nu}_{9,I,U}) \text{ for 9-months follow-up.}$$

6.5.2 Relapse Rate

The relapse rate is defined to be the probability of a good response at 6-months minus the probability of a good response at 9-months as described in Section 5.6.2. Let

θ_7 be the relapse rate,

$\hat{\theta}_7$ be the unadjusted estimate,

$\tilde{\theta}_7$ be the bias adjusted estimate.

This is converted in to a monthly probability using

$$\theta_7 = 1 - e^{\{-(-\frac{1}{4}log(\nu_{6,I} - \nu_{9,I}))\}}. \tag{6.32}$$

If $\theta_7 < 0$ this is set to zero as a negative relapse rate is not possible.

6.5.2.1 Point Estimate

The unadjusted relapse rate ($\hat{\theta}_7$) is calculated using the unadjusted probabilities of a good response at 6 and 9-months in Equation 6.32. The adjusted estimate ($\tilde{\theta}_7$) uses the adjusted probabilities at 6 and 9-months in Equation 6.32.

6.5.2.2 Confidence Interval

An adjusted 95% confidence interval for the relapse rate can be approximated using bootstrapping as follows

1. Sample trial data with replacement for each arm independently.

2. Apply the trial analysis for the chosen design to see when the trial would have stopped early.

3. Calculate an adjusted probability of a good response at 6 and 9-months using the steps outlined in Section 6.5.1.

4. Calculate the adjusted relapse rate using the steps outlined in Section 6.5.2.1.

5. Repeat 1000 times.

6. Rank the adjusted probabilities of good response.

7. Take the 25^{th} and 975^{th} values as the lower and upper adjusted confidence intervals.

These values are denoted by $(\tilde{\theta}_{7,L}, \tilde{\theta}_{7,U})$.

6.5.3 Baseline Utility

Let $u_{b,C,j}$ be the observed utility score at baseline for participant j in the control arm, with mean $\mu_{ub,C}$ and standard deviation σ_u. The true mean baseline utility is denoted by $\nu_{8,b,C}$.

6.5.3.1 Point Estimate

The unadjusted point estimate is the MLE calculated directly from the trial data calculated using

$$\hat{\nu}_{8,b,C} = \sum_{j=1}^{n_I} \frac{u_{b,C,j}}{n_C}. \tag{6.33}$$

As this is a baseline parameter collected before the adaptive trial begins it is not necessary to adjust the estimate for the adaptive nature of the trial. Therefore the adjusted point estimate, denoted by $\tilde{\nu}_{8,b,C}$, is set to be equal to the unadjusted estimate,

$$\tilde{\nu}_{8,b,C} = \hat{\nu}_{8,b,C}. \tag{6.34}$$

6.5.3.2 Confidence Interval

The adjusted upper and lower limits of the 95% confidence interval are equal to the unadjusted estimates $(\hat{\nu}_{8,b,C,L}, \hat{\nu}_{8,b,C,U})$.

6.5.4 Utility Improvement

The utility improvement model parameter is defined to be the mean improvement in utility from baseline to 6-month follow-up for those participants that responded to treatment across both arms minus the mean utility improvement for non-responders across both arms. Participants were classed as 'responding' if they had an improvement in the percentage words named correctly of 17% or more. Let $u_{6,Ij}$ be the observed utility of participant j at 6-months follow-up in the intervention arm. The observed utility improvement for the participant is given by

$$u_{I,j} = u_{6,Ij} - u_{b,Ij}. \tag{6.35}$$

The true difference in mean utility improvement between responders and non-responders is denoted by θ_8^*.

6.5.4.1 Point Estimate

An unadjusted estimate of the utility improvement is calculated by determining which participants had a good response to treatment and then calculating the mean utility for responders and mean utility of non-responders, irrespective of treatment arm using the MLE,

$$\hat{\theta}_8^* = \sum_{j=1}^{n_R} \frac{u_{I,j}}{n_R} - \sum_{j=1}^{n_{NR}} \frac{u_{I,j}}{n_{NR}}, \tag{6.36}$$

where n_R is the number of responders and n_{NR} is the number of non-responders.

An adjusted utility improvement, denoted by $\tilde{\theta}_8$ is calculated using

$$\tilde{\theta}_8^* = \hat{\theta}_8^* - \hat{\rho}_{18}\frac{\hat{\sigma}_8}{\hat{\sigma}_1}(\hat{\theta}_1 - \tilde{\theta}_1), \tag{6.37}$$

where

$\hat{\rho}_{18}$ is the estimated pooled Spearman rank correlation
between the primary outcome and utility improvement,
$\hat{\sigma}_1$ is the estimated pooled standard deviation for the primary outcome,
$\hat{\sigma}_8$ is the estimated pooled standard deviation for utility improvement.

The standard deviation of improvement in utility ($\hat{\sigma}_8$) from baseline and 6-month follow-up between treatment arms (rather than between responders and non-responders) is used to calculate the adjusted estimate in Equation 6.37. This is a pooled standard deviation between the intervention and control arms and calculated in a similar way to the pooled standard deviation for the primary outcome ($\hat{\sigma}_1$). Likewise the correlation $\hat{\rho}_{18}$ is a pooled correlation between intervention arms rather than responders and non-responders. These estimates are considered an approximation to the estimates based on responders and non-responders and are used as this reflects how the adjusted and unadjusted primary outcome are calculated.

6.5.4.2 Confidence Interval

To estimate an adjusted confidence interval for utility improvement the Skalland approach described in Section 6.4.3.2 using the pooled correlation between the primary outcome and utility improvement between arms and the standard error of the utility improvement pooled across treatment arms as in the point estimate adjustment. These values are denoted by $(\tilde{\theta}^*_{8,L}, \tilde{\theta}^*_{8,U})$.

6.5.5 Resource Use

The resource use in the intervention and control arm of the trial includes all costs incurred during the trial by the NHS but not including the cost of the intervention itself.

6.5.5.1 Point Estimate

The unadjusted point estimates and confidence intervals are calculated directly from the trial data using the MLE and denoted by $\hat{\nu}_{10,I}, \hat{\nu}_{10,C}$. Adjusted estimates are calculated using Equation 2.6 and denoted by $\tilde{\nu}_{10,I}, \tilde{\nu}_{10,C}$.

6.5.5.2 Confidence Interval

Adjusted estimates of the upper and lower limits of the 95% confidence interval are calculated using the approach described in Section 6.4.3.2 and denoted by

$$(\widetilde{\nu}_{10,I,L}, \widetilde{\nu}_{10,I,U}) \text{ for the intervention arm,} \tag{6.38}$$

$$(\widetilde{\nu}_{10,C,L}, \widetilde{\nu}_{10,C,U}) \text{ for the control arm.} \tag{6.39}$$

6.5.6 Uncertainty Analysis

An important part of a health economic analysis is understanding the uncertainty in the results. When this analysis follows an adaptive design it is important to appropriately adjust approaches to allow for the adaptive nature of the trial.

6.5.6.1 Parametric Probabilistic Sensitivity Analysis

As described in Section 2.7.8, a parametric probabilistic sensitivity analysis (PSA) assigns a distribution to each of the parameters in the health economic model. However, when an adaptive design is used, the distribution of parameters estimated from the trial is affected by the adaptive nature of the trial. To accurately capture the distribution of the parameter adjusted percentiles can be calculated using the same steps used to calculate the adjusted probability of good response in Section 6.5.1 and summarised in Figure 6.2.

The steps described in Section 2.7.8.1 can then be followed (Saltelli *et al.*, 2000; Briggs *et al.*, 2006):

1. Assign a distribution to each of the parameters in the health economic model.

2. Randomly sample N_{PSA} values from the assigned distribution. Repeat for all parameters in the model to create a PSA sample.

3. Evaluate the output (usually the INB) for each of the N_{PSA} set of parameters.

FIGURE 6.2: Summary of the steps required to sample from an adjusted distribution using adjusted percentiles for the given parameter for use in a parametric probabilistic sensitivity analysis

1. Calculate the adjusted percentiles using the approaches described in Section 6.3 by varying the significance level from zero to one.
2. Create a 'look-up' table of percentiles and the estimated parameter value at the percentile.
3. Randomly sample values from a Uniform distribution between zero and one.
4. Choose the estimated parameter value associated with the percentile closest to the sampled Uniform value. For example, if the rounded sampled uniform value is then the sampled parameter value will correspond to the 1st percentile, if the sampled value is 0.5 then the sampled parameter value will correspond to the 50^{th} percentile.
5. Sampled parameter values then form a sampled dataset of values from the chosen distribution.

4. Use the sample of outputs to summarise its distribution for example the mean INB,

$$\frac{1}{N_{PSA}} \sum_{i=1}^{N_{PSA}} {\theta_4}^{(i)}. \tag{6.40}$$

This approach does not capture the correlation between the mode parameters; instead, it assumes they are all independent which is likely to be unrealistic. Additionally, the calculation of the adjusted percentiles and sampling process described in Figure 6.2 can be computationally intensive, especially if there are a large number of model parameters estimated from the trial to be included in the PSA.

6.5.6.2 Bootstrapped Probabilistic Sensitivity Analysis

An alternative to the parametric PSA is to use a bootstrapping approach, as discussed in Section 2.7.8. When the trial uses an adaptive design, the steps required for a bootstrapped PSA are as follows:

1. Sample data with replacement for each arm independently.
2. Apply the trial analysis for chosen design to see when the bootstrapped trial would have stopped.

3. Calculate adjusted or unadjusted model parameters.

4. Run health economic model with the bootstrapped parameters to give the INB.

5. Repeat N_{boot} times.

6. Use the sample of outputs to summarise its distribution.

Step 2 requires applying the trial analysis to the bootstrapped dataset. This step is an additional requirement to reflect the adaptive nature of the trial design. This can be computationally demanding, especially when repeated a large number of times. However, this approach captures the correlation structure between the variables and is less computationally demanding than calculating the adjusted percentiles for each of the model parameters to then estimate their distributions. This approach is adopted in the analyses in the simulation studies of Chapter 7 and Chapter 8.

6.6 Chapter Summary

In this chapter, the bias adjustment methods introduced in Chapter 2 have been extended to the context of a within trial and a model based health economic analysis following a group sequential design. All parameter estimates are summarised in Table 6.2. The within trial analysis requires a straightforward application of the existing BAMLE adjustment for secondary outcomes and the SMO approach for adjusted confidence intervals. These results are generalisable to any within trial health economic analysis that collects continuous outcome data on costs and benefits in each treatment arm.

The model based analysis, based on the CACTUS health economic model, described in Chapter 5, demonstrates how model parameters estimated from trial data may require adaptation of the existing adjustment methods. However, approximate adjustments can be made that account for the adaptive nature of the trial.

This chapter demonstrates how bias adjustments may be operationalised in a health economic analysis following a group sequential design, in part achieving the research aim to explore the potential for an adaptive design to introduce bias into the health economic analysis following a clinical trial. In Chapter 7 this research aim is explored further using a simulation study to

assess the extent to which bias may affect the health economic analysis and how well the adjustment methods described in this chapter control for this. The bias adjustments are then applied in Chapter 8 where EVSI analysis methods are used to guide the design of adaptive clinical trials to ensure the analysis appropriately accounts for the adaptive nature of the proposed trial designs.

Parameter	**Unadjusted Point Estimate (95% CI)**	**Adjusted Point Estimate (95% CI)**
	Within trial analysis	
Incremental net benefit	$\hat{\theta}_4(\hat{\theta}_{4,L}, \hat{\theta}_{4,U})$	$\tilde{\theta}_4(\tilde{\theta}_{4,L}, \tilde{\theta}_{4,U})$
Difference in Cost	$\hat{\theta}_3(\hat{\theta}_{3,L}, \hat{\theta}_{3,U})$	$\tilde{\theta}_3(\tilde{\theta}_{3,L}, \tilde{\theta}_{3,U})$
Difference in QALY	$\hat{\theta}_2(\hat{\theta}_{2,L}, \hat{\theta}_{2,U})$	$\tilde{\theta}_2(\tilde{\theta}_{2,L}, \tilde{\theta}_{2,U})$
Primary Outcome	$\hat{\theta}_1(\hat{\theta}_{1,L}, \hat{\theta}_{1,U})$	$\tilde{\theta}_1(\tilde{\theta}_{1,L}, \tilde{\theta}_{1,U})$
	Model-based analysis	
Probability of a good response at 6-months	$\hat{\nu}_6$ $(\hat{\nu}_{6,I,L}, \hat{\nu}_{6,I,U})$	$\tilde{\nu}_{6,I}$ $(\tilde{\nu}_{6,I,L}, \tilde{\theta}_{6,U})$
Relapse rate	$\hat{\theta}_7$ $(\hat{\theta}_{7,L}, \hat{\theta}_{7,U})$	$\tilde{\theta}_7$ $(\tilde{\theta}_{7,L}, \tilde{\theta}_{7,U})$
Baseline utility control arm	$\hat{\nu}_{8,b,C}$ $(\hat{\nu}_{8,b,C,L}, \hat{\nu}_{8,b,C,U})$	$\hat{\nu}_{8,b,C}$ $(\hat{\nu}_{8,b,C,L}, \hat{\nu}_{8,b,C,U})$
Utility improvement	$\hat{\theta}^*_8$ $(\hat{\theta}^*_{8,L}, \hat{\theta}^*_{8,U})$	$\tilde{\theta}^*_8$ $(\tilde{\theta}^*_{8,L}, \tilde{\theta}^*_{8,U})$
Resource use intervention arm	$\widehat{\nu}_{10,I}$ $(\widehat{\nu}_{10,I,L}, \widehat{\nu}_{10,I,U})$	$\widetilde{\nu}_I$ $(\widetilde{\nu}_{I,L}, \widetilde{\nu}_{I,U})$
Resource use control arm	$\widehat{\nu}_{10,C}$ $(\widehat{\nu}_{10,C,L}, \widehat{\nu}_{10,C,U})$	$\widetilde{\nu}_{10,C}$ $(\widetilde{\nu}_{10,C,L}, \widetilde{\nu}_{10,C,U})$

TABLE 6.2: Summary of notation for adjusted and unadjusted within trial and model-based health economic parameters

Chapter 7

Simulation Study to Assess the Bias in a Health Economic Analysis Following a Group Sequential Design

7.1 Introduction

The analysis following a group sequential design requires adjustment to account for the adaptive nature of the trial. Chapter 3 demonstrated that no clinical trials with an adaptive design and health economic analysis adjusted their analyses. This is not surprising given the lack of methods research discussing the impact and proposed adjustments to a health economic analysis following this type of trial.

To address this gap in the research, Chapter 6 extends the existing theory on adjustments following a group sequential design to the specific health economic context. Adjustments to the point estimates and confidence intervals for a within trial analysis are discussed, applying the existing BAMLE and SMO approaches to health economic outcomes. Adjustments to health economic model parameters from the trial data are described in the context of the CACTUS case study, extending the existing methods to allow calculation of adjusted estimates. This demonstrates how adjustments can be operationalised from a theoretical perspective.

Whitehead, 1986a; Whitehead, 1986b; Pinheiro, 1997 discuss how the level of bias in the analysis following a group sequential trial depends on the stopping rule chosen, the number and timing of interim analyses and the correlation between primary and secondary outcomes. To

date, these trial and data characteristics have not been explored in the context of a health economic analysis.

7.2 Chapter Aims

This chapter considers the extent to which bias may affect a health economic analysis and how well existing adjustments allow for this under a range of scenarios. Simulations are used to achieve the third research aim; to explore the potential for an adaptive design to impact the health economic analysis following a clinical trial. This investigation is split into two parts.

The first investigation considers a simplified context where data are simulated using a multivariate Normal distribution for the primary outcome, costs and QALY. The bias in the within trial health economic analysis is assessed under a range of trial designs and correlations between parameters.

The second investigation extends this to a more realistic setting anchored in the CACTUS case study, described in Chapter 5. The assumption of Normality for the trial data is dropped and pairwise correlations are allowed to vary between parameters. Both a within trial and model-based health economic analysis are considered using the adjustments discussed in Chapter 6.

7.3 Simulation Study Methods Overview

A simulation study allows an understanding of the 'truth' for the primary and secondary parameters of interest as they are known from the data generating process (Morris *et al.*, 2019). It is possible to estimate the deviation from the truth for each of the designs and adjustments considered. Adjustments are compared for trials with a fixed sample size design (FIX), Pocock stopping rule (POC) and O'Brien-Fleming stopping rule (OBF) with two and five interim analyses. The correlation between primary and secondary outcomes is varied from zero to one.

The first investigation (Simulation Study One) considers a simplified context where a primary outcome, costs and QALYs are simulated using a multivariate Normal distribution with a common pairwise correlation between the variables in a within trial health economic analysis. The

impact of the choice of stopping rule, number of interim analyses and correlation between parameters are explored. Adjusted and unadjusted approaches to calculate point estimates and confidence intervals are compared to assess how well they handle the bias introduced by the adaptive design. The CACTUS pilot data described in Chapter 5 are used to inform the simulation parameters. The findings of the study are used to refine and guide the scenarios considered in the second investigation.

The second investigation (Simulation Study Two) considers a more realistic simulation of the trial data, allowing for non-Normal marginal distributions and for pairwise correlations to have a different sign for different parameters. The impact of the choice of stopping rule, number of interim analyses and correlation between parameters on the bias in a model-based health economic analysis are considered. The CACTUS pilot data are used to inform the simulation parameters and the health economic analyses described in Section 5.6.1 are used to conduct the model-based analysis.

Two thousand repetitions are used in the simulation study. This balances the computation time and the accuracy of the results. Based on the pilot data in Table 7.5, an estimate of the bias in the primary outcome has a Monte Carlo standard error of approximately 0.0075. An estimate of the coverage of the 95% confidence interval has a Monte Carlo standard error of 0.2375 calculating using the methods described by Morris *et al.*, 2019.

The methods for Simulation Study One are described and results are presented, followed by Simulation Study Two. The discussion then summarises and draws conclusions from all analyses.

7.3.1 Software

The simulation studies are carried out in the statistical package `R`. The `RCTdesign` package is used for the design and analysis of the group sequential trials considered as described in Section 2.6, and the `HEEMOD` package is used to conduct the model based analysis, as described in Section 5.7.1.

FIGURE 7.1: Summary of the steps in the simulation of trial outcome data to explore the impact of trial and data characteristics on bias in the health economic analysis following an adaptive design

1. Choose the data characteristics under consideration including:
   (a) Correlation between trial outcomes.
   (b) Mean and standard deviation for each trial outcome.

2. Choose the trial characteristics under consideration including:
   (a) Type I and type II error rates.
   (b) Stopping rule.
   (c) Number of interim analyses.

3. Calculate the required sample size n_{max} and boundary values at each planned interim analysis.

4. Simulate outcome data for the n_{max} individuals from a multivariate Normal distribution and calculate their net benefit.

5. Apply the trial analysis for the design chosen in step 2 to determine when the trial would have stopped, giving the trial analysis dataset.

6. Calculate adjusted and unadjusted point estimates of the primary and secondary outcomes.

7. Repeat N_{SIM} times.

8. Performance measures - Compare average estimates over N_{SIM} simulated trials to the 'true' values used to simulate the data in step 3 by calculating the percentage bias.

7.4 Simulation Study One - CACTUS Within Trial Analysis using the Multivariate Normal Distribution

The setting of the CACTUS case study is used to illustrate the potential impact of adaptive designs with varying characteristics on a within trial health economic analysis. The trial is assumed to have two arms (computer-based intervention and usual care control). The following sections describe the methods of the simulation study with a summary in Figure 7.1.

7.4.1 Data Generating Mechanism

Data representing the expected trial outcomes for individuals randomised to each intervention are simulated using the multivariate Normal distribution. Simulated outcomes include:

1. Primary outcome (θ_1) - improvement in percentage of words named correctly between baseline and 6-month follow-up.

2. QALY (θ_2) - quality adjusted life year from baseline to 6-month follow-up.

3. Cost (θ_3) - total costs (including resource costs and intervention costs) from baseline to 6-month follow-up.

The notation used in Chapter 6 is adopted in this chapter, as summarised in Table 6.2.

Each pair of outcomes are assumed to have the same correlation, denoted by

$$\rho = \rho_{12} = \rho_{13} = \rho_{23}. \tag{7.1}$$

The values in Table 7.1 are used as the 'true' parameter values to simulate trial result datasets. Individual level data in the intervention arm are simulated from the following multivariate Normal distribution,

$$\begin{pmatrix} \theta_{1I} \\ \theta_{2I} \\ \theta_{3I} \end{pmatrix} \sim MVN \left(\begin{pmatrix} \mu_{1,I} \\ \mu_{2,I} \\ \mu_{3,I} \end{pmatrix}, \begin{pmatrix} \sigma_1^2 & \rho\sigma_1\sigma_2 & \rho\sigma_1\sigma_3 \\ \rho\sigma_2\sigma_1 & \sigma_2^2 & \rho\sigma_2\sigma_3 \\ \rho\sigma_3\sigma_1 & \rho\sigma_3\sigma_2 & \sigma_3^2 \end{pmatrix} \right). \tag{7.2}$$

Individual level data in the control arm are simulated from the following multivariate Normal distribution,

$$\begin{pmatrix} \theta_{1C} \\ \theta_{2C} \\ \theta_{3C} \end{pmatrix} \sim MVN \left(\begin{pmatrix} \mu_{1,C} \\ \mu_{2,C} \\ \mu_{3,C} \end{pmatrix}, \begin{pmatrix} \sigma_1^2 & \rho\sigma_1\sigma_2 & \rho\sigma_1\sigma_3 \\ \rho\sigma_2\sigma_1 & \sigma_2^2 & \rho\sigma_2\sigma_3 \\ \rho\sigma_3\sigma_1 & \rho\sigma_3\sigma_2 & \sigma_3^2 \end{pmatrix} \right). \tag{7.3}$$

7.4.2 Trial Design and Data Characteristics

The scenarios considered in all the investigations are summarised in Table 7.2. The FIX design and two group sequential stopping boundaries are considered; POC and OBF, described in Section 2.4. OBF and POC have been chosen because they have contrasting characteristics.

Parameter	Description	Value	Source
$\mu_{1,I}$	mean improvement in word naming ability in the intervention arm	0.21	CACTUS pilot data
$\mu_{2,I}$	mean QALY intervention arm	0.31	CACTUS pilot data
$\mu_{3,I}$	mean cost intervention arm	972.33	CACTUS pilot data
$\mu_{1,C}$	mean improvement in word naming ability in the control arm	0.08	CACTUS pilot data
$\mu_{2,C}$	mean QALY control arm	0.28	CACTUS pilot data
$\mu_{3,C}$	mean cost control arm	270.97	CACTUS pilot data
σ_1	standard deviation of improvement in word naming ability (pooled over treatment arms)	0.34	CACTUS pilot data
σ_2	standard deviation of QALY (pooled over treatment arms)	0.12	CACTUS pilot data
σ_3	standard deviation of cost (pooled over treatment arms)	284.24	CACTUS pilot data

TABLE 7.1: Summary of the parameters used to simulate 2,000 trial result datasets in the simulation studies

Pinheiro, 1997 has shown that the OBF rule introduces less bias in the primary outcome for smaller effect sizes compared to POC, but greater bias for larger effect sizes.

Todd *et al.*, 2001 suggest between four and eight interim analyses are appropriate and Pocock, 1983 suggests there is little statistical benefit to conducting more than five interims. Therefore, up to five equally spaced analyses are considered. As discussed in Chapter 2, the correlation between the primary and secondary outcomes in a group sequential design can influence the level of bias in the point estimate of the secondary outcomes (Whitehead, 1986b). To explore the impact this has on the health economic outcomes the correlation ρ is varied from zero to one.

The sample size for a given design is determined by the choice of stopping rule, number of interim analyses in Table 7.1. A type I error rate of 0.05 and type II error rate of 0.1 are chosen based on the Big CACTUS trial design described in Chapter 5. The minimally important difference and standard deviation for the primary outcome are estimated from the CACTUS pilot data. Boundary values are calculated based on the design parameters using `RCTdesign` as outlined in Section 2.4. These boundaries are not updated based on observed data at the interim analyses, as this is computationally demanding in the simulation study. When fitting the group sequential design to a real-world trial the boundaries would only need to be updated a small number of times (dependent on the number of analyses), and so it will be possible to update

Parameter	Description	Value
Stopping rule	Group sequential design stopping boundary	FIX, OBF, POC
Number of analyses	Number of analyses	1,2,5
ρ	Correlation between outcomes	0.0, 0.2, 0.4, 0.6, 0.8 1.0

TABLE 7.2: Summary of the 30 scenarios considered in Simulation Study One and 25 scenarios considered in Simulation Study Two. Correlations of 1.0 are not used in Simulation Study Two

the boundaries based on the interim data rather than using the same boundaries throughout the study.

7.4.3 Trial Result Estimates

For each of the 30 designs in Table 7.2, the sample size and boundary values are calculated. The trial outcome dataset represents the data collected as part of this trial as if it was being conducted in the real-world setting. The first group of simulated participants form the analysis dataset for the first interim analysis. The primary endpoint is estimated ($\hat{\theta}_1$) and compared to the pre-specified boundary value. If the estimate crosses the boundary, the trial is stopped and the data collected up to that point become the trial analysis dataset. If the trial reaches the final analysis without crossing the boundary and stopping early, the final analysis becomes the trial analysis dataset.

The results of interest in the simulation study are adjusted and unadjusted point estimates and 95% confidence intervals for the primary outcome (θ_1), QALY (θ_2), costs (θ_3) and INB (θ_4). The average pooled correlations between the primary outcome and the health economic outcomes across the simulations are also summarised. These results are calculated using the methods described in Section 6.4 and summarised in Table 6.1. Only unadjusted estimates and intervals are presented for the FIX designs as there are no early examinations of the data.

7.4.4 Performance Measures of Point Estimates and Confidence Intervals

Adjusted and unadjusted point estimates are calculated for $N_{SIM} = 2,000$ trials with the design under consideration. The average adjusted and unadjusted point estimates and confidence

intervals for each outcome are estimated across the N_{SIM} trials, given by

$$\bar{\tilde{\theta}} = \sum_{i=1}^{N_{SIM}} \frac{\tilde{\theta}_i}{N_{SIM}}, \tag{7.4}$$

$$\bar{\hat{\theta}} = \sum_{i=1}^{N_{SIM}} \frac{\hat{\theta}_i}{N_{SIM}}, \tag{7.5}$$

respectively.

7.4.4.1 Point Estimates

The average estimates are then compared to the 'true' parameter values given in Table 7.1 by calculating the expected bias, given by

$$\delta(\theta) = \bar{\tilde{\theta}} - \theta, \tag{7.6}$$

where

θ is the 'true' value,

$\bar{\tilde{\theta}}$ is the average adjusted estimate of θ across the simulated trials.

This is repeated for the unadjusted point estimates $\hat{\theta}$.

As each estimate has a different scale the standardised bias is calculated to allow a comparison of the bias in each outcome,

$$sb(\theta) = \frac{\delta(\theta)}{SE(\theta)} \times 100\%, \tag{7.7}$$

where

$SE(\theta)$ is the standard error of the estimates across the simulations.

The mean square error is calculated for each outcome using

$$MSE(\theta) = \delta(\theta)^2 + SE(\theta)^2. \tag{7.8}$$

The mean square error is a measure of the overall accuracy incorporating a measure of both bias and variability (Burton *et al.*, 2006).

7.4.4.2 Confidence Intervals

Coverage and width are used to assess the performance of the unadjusted and adjusted confidence intervals. The coverage of the adjusted and unadjusted confidence intervals is calculated using the upper and lower limits of the 95% confidence interval and counting the proportion of times the interval contains the true value across the simulated trials. The average width of the 95% confidence interval is the average difference between the upper and lower limits of the interval across the simulated trials. Ideally, the confidence interval will have coverage close to 95% and a narrow width.

7.5 Results

Results for Simulation Study One are presented in the following sections, first discussing the average correlations across the simulations, secondly the point estimates and finally the confidence intervals.

7.5.1 Correlations between Primary and Health Economic Outcomes

Table 7.3 gives the average correlations between the variables for the FIX design, for the true correlation values of 0, 0.4, 0.8. The four adaptive designs (each of POC and OBF with two and five analyses) have similar average correlations. As expected, the correlation between the primary outcome (θ_1) and the health economic outcomes (θ_2 to θ_4) increases with the true correlation imposed by the data generating mechanism described in Section 7.4.1.

There is near perfect correlation between QALY (θ_2) and net benefit (θ_4) in all cases, regardless of the true correlation between QALY and the primary outcome. The correlation between the primary outcome and net benefit is close to the true value in all cases. The correlation between costs and net benefit is also similar to the true value.

0.0 correlation				
	θ_1	θ_2	θ_3	θ_4
θ_1	1.000	0.001	-0.001	0.001
θ_2		1.000	0.001	0.992
θ_3			1.000	-0.107
θ_4				1.000
0.4 correlation				
	θ_1	θ_2	θ_3	θ_4
θ_1	1.000	0.383	0.383	0.353
θ_2		1.000	0.382	0.992
θ_3			1.000	0.283
θ_4				1.000
0.8 correlation				
	θ_1	θ_2	θ_3	θ_4
θ_1	1.000	0.783	0.782	0.759
θ_2		1.000	0.782	0.996
θ_3			1.000	0.733
θ_4				1.000

TABLE 7.3: Pooled correlation between primary and health economic parameters averaged across the 2,000 simulated trials with the fixed sample size design

7.5.2 Point Estimates

The standardised bias (defined in Equation 7.7) for each trial design and each correlation is given in Figure 7.2. The blue points and lines represent the adjusted estimates and the red points and lines the unadjusted estimates.

The first row of Figure 7.2 gives the standardised bias for the primary outcome for each of the five designs considered. The bias for the adjusted and unadjusted estimates is mostly unaffected by the correlation between the primary and health economic outcomes. The bias in the unadjusted estimates of the primary outcome is higher for POC compared to OBF and for five analyses compared to two.

In rows two to four of Figure 7.2 it is clear that as the correlation increases the bias in each of the health economic outcomes increases for the unadjusted estimates. The magnitude of this bias is similar for each of the outcomes for each scenario considered. As for the primary outcome, the bias is highest for POC with up to five analyses. The BAMLE (adjusted) point estimates have a much smaller bias compared to the unadjusted estimates. The adjusted estimates are less affected by the correlation between the primary and health economic outcome, however there is still some increase in the bias as the correlation increases.

Figure 7.3 summarises the square root of the mean square error (RMSE) for each within trial health economic outcome and the primary outcome. In the first row, the RMSE for the primary outcome is lowest for FIX. The RMSE is highest for the unadjusted estimates from POC with five analyses. The adjusted estimates provide some reduction in RMSE. In rows two to four, the RMSE is highest for the health economic outcomes when POC with five analyses is used, with a slight reduction for the adjusted estimates. The RMSE in the unadjusted case increases for higher correlation values. This is likely to be a consequence of the increased bias in the point estimates at these correlation values as shown in Figure 7.2. There is little difference between the RMSE for adjusted and unadjusted approaches for POC with two analyses. This is can be explained by the similar standard error for the estimates across the simulations that outweigh any differences in the bias.

Table 7.4 summarises the within trial health economic analysis for POC with five analyses and correlations between primary and health economic outcomes of zero, 0.4, 0.8 and 1.0. The tables gives the average estimate of the mean cost and QALY in each arm over the 2,000 simulated trials. The average difference in costs, difference in QALY and INB for a willingness to pay threshold of £20,000 per QALY are summarised. Both adjusted and unadjusted estimates are presented for comparison. This design is chosen as it appears to introduce the highest levels of bias into the within trial health economic analysis.

When there is no correlation, the unadjusted and adjusted point estimates are similar with the INB equal to -£129.14 and -£129.43 per QALY respectively. However, as the correlation increases the unadjusted estimate of the INB gets closer to zero. When there is perfect correlation between the primary and health economic outcomes the INB is £21.28 and -£102.75 per QALY for the unadjusted and adjusted analyses. These estimates have a different interpretation, with the adjusted analysis suggesting the intervention is not cost-effective but the unadjusted analysis suggesting the intervention is cost-effective. This highlights the importance of the bias adjustment following the group sequential design.

7.5.3 Confidence Intervals

The width of the confidence intervals is illustrated in Figure 7.4. For the primary outcome, in the first row, there are small differences in the average width of the confidence intervals. As expected, the intervals are narrowest for FIX and widest for the adaptive designs, especially POC

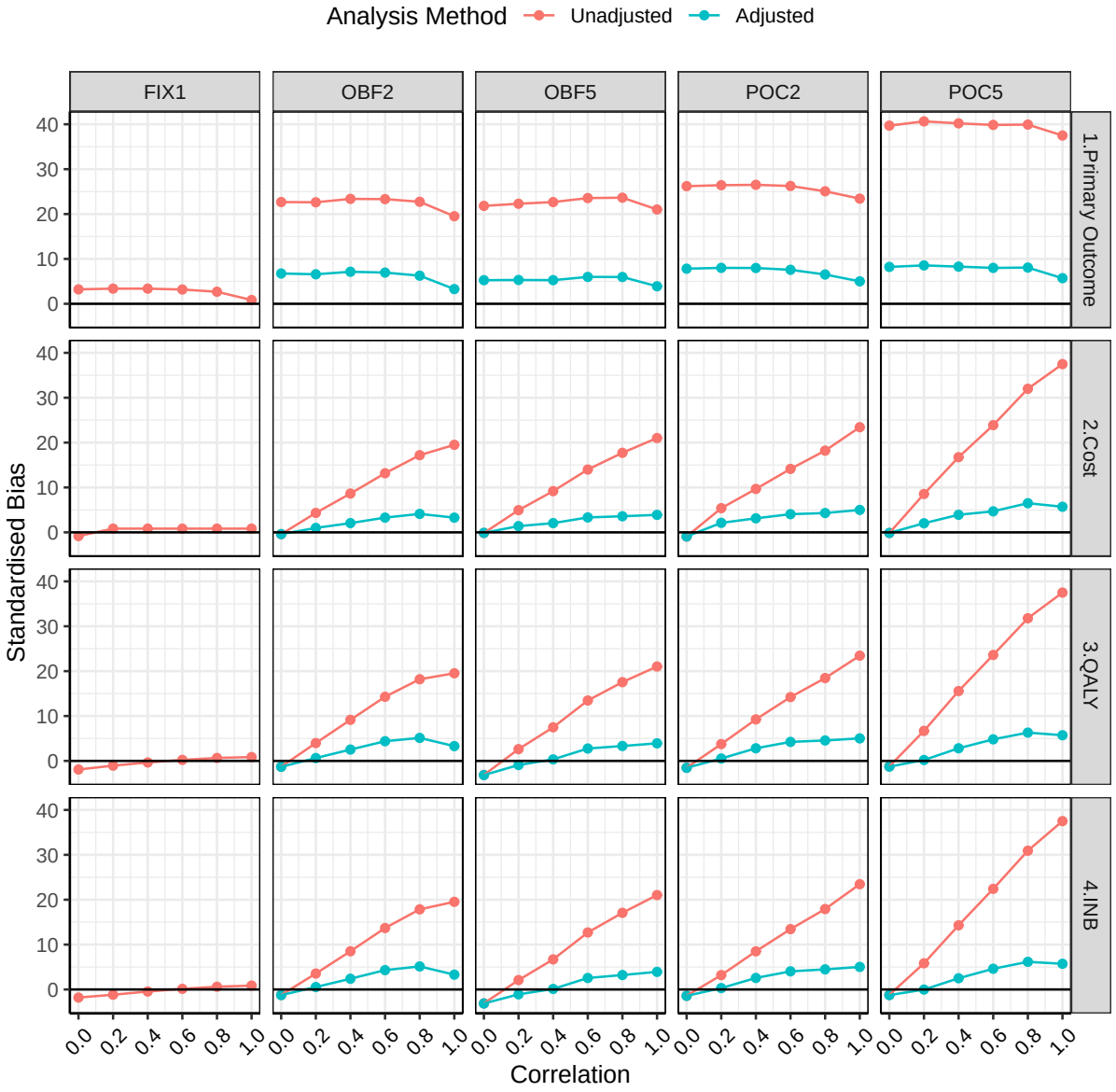


FIGURE 7.2: Results for the standardised bias for the five designs considered under a range of correlations between outcomes for Simulation Study One. FIX; fixed sample size design, POC; Pocock stopping rule, OBF; O'Brien-Fleming stopping rule, INB; Incremental net benefit, QALY; Quality adjusted life year

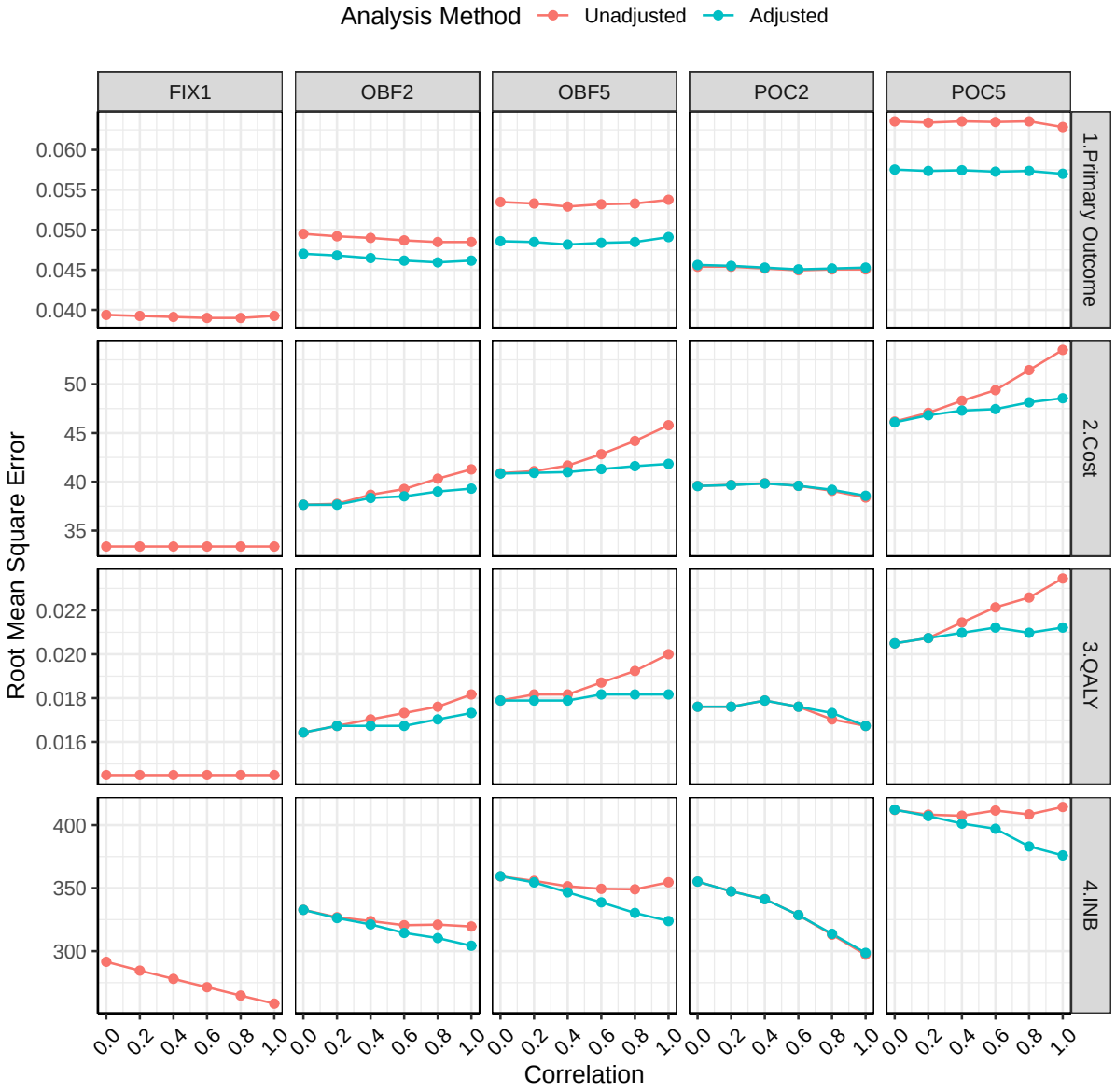


FIGURE 7.3: Results for the root mean square error for the five designs considered under a range of correlations between outcomes for Simulation Study One. FIX; fixed sample size design, POC; Pocock stopping rule, OBF; O'Brien-Fleming stopping rule, INB; Incremental net benefit, QALY; Quality adjusted life year

	Unadjusted					Adjusted				
	Per Person		Incremental			Per Person		Incremental		
	Cost	**QALY**	**Cost**	**QALY**	**INB**	**Cost**	**QALY**	**Cost**	**QALY**	**INB**
0.0 Correlation										
Control	270.50	0.28	-	-	-	270.48	0.28	-	-	-
Intervention	971.78	0.31	701.28	0.03	-129.14	971.78	0.31	701.30	0.03	-129.43
0.4 Correlation										
Control	267.06	0.28	-	-	-	270.38	0.28	-	-	-
Intervention	976.40	0.31	709.34	0.03	-66.58	973.58	0.31	703.21	0.03	-114.21
0.8 Correlation										
Control	263.26	0.28	-	-	-	269.94	0.28	-	-	-
Intervention	980.29	0.31	717.03	0.04	-3.58	974.42	0.31	704.48	0.03	-100.70
1.0 Correlation										
Control	261.11	0.28	-	-	-	269.65	0.28	-	-	-
Intervention	981.25	0.31	720.14	0.04	21.28	973.77	0.31	704.12	0.03	-102.75

TABLE 7.4: Summary of within trial incremental health economic analysis for trials with a Pocock stopping rule and five analyses. INB; Incremental net benefit, QALY; Quality adjusted life year

with five analyses. The adjusted intervals are slightly wider for each design, highlighting the greater uncertainty in the data from the adaptive trials with a smaller sample size, on average, compared to FIX.

The unadjusted confidence intervals for the health economic outcomes follow a similar pattern and are consistent for each correlation value for a given design, with the exception of the INB. For this outcome, the width of the intervals gets narrower as the correlation increases. This suggests that as the correlation between the QALY and cost increases the uncertainty in the INB decreases giving narrower confidence intervals. The adjusted intervals are wider compared to the unadjusted intervals for the health economic outcomes. The adjusted intervals are affected by the higher levels of correlation. For POC with two analyses, this results in intervals that are slightly narrower than the unadjusted intervals.

The coverage of the confidence intervals is illustrated in Figure 7.5 for adjusted and unadjusted estimates. Ideally, for a 95% confidence interval the coverage will be 0.95. The coverage for the unadjusted intervals of the primary outcome (row one) decreases for the adaptive designs with up to five analyses. The adjusted intervals have coverage close to 0.95 for all the adaptive designs.

In rows two to four, the impact of the adaptive design on the coverage of the unadjusted intervals is reduced. The coverage for designs with five analyses is lower compared to those

with two analyses; however, the coverage is still close to 0.95. The adjusted confidence intervals provide coverage higher than 0.95 especially for lower correlation values and for OBF. The coverage for POC when the correlation values are greater than 0.4 drops below 0.95, and in some cases below the coverage of the unadjusted intervals. However, these values are still close to the desired 0.95 value.

7.5.4 Summary of Results

Simulation Study One assumes a multivariate Normal distribution to simulate correlated primary and health economic outcomes required for a within trial health economic analysis. The study compares adaptive designs using POC or OBF stopping rules with up to five analyses and for pairwise correlations between outcomes taking values between zero and one.

High levels of correlation between primary and health economic outcomes results in bias in the point estimates potentially changing the conclusions drawn about the short-term cost-effectiveness of an intervention. The bias is higher when there are more interim analyses and for POC compared to OBF with the same number of analyses. The BAMLE does reduce the bias in the point estimates but it is important to note that it does not eradicate it completely. Adjusted confidence intervals provide wider intervals and coverage close to the desired 0.95 level for all designs.

Simulation Study One has made a number of limiting assumptions to illustrate the potential for an adaptive design to influence a within trial health economic analysis, for example, the assumption of a multivariate Normal distribution and an assessment of the short-term cost-effectiveness only. Simulation Study Two aims to address some of these limitations.

7.6 Simulation Study Two - CACTUS Case Study Model Based Analysis

Simulation Study One demonstrates the potential for an adaptive design to influence a within trial health economic analysis where the variables are Normally distributed with equal correlation between variables. This is an idealistic scenario as, in reality, outcomes are not always Normally distributed, there may be different correlations between variables and a model-based

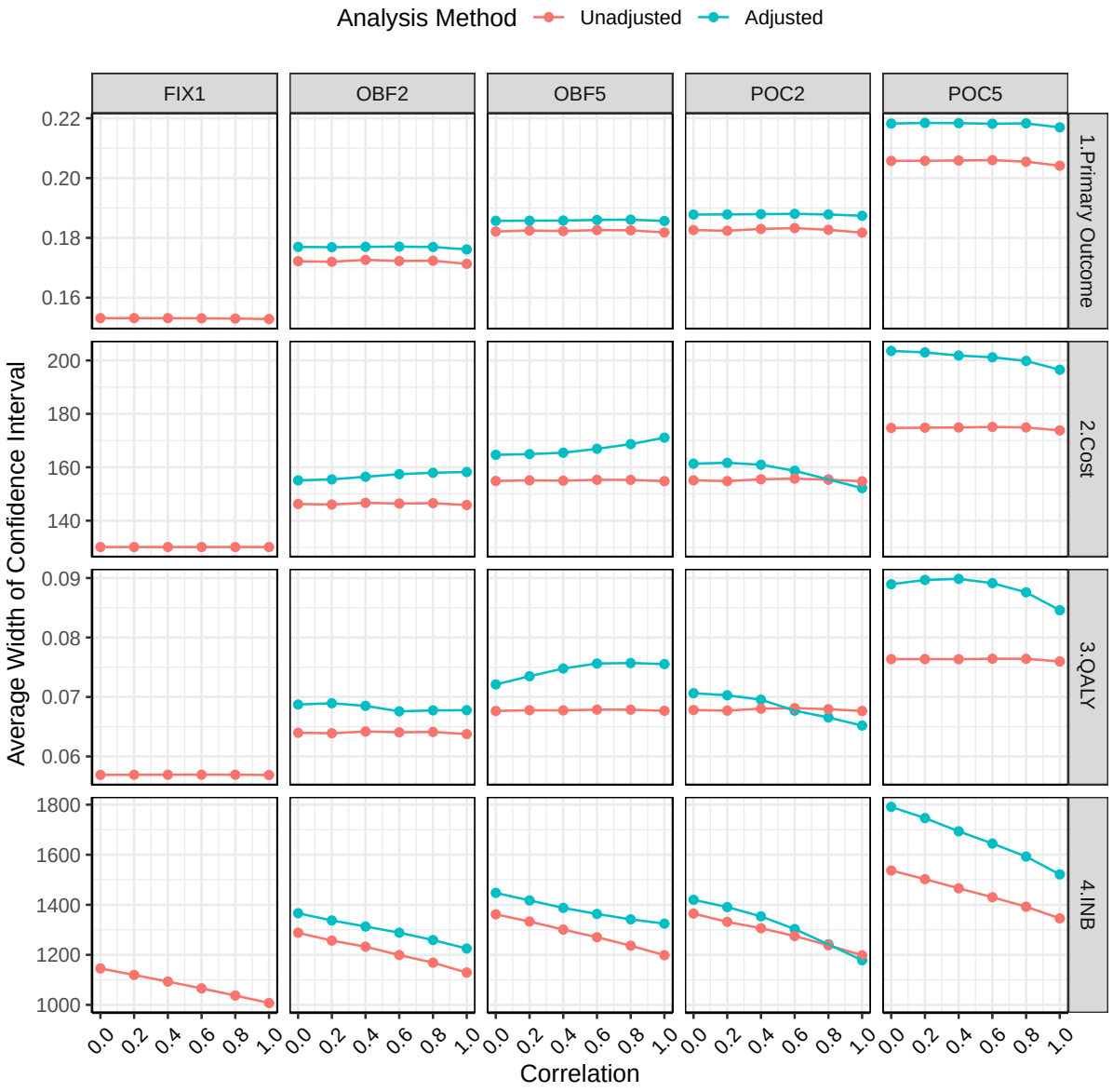


FIGURE 7.4: Results for the average width of the confidence intervals for the five designs considered under a range of correlations between outcomes for Simulation Study One. FIX; fixed sample size design, POC; Pocock stopping rule, OBF; O'Brien-Fleming stopping rule, INB; Incremental net benefit, QALY; Quality adjusted life year

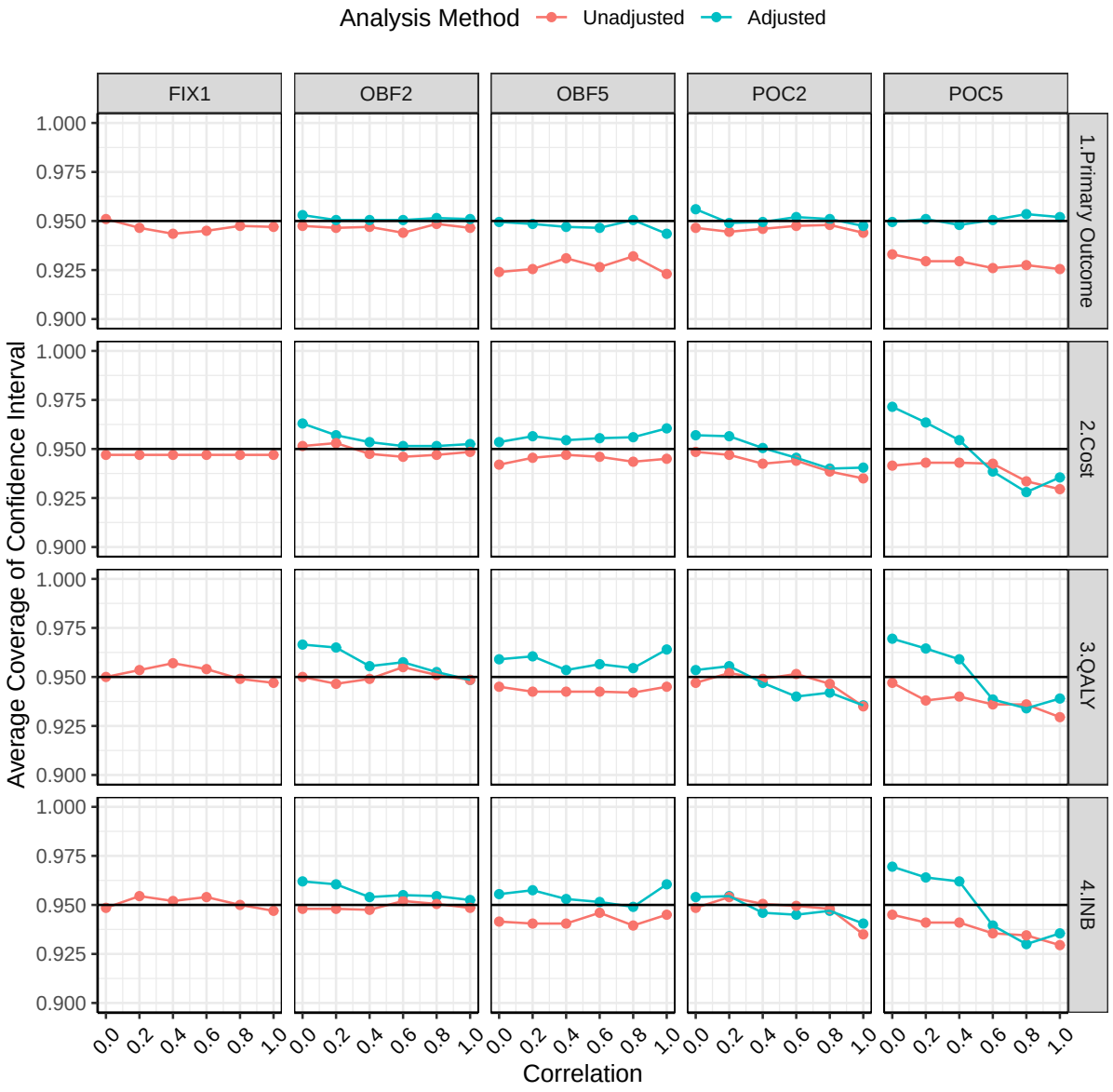


FIGURE 7.5: Results for the average coverage of the 95% confidence intervals for the five designs considered under a range of correlations between outcomes for Simulation Study One. FIX; fixed sample size design, POC; Pocock stopping rule, OBF; O'Brien-Fleming stopping rule, INB; Incremental net benefit, QALY; Quality adjusted life year

analysis is likely to be required to establish the long-term cost-effectiveness of an intervention.

Simulation Study Two considers a more realistic simulation of trial data representative of a full scale trial following the CACTUS pilot described in Chapter 5. The simulation study explores the impact of bias on the estimation of health economic model parameters and the deterministic model-based analysis, described in Section 6.5, for the trial designs considered in Simulation Study One.

7.6.1 Data Generating Mechanism

A dataset of the expected trial outcomes for individuals randomised to the computer-based intervention arm and usual care control arm of the trial is simulated. A multivariate distribution that allows the marginal distributions to be non-Normal and where a correlation between outcomes can be specified is used. Simulated outcomes include

1. Percentage of words named correctly measured at baseline, 6 and 9-months follow-up.

2. Resource costs - resource costs incurred from baseline to 6-month follow-up (not including the cost of delivering the computer-based intervention).

3. Utility - EQ-5D score (described in Section 2.7.3) measured at baseline, 6 and 9-months follow-up.

Let $x_{1Ib,j}, x_{1I6,j}, x_{1I9,j}$ be the observed percentage words named correctly at baseline, 6 and 9-months follow-up for participant j in the intervention arm. Let $x_{1Cb,j}, x_{1C6,j}, x_{1C9,j}$ be the percentage words named correctly at baseline, 6 and 9-months follow-up for participant j in the control arm. The true means are $\mu_{x_1b,I}, \mu_{x_16,I}, \mu_{x_19,I}$ in the intervention arm and $\mu_{x_1b,C}, \mu_{x_16,C}, \mu_{x_19,C}$ in the control arm with standard deviation σ_{x_1I} and σ_{x_1C} respectively.

Let $u_{Ib,j}, u_{I6,j}, u_{I9,j}$ be the observed utility score at baseline, 6 and 9-months follow-up for participant j in the intervention arm. Let $u_{Cb,j}, u_{C6,j}, u_{9C,j}$ be the observed utility score at baseline, 6 and 9-months follow-up in the control arm. Let $\mu_{ub,I}, \mu_{u6,I}, \mu_{u9,I}$ be the true mean utility score in the intervention arm and $\mu_{ub,C}, \mu_{u6,C}, \mu_{u9,C}$ the true mean utility score in the control arm with standard deviations σ_{uI} and σ_{uC} respectively.

Let $y_{I,j}$ and $y_{C,j}$ be the observed total resource costs incurred up to the 6-months follow-up for participant j in the intervention and control arms respectively. Let $\mu_{y,I}$ be the mean resource cost in the intervention arm and $\mu_{y,C}$ the mean in the control arm with standard deviations σ_{yI} and σ_{yC} respectively.

7.6.1.1 Correlations

Correlations between repeated outcomes measured on the same participant at different time points are commonly correlated (Walters *et al.*, 2019). For example, a patient who has a high percentage of words named correctly at baseline is likely to have a high percentage of words named correctly at six-months follow-up; a positive correlation. Let ρ_t denote the correlation within a given outcome measured at different time points. This is fixed to be $\rho_t = 0.5$ for all outcomes at each follow-up.

Correlation values are varied to explore the effect of correlation between the primary and health economic outcomes on the level of bias in the point estimates. Let ρ denote the correlation. It is assumed that outcomes measured at the same period have correlation ρ, for example baseline costs and baseline utility have correlation ρ. Outcomes measured at one follow-up time point apart have correlation ρ^3 and two follow-up time points have correlation ρ^4. For example, the correlation between baseline costs and utility at 6-months follow-up have correlation ρ^3. This structure is chosen as if the same correlation value is chosen for all periods these terms cancel out and give a correlation of zero. This is discussed in Appendix D.

At each time point, the correlation between the percentage of words named correctly and utility score is assumed positive as it is feasible that as a person's word naming ability improves their quality of life and hence their utility score improves. The correlation between the percentage of words named correctly and resource costs is assumed negative as it is feasible that as a person's word naming ability improves they incur fewer costs as they need less speech and language therapy. The correlation between the utility score and costs is assumed negative as it is feasible that as a person's quality of life improves they incur fewer costs.

Individual level data in the intervention arm are simulated using the mean vector and covariance matrix given by

$$\left(\begin{pmatrix}\mu_{x_1b,I}\\ \mu_{x_16,I}\\ \mu_{x_19,I}\\ \mu_{ub,I}\\ \mu_{u6,I}\\ \mu_{u9,I}\\ \mu_{y,I}\end{pmatrix}, \begin{pmatrix}\sigma^2_{x_1,I} & \rho_t\sigma_{x_1I}\sigma_{x_1I} & \rho_t\sigma_{x_1I}\sigma_{x_1I} & \rho\sigma_{x_1I}\sigma_{uI} & \rho^3\sigma_{x_1I}\sigma_{uI} & \rho^4\sigma_{x_1I}\sigma_{uI} & -\rho^3\sigma_{x_1I}\sigma_{yI}\\ & \sigma^2_{x_1I} & \rho_t\sigma_{x_1I}\sigma_{x_1I} & \rho^3\sigma_{x_1I}\sigma_{uI} & \rho\sigma_{x_1I}\sigma_{uI} & \rho^3\sigma_{x_1I}\sigma_{uI} & -\rho\sigma_{x_1I}\sigma_{yI}\\ & & \sigma^2_{x_1I} & \rho^4\sigma_{x_1I}\sigma_{uI} & \rho^3\sigma_{x_1I}\sigma_{uI} & \rho\sigma_{x_1I}\sigma_{uI} & -\rho^3\sigma_{x_1I}\sigma_{yI}\\ & & & \sigma^2_{uI} & \rho_t\sigma_{uI}\sigma_{uI} & \rho_t\sigma_{uI}\sigma_{uI} & -\rho^3\sigma_{uI}\sigma_{yI}\\ & & & & \sigma^2_{uI} & \rho_t\sigma_{uI}\sigma_{uI} & -\rho\sigma_{uI}\sigma_{yI}\\ & & & & & \sigma^2_{uI} & -\rho^3\sigma_{uI}\sigma_{yI}\\ & & & & & & \sigma^2_{yI}\end{pmatrix}\right). \tag{7.9}$$

Data in the control arm are simulated using a similar mean and covariance matrix structure, replacing parameters with the corresponding control arm values.

7.6.1.2 Marginal Distributions

The marginal distributions of the simulated parameter values are allowed to be non-Normal. The chosen distributions are summarised in Table 7.5 with the 'true' parameters used to simulate the data based on the CACTUS pilot data described in Chapter 5.

A Beta distribution is used to simulate the primary outcome data (percentage of words named correctly) at baseline and follow-up time points. This distribution does not allow values below zero and above one as is required for a percentage outcome. The parameters for the Beta distribution are estimated from the mean and standard deviations in Table 7.5 using

$$a = \mu^2\left(\frac{(1-\mu)}{\sigma^2 - \frac{1}{\mu}}\right), \tag{7.10}$$

$$b = \alpha\left(\frac{1}{\mu - 1}\right). \tag{7.11}$$

EQ-5D utility scores are simulated for baseline and follow-up time points by first simulating disutilities and transforming on to the utility scale by subtracting from one. Simulating disutilities (1 minus the utility) allows utility scores to be negative, representing health states considered to be worse than death but imposing an upper limit of one (full health) (Patrick *et al.*, 1994).

The logNormal distribution represents a small number of patients with high disutility values (low utility values) and the majority of patients with a score lying closer to one. The parameters for the logNormal distribution are estimated from the means and standard deviations, given in Table 7.5 using

$$(1-m) = log\left(\frac{(1-\mu)^2}{\sqrt{\sigma^2+(1-\mu)^2}}\right), \tag{7.12}$$

$$s = \sqrt{\frac{log(1+\sigma^2)}{(1-\mu)^2}}. \tag{7.13}$$

Resource cost data are simulated using a logNormal distribution to represent the right skew commonly seen, where a small number of participants have high costs (Briggs *et al.*, 2006). The parameters for the logNormal distribution are estimated from the means and standard deviations given in Table 7.5 using

$$m = log\left(\frac{\mu^2}{\sqrt{\sigma^2+\mu^2}}\right), \tag{7.14}$$

$$s = \sqrt{\frac{log(1+\sigma^2)}{\mu^2}}. \tag{7.15}$$

7.6.1.3 Using Copulas to Simulate Non-Normal Marginal Distributions

To simulate the multivariate non-Normal distribution, copulas are used. By using copulas, it is possible to simulate correlated variables that have different marginal distributions. For the case of three correlated variables denoted by x_1, x_2, x_3 the copula method has the following steps (Nelsen, 2007; Hofert *et al.*, 2011):

1. Simulate three multivariate Normal variables. Let these variables be denoted by

$$\mathbf{y} = (y_1, y_2, y_3). \tag{7.16}$$

   These variables have the desired Spearman Rank Correlation. Let this correlation be denoted by ρ.

Parameter	Description	Value	Distribution	Source
		Correlation		
ρ_t	outcome measured a different time points	0.50	Fixed	Walters *et al.*, 2019
ρ		0.00, 0.20, 0.40, 0.60, 0.80	Fixed	-
Parameter	**Description**	**Mean (SD)**	**Distribution**	**Source**
		Percentages words named correctly		
$\mu_{x_1b,I}(\sigma_{x_{1I}})$	baseline	0.355 (0.3589)	Beta	CACTUS pilot
$\mu_{x_16,I}(\sigma_{x_{1I}})$	6-months	0.563 (0.3589)	Beta	CACTUS pilot
$\mu_{x_19,I}(\sigma_{x_{1I}})$	9-months	0.551 (0.3589)	Beta	CACTUS pilot
$\mu_{x_1b,C}(\sigma_{x_{1C}})$	baseline	0.517 (0.3088)	Beta	CACTUS pilot
$\mu_{x_16,C}(\sigma_{x_{1C}})$	6-months	0.599 (0.3088)	Beta	CACTUS pilot
$\mu_{x_19,C}(\sigma_{x_{1C}})$	9-months	0.706 (0.3088)	Beta	CACTUS pilot
		EQ-5D score		
$\mu_{ub,I}(\sigma_{uI})$	baseline	0.629 (0.2800)	logNormal	CACTUS pilot
$\mu_{u6,I}(\sigma_{uI})$	6-months	0.608 (0.2800)	logNormal	CACTUS pilot
$\mu_{u9,I}(\sigma_{uI})$	9-months	0.490 (0.2800)	logNormal	CACTUS pilot
$\mu_{ub,C}(\sigma_{uC})$	baseline	0.551 (0.4015)	logNormal	CACTUS pilot
$\mu_{u6,C}(\sigma_{uC})$	6-months	0.570 (0.4015)	logNormal	CACTUS pilot
$\mu_{u9,C}(\sigma_{uC})$	9-months	0.468 (0.4015)	logNormal	CACTUS pilot
		Total resource cost		
$\mu_{y,I}(\sigma_{yI})$		203.08 (346.17)	logNormal	CACTUS pilot
$\mu_{y,C}(\sigma_{yC})$		270.97 (222.30)	logNormal	CACTUS pilot

TABLE 7.5: Summary of the distributions and parameter values used to simulate the 2,000 trials in Simulation Study Two

2. Each variable ($\mathbf{y}$) is transformed onto the Uniform distribution using,

$$\mathbf{u} = \Phi(\mathbf{y}), \tag{7.17}$$

where $\Phi(\cdot)$ is the cumulative density function for the Normal distribution.

3. The Uniform variables ($\mathbf{u}$) are then transformed onto the chosen marginal distributions (such as the logNormal distribution) denoted by $F(\cdot)$ with the desired Spearman Rank Correlation by taking the inverse of the cumulative density function of the marginal distribution. For example for u_1 with marginal distribution $F_1(\cdot)$,

$$x_1 = F_1^{-1}(u_1). \tag{7.18}$$

The `Copula` package is used to implement this approach in `R`.

7.6.2 Trial Design and Data Characteristics

The designs from Simulation Study One are considered and include FIX, POC and OBF with two and five analyses. Each design has a type I error rate of 0.05 and type II error rate of 0.1.

The sample size for a given design is determined by the choice of stopping rule, number of interim analyses, type I error rate and type II error rate. The minimally important difference and standard deviation for the primary outcome, are estimated from the CACTUS pilot data using the values in Table 7.5. Stopping rule boundary values are calculated based on the design parameters using `RCTdesign` as outlined in Section 2.4.

Correlation values are varied from 0.0 to 0.8 to represent a low correlation structure to a high correlation structure. The highest correlation value is chosen to be 0.80 as this is the highest feasible correlation that can simulate a dataset for the chosen data structure and mean and standard deviation values with a valid covariance matrix. A covariance matrix is required to be positive semi-definite. Values of ρ from Table 7.5 greater than 0.8 violate this property and are not considered.

7.6.3 Trial Result Estimates

The estimates of interest are adjusted and unadjusted:

1. Within trial point estimates,
2. Model parameter estimates,
3. Deterministic model results.

Unadjusted analyses are presented for FIX.

Confidence intervals are not calculated for the within trial or model parameters. Confidence intervals for the within trial analysis have been discussed in Section 7.5.3. It is possible calculate adjusted confidence intervals for the health economic model parameters using the bootstrapping approach described in Section 6.5. The bootstrapping method and the Skalland approach (Figure 6.1) are computationally intensive, drastically increasing the computation time of the simulation study. For the case of a single trial it would be possible to calculate adjusted confidence intervals or an uncertainty analysis, as described in Section 6.5.6, to estimate the uncertainty in the model-based results.

7.6.3.1 Within Trial Analysis

As defined in Section 6.4 the observed improvement in percentage words named correctly at 6-months (primary outcome) for participant j in the intervention and control arms respectively are given by

$$x_{1I,j} = (x_{6I,j} - x_{bI,j}), \tag{7.19}$$

$$x_{1C,j} = (x_{6C,j} - x_{bC,j}). \tag{7.20}$$

The primary outcome is defined to be the difference in the mean improvement in percentage words named correctly from baseline to 6-months between the intervention and control arm. An unadjusted estimate of the primary outcome is given in Equation 6.7 and denoted by $\hat{\theta}_1$ and an adjusted estimate given in Equation 6.14 denoted by $\tilde{\theta}_1$.

The QALY at 6-months for individual j in the intervention and control arms respectively is denoted by $x_{2I,j}$ and $x_{2C,j}$. These values are calculated by linear interpolation using

$$x_{2I,j} = 0.5(0.5)(u_{bI,j} + u_{6I,j}), \tag{7.21}$$

$$x_{2C,j} = 0.5(0.5)(u_{bC,j} + u_{6C,j}). \tag{7.22}$$

An unadjusted within trial estimate of QALY ($\hat{\theta}_2$) is given in Equation 6.8 and an adjusted within trial estimate of QALY ($\tilde{\theta}_3$) is given in Equation 6.15.

In the CACTUS case study the fixed costs associated with delivering the intervention incurred by participants in the control arm was £0 as this was the usual care arm. The fixed cost of delivering the computer-based intervention was £769.25. Let $x_{3I,j}$ and $x_{3C,j}$ be the total resource costs incurred by participant j calculated as

$$x_{3I,j} = y_{I,j} + 769.25, \tag{7.23}$$

$$x_{3C,j} = y_{C,j} + 0. \tag{7.24}$$

An unadjusted within trial estimate of cost ($\hat{\theta}_3$) is given in Equation 6.9 and an adjusted within trial estimate of cost ($\tilde{\theta}_3$) is given in Equation 6.17.

Following the steps outlined in Section 6.4 an unadjusted within trial estimate of the INB ($\hat{\theta}_4$) is given in Equation 6.10 and an adjusted within trial estimate ($\tilde{\theta}_4$) given in Equation 6.18.

7.6.3.2 Health Economic Model Analysis

The parameters required for the health economic model analysis are described in Section 6.5 and include adjusted and unadjusted estimates of:

- Probability of good response,
- Relapse rate,
- Utility improvement,
- Resource use.

These parameters are used in the health economic model described in Chapter 5. The health economic model is evaluated using unadjusted (adjusted) parameter estimates to give an unadjusted (adjusted) deterministic model-based estimate of total costs, QALY and INB.

7.6.4 Performance Measures of Point Estimates

The same performance measures described for Simulation Study One in Section 7.4.4 are used. Where there is no 'true' parameter for the model parameters, the estimate from the CACTUS pilot summarised in Chapter 5 is used. The results of the deterministic health economic model analysis are compared to the unadjusted analysis of the FIX design as this represents how the analysis would be conducted in practice and the 'gold standard' design if there were unlimited resources and no ethical concerns.

7.7 Results

The results for Simulation Study Two are presented in the following sections, first discussing the average correlations across the simulations, secondly trial based point estimates and finally deterministic model results.

7.7.1 Correlation between Primary and Health Economic Outcomes

Table 7.6 gives the average correlation, pooled across treatment arms, between the primary outcome and the within trial health economic outcomes (costs, QALY and net benefit) for FIX. The correlation between the primary outcome (θ_1) and QALY (θ_2) is low regardless of the true correlation value used to simulate the trial data. This occurs because the covariance terms in the calculation of the correlation cancel out, giving an estimated correlation close to zero. This result is shown mathematically in Appendix D. This differs from the results in Simulation Study One (Table 7.3) where the differences were simulated directly with the chosen correlation value.

The correlation between the primary outcome (θ_1) and costs (θ_3) is also affected by this result. The correlation is highest when the difference between correlations at each time point is high rather than when the overall correlation level is high. For example, when the difference between the correlation ρ and ρ^3 is high rather than the value of ρ. This explains why

0.0 correlation				
	θ_1	θ_2	θ_3	θ_4
θ_1	1.000	0.001	0.001	0.000
θ_2		1.000	0.001	0.988
θ_3			1.000	-0.095
θ_4				1.000
0.4 correlation				
	θ_1	θ_2	θ_3	θ_4
θ_1	1.000	0.004	-0.297	0.030
θ_2		1.000	-0.251	0.992
θ_3			1.000	-0.329
θ_4				1.000
0.8 correlation				
	θ_1	θ_2	θ_3	θ_4
θ_1	1.000	0.007	-0.270	0.025
θ_2		1.000	-0.731	0.998
θ_3			1.000	-0.761
θ_4				1.000

TABLE 7.6: Pooled correlation between primary and health economic parameters averaged across the 2,000 simulated trials with the fixed sample size design

the correlation between the primary outcome and costs is higher for correlation of 0.4 than for 0.8.

The correlation between the primary outcome and net benefit is related to the correlations between the QALY and costs. This correlation is much smaller than seen in Simulation Study One, however it increases as the correlation between the primary outcome and costs increases. As in Simulation Study One the correlation between the QALY and net benefit is close to one regardless of the correlation value used for the data generation. The correlation between net benefit and costs and QALY and costs increases as the true correlation value increases.

Table 7.7 gives the average pooled correlation (across treatment arms) between the primary outcome and the parameters estimated from the trial data for the health economic model. As expected, the resource costs (ν_{10}) have the same correlation values as the total costs in Table 7.6. The correlation between the primary outcome and resource costs have a higher correlation when the true correlation is 0.6 than 0.8. Baseline utility ($\nu_{8,b}$) follows a similar pattern to resource costs as expected from the theory described in Appendix D.

The correlation between utility improvement (θ_8^*) and the primary outcome increases as the true correlation value increases, until a true value of 0.8. The correlation at this level is 0.486

0.0 Correlation					
	θ_1	ν_{10}	ν_{8b}	θ_8^*	θ_9
θ_1	1.000	0.001	0.000	0.001	0.491
0.2 Correlation					
	θ_1	ν_{10}	ν_{8b}	θ_8^*	θ_9
θ_1	1.000	-0.168	-0.168	0.320	0.491
0.4 Correlation					
	θ_1	ν_{10}	ν_{8b}	θ_8^*	θ_9
θ_1	1.000	-0.297	-0.295	0.569	0.490
0.6 Correlation					
	θ_1	ν_{10}	ν_{8b}	θ_8^*	θ_9
θ_1	1.000	-0.346	-0.342	0.656	0.490
0.8 Correlation					
	θ_1	ν_{10}	ν_{8b}	θ_8^*	θ_9
θ_1	1.000	-0.270	-0.263	0.486	0.491

TABLE 7.7: Correlation of health economic model parameters estimated from trial data with the primary outcome (improvement in percentage words named correctly at 6-months) for the fixed sample size design

compared to 0.656 when the correlation is 0.6. Again, this is a nuance of the relationship between the variables in the data generation. The higher levels of correlation are expected for this variable as utility improvement is related to the primary outcome as described in Section 6.5.4.1.

It is not possible to calculate a correlation between the primary outcome and the probability of a good response at 6-months (θ_6) and relapse rate (θ_7), however, these parameters are estimated using the improvement in percentage words named correctly at 6 and 9-months follow-up. Table 7.7 shows that the correlation between the primary outcome and the word improvement at 9-months (θ_9) have correlation approximately equal to 0.5 as expected from the data generating mechanism.

Similar correlation values were observed for the adaptive designs.

7.7.2 Within Trial Analysis without the Normality Assumption

Figure 7.6 gives the standardised bias for the within trial health economic parameters in Simulation Study Two. As in Simulation Study One (Figure 7.2) the bias is small for all outcomes for FIX. The bias in the primary outcome is highest for POC with five analyses, with the adjusted point estimates offering a reduction in the bias.

The within trial estimate of costs is underestimated for the unadjusted analyses of the adaptive designs, reflecting the negative correlation imposed in the data generation. The bias increases as the correlation increases to 0.6 with a small reduction in bias when the correlation reaches 0.8. The bias is highest for POC with five analyses and the adjusted point estimates are less biased for each adaptive design.

In contrast to Simulation Study One, the bias for the QALY is small and close to zero for all designs. There is a small increase in bias for POC with five analyses when the correlation is higher than 0.6. However, this is small relative to the bias in the cost and primary outcome. The INB is slightly over estimated for the adaptive designs with five analyses with the adjusted estimate providing some correction for this. Compared to Simulation Study One the impact of the adaptive design on the point estimate is small. This reflects the lower correlations between the primary outcome and health economic outcomes shown in Table 7.6.

Figure 7.7 summarises the RMSE for the within trial point estimates in Simulation Study Two. The RMSE is higher for the unadjusted point estimates in the adaptive designs for the primary outcome. The RMSE is similar for the adjusted and unadjusted health economic outcomes for the adaptive designs.

7.7.3 Health Economic Model Results

Figure 7.8 gives the standardised bias for the health economic model parameters estimated from the trial data. In the first row, the resource costs in the intervention arm are slightly under estimated for both the adjusted and unadjusted estimates. However, the adjusted estimates provide a reduction in this bias. The bias increases as the correlation increases to 0.6 and then there is a decrease in the bias for correlation equal to 0.8. The bias is greatest for the Pocock stopping rule with five analyses. The resource costs in the control arm (row two) follow a similar pattern but with the point estimates over-estimating the costs.

The third row of Figure 7.8 gives the standardised bias for the point estimates of the utility improvement used in the health economic model. This parameter is an estimate of the utility gain for participants who responded to treatment compared to those who did not respond to treatment, defined in Section 6.5.4.1. This variable was not specified in the data generation described in Section 7.6.1, instead this was estimated from the trial data. The bias in this variable

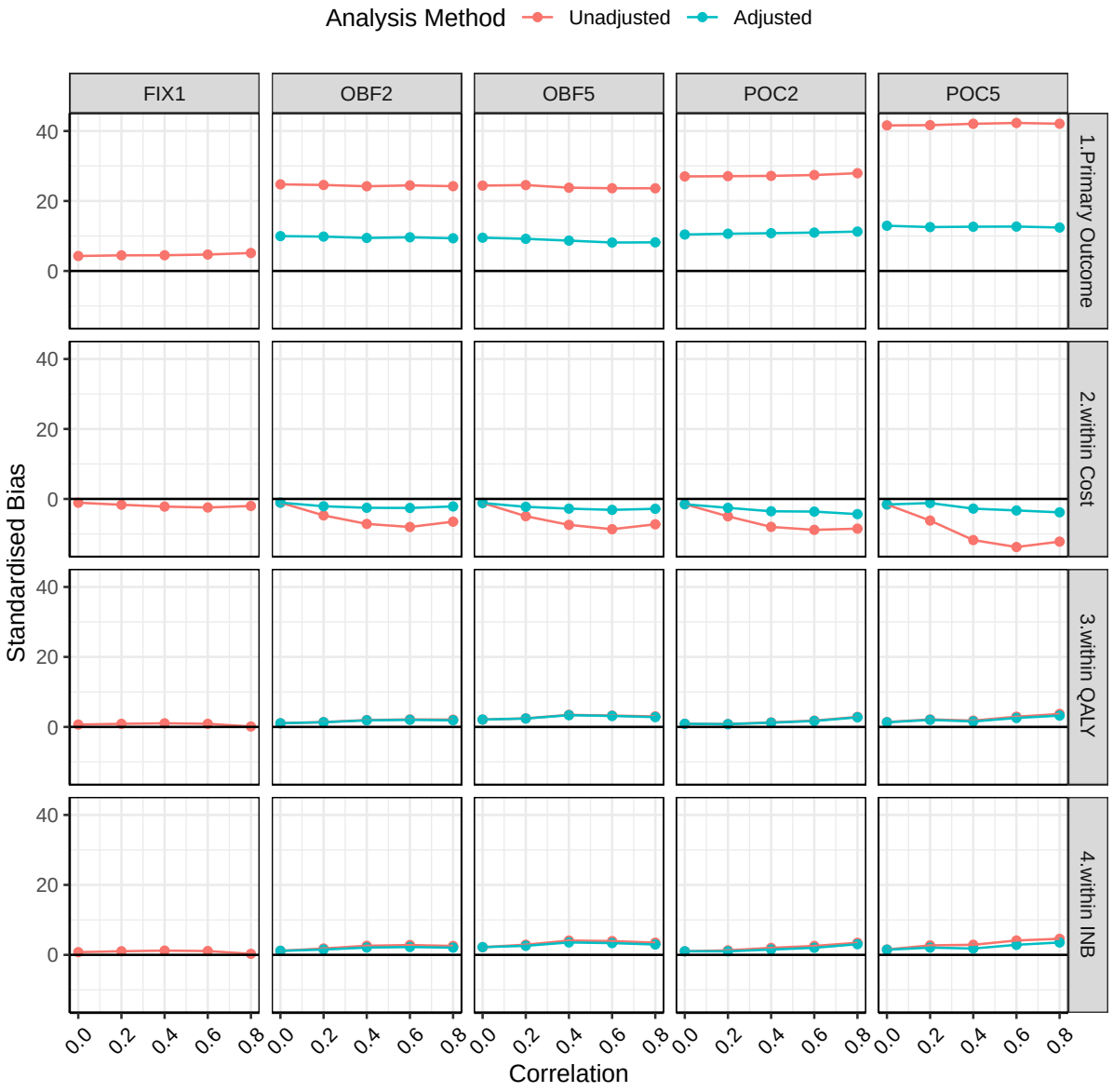


FIGURE 7.6: Results for the standardised bias for within trial health economic analysis outcomes for the five designs considered under a range of correlations between outcomes for Simulation Study Two. FIX; fixed sample size design, POC; Pocock stopping rule, OBF; O'Brien-Fleming stopping rule, INB; Incremental net benefit, QALY; Quality adjusted life year

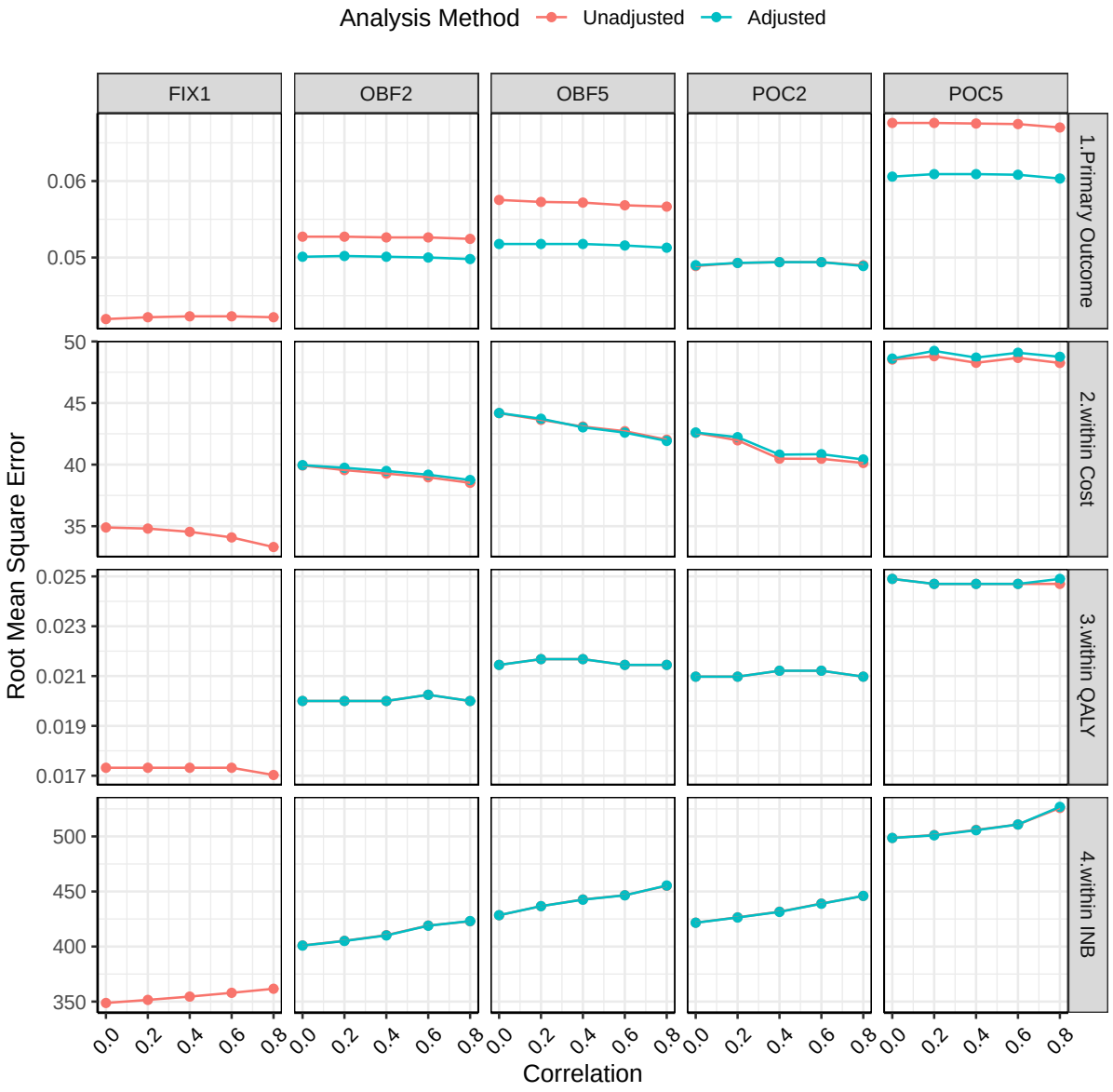


FIGURE 7.7: Results for the root mean square error for the within trial health economic analysis outcomes for the five designs considered under a range of correlations between outcomes for Simulation Study Two. FIX; fixed sample size design, POC; Pocock stopping rule, OBF; O'Brien-Fleming stopping rule, INB; Incremental net benefit, QALY; Quality adjusted life year

has been estimated using the value taken from the CACTUS pilot of 0.07 (described in Section 5.6.3). There are high levels of bias compared to the pilot trial estimate in the simulated trials, perhaps suggesting that the pilot did not provide a good estimate of the parameter. The bias in the utility improvement increases with the correlation between primary and health economic outcome, for both the adjusted and unadjusted estimates.

The fourth row of Figure 7.8 gives the standardised bias for the probability of good response at 6-months parameter for the health economic model. This parameter, described in Section 6.5.1.1 is the probability that a participant in the intervention arm has an improvement of 17% or more between baseline and 6-months follow-up. Again this parameter was not specified in the data generation and is compared to the value estimated from the CACTUS pilot of 0.533 (see Section 5.6.2). The model-based parameter from the simulation study is under estimated compared to the value from the pilot trial for both adjusted and unadjusted estimates. This suggests that the pilot trial may not have provided a true estimate of the probability of a good response given its small sample size. The unadjusted and adjusted estimates do not appear to be affected by the correlation chosen. The bias is greatest for the unadjusted estimates compared to adjusted estimates.

Row five gives the standardised bias for the relapse rate. As with the probability of good response at 6-months this is not specified in the data generation, instead this model parameter is estimated from the probability of a good response at 6 and 9-months. The unadjusted estimates have a smaller bias in this case, with the exception of POC with five analyses, where the adjusted estimates are close to the estimate of 0.08% from the pilot trial.

Row six gives the baseline utility model parameter for each of the scenarios considered. Here the adjusted and unadjusted estimates are set to be equal as this is a baseline parameter it is not expected to be affected by the adaptive design. The bias in this parameter is small for all scenarios, with some over estimation for high correlations perhaps suggesting these scenarios differ from the true underlying population that was captured in the CACTUS pilot trial.

The final row of Figure 7.8 gives the standardised bias for the improvement in percentage words named correctly from baseline to 9-months follow-up. This parameter is used to estimate the probability of good response at 9-months as described in Section 6.5.1.1. The unadjusted

estimate of this parameter has greatest bias for POC with five analyses and high correlation values. The adjusted estimates provide a good reduction in the bias in the estimate.

Figure 7.9 gives the RMSE for the health economic model parameters estimated from the trial data. The RMSE is similar for each of the adjusted and unadjusted point estimates of the parameters specified in the data generation (resource costs, baseline utility and improvement in percentage words at 9-months). The RMSE for the utility improvement follows a similar pattern as for the standardised bias with increasing RMSE for correlations up to 0.6 followed by a reduction when the correlation is 0.8. The adjusted estimates of probability of a good response and relapse rate have a slightly lower RMSE compared to the unadjusted estimates. This may potentially reflect their different estimation methods as discussion in Section 6.5.1.1.

Table 7.8 gives the unadjusted incremental analysis for the deterministic model based results for FIX and adjusted and unadjusted analysis for the adaptive designs with five analyses. Correlation values of $0.0, 0.4, 0.6, 0.8$ are reported.

For FIX, as the correlation increases so does the deterministic model estimate of INB. This is driven by changes in the QALY estimate in the intervention arm. As the correlation increases this increases the correlation between the primary outcome and the estimate of the model parameter utility improvement as shown in Table 7.6. By the design of the health economic model, described in Chapter 5, as the estimate of the utility improvement increases the QALY in the intervention arm increases, hence increasing the INB. This pattern is mirrored in the adaptive designs considered too.

Comparing the adaptive designs to FIX, the adjusted and unadjusted adaptive designs tend to overestimate the estimate of the INB. For all designs, these differences are small and would not lead researchers to draw different conclusions about the cost-effectiveness of the computer-based intervention compared to the control.

Figure 7.10 compares the adjusted and unadjusted estimate of the INB from the health economic model. The percentage difference between estimates is small for all designs (less than approximately 4.5%). Patterns of differences across correlations are similar for each design with differences greatest when correlations reach 0.6 and decreasing again for 0.8 correlation. This reflects the patterns seen in Figure 7.8. POC with five analyses appears to the smallest

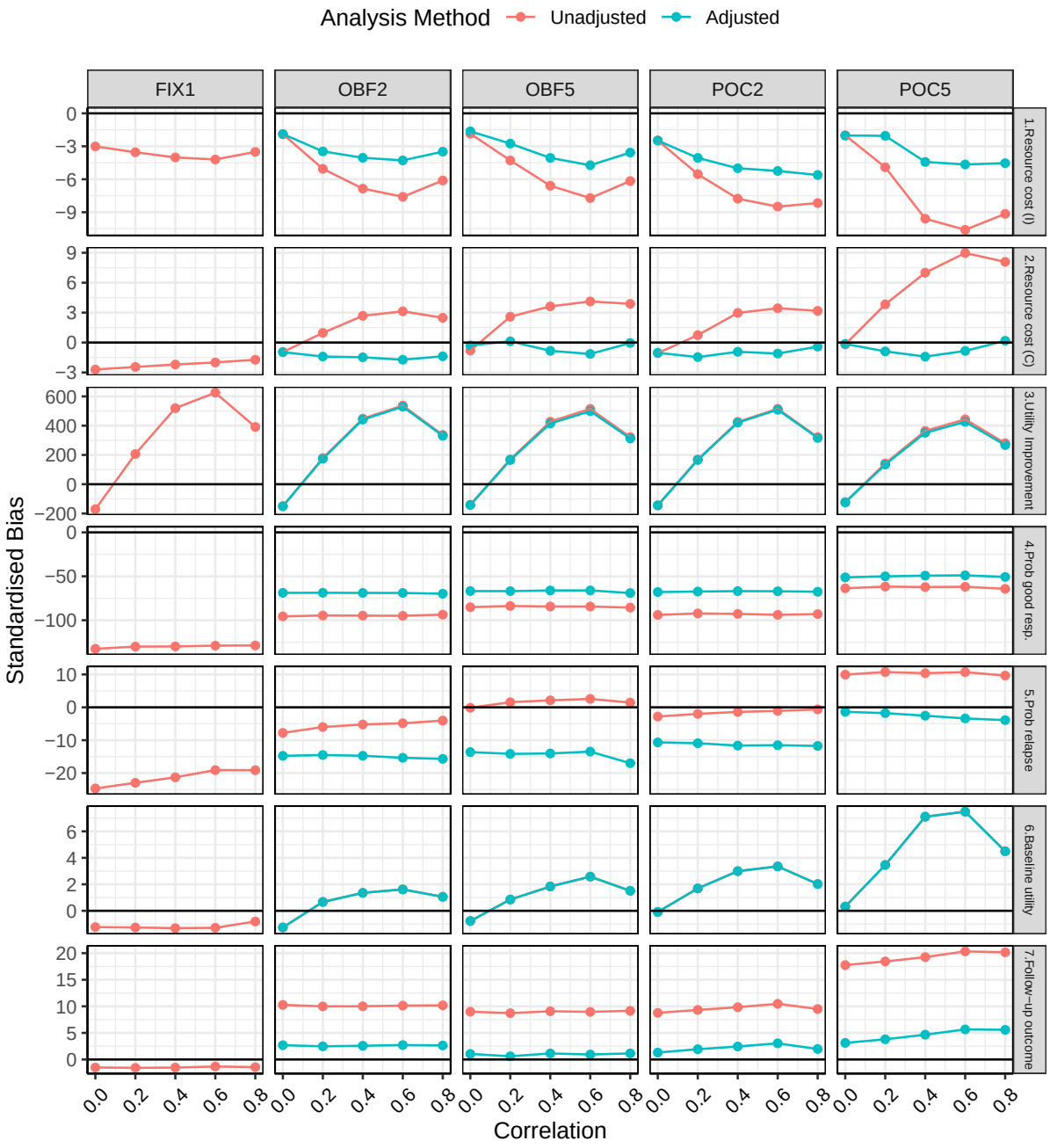


FIGURE 7.8: Results for the standardised bias for health economic model parameters for the five designs considered under a range of correlations between outcomes for Simulation Study Two. FIX; fixed sample size design, POC; Pocock stopping rule, OBF; O'Brien-Fleming stopping rule

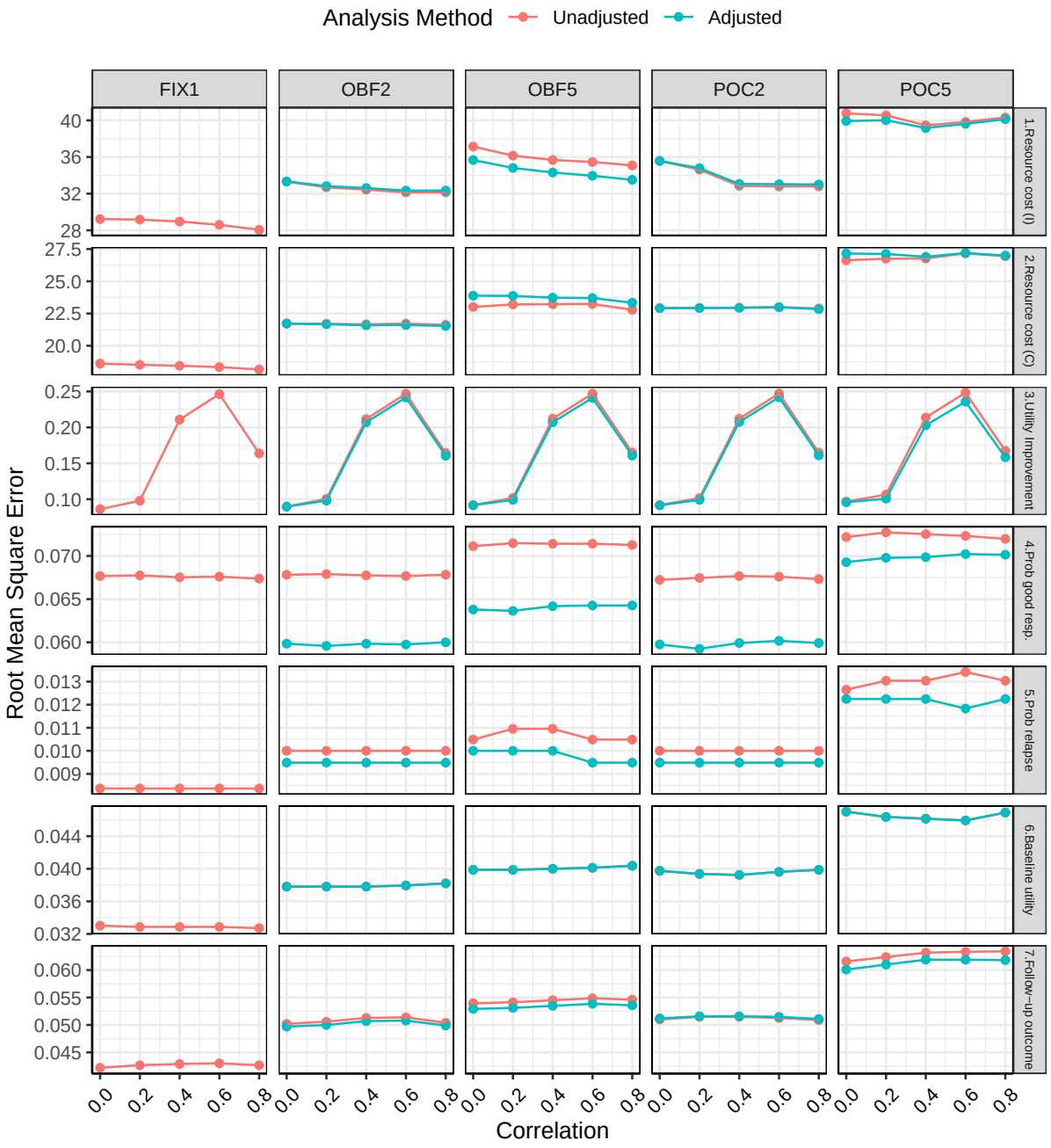


FIGURE 7.9: Results for the root mean square error for health economic model parameters for the five designs considered under a range of correlations between outcomes for Simulation Study Two. FIX; fixed sample size design, POC; Pocock stopping rule, OBF; O'Brien-Fleming stopping rule

	Unadjusted					Adjusted				
	Per Person		Incremental			Per Person		Incremental		
	Cost	QALY	Cost	QALY	INB	Cost	QALY	Cost	QALY	INB
FIX 0.0 Correlation										
Control	18654.53	3.07	-	-	-	-	-	-	-	-
Intervention	19089.96	3.10	435.43	0.03	214.51	-	-	-	-	-
FIX 0.4 Correlation										
Control	18661.19	3.07	-	-	-	-	-	-	-	-
Intervention	19094.77	3.65	433.58	0.58	11183.00	-	-	-	-	-
FIX 0.6 Correlation										
Control	18663.94	3.07	-	-	-	-	-	-	-	-
Intervention	19097.14	3.72	433.20	0.65	12630.82	-	-	-	-	-
FIX 0.8 Correlation										
Control	18667.58	3.07	-	-	-	-	-	-	-	-
Intervention	19101.58	3.55	434.00	0.48	9105.62	-	-	-	-	-
OBF5 0.0 Correlation										
Control	18676.45	3.07				18684.45	3.07			
Intervention	19111.27	3.11	434.81	0.04	320.98	19119.20	3.11	434.76	0.04	357.98
OBF5 0.4 Correlation										
Control	18747.19	3.08	-	-	-	18675.61	3.08	-	-	-
Intervention	19168.89	3.64	421.69	0.56	10869.64	19107.05	3.67	431.44	0.59	11409.11
OBF5 0.6 Correlation										
Control	18755.15	3.08	-	-	-	18670.50	3.08	-	-	-
Intervention	19174.42	3.72	419.26	0.64	12307.29	19101.27	3.75	430.77	0.67	12916.92
OBF5 0.8 Correlation										
Control	18750.14	3.08	-	-	-	18688.41	3.08	-	-	-
Intervention	19172.54	3.54	422.40	0.47	8879.79	19119.89	3.57	431.49	0.49	9344.47
POC5 0.0 Correlation										
Control	18685.75	3.07	-	-	-	18686.64	3.07	-	-	-
Intervention	19119.28	3.12	433.54	0.04	419.75	19120.15	3.12	433.51	0.04	444.56
POC5 0.4 Correlation										
Control	18818.12	3.09	-	-	-	18663.02	3.09	-	-	-
Intervention	19227.83	3.66	409.71	0.56	10875.81	19093.69	3.67	430.67	0.58	11087.40
POC5 0.6 Correlation										
Control	18856.64	3.09	-	-	-	18673.43	3.09	-	-	-
Intervention	19261.50	3.73	404.87	0.64	12318.45	19102.81	3.74	429.38	0.65	12600.78
POC5 0.8 Correlation										
Control	18839.01	3.09	-	-	-	18692.50	3.09	-	-	-
Intervention	19247.70	3.55	408.69	0.47	8928.87	19120.66	3.56	428.15	0.48	9084.75

TABLE 7.8: Summary of the unadjusted and adjusted deterministic model based incremental analysis for the the fixed sample size design and adaptive designs with correlations 0, 0.4, 0.6, 0.8 based on 2,000 simulated trials for Simulation Study Two FIX; fixed sample size design, POC; Pocock stopping rule, OBF; O'Brien-Fleming stopping rule

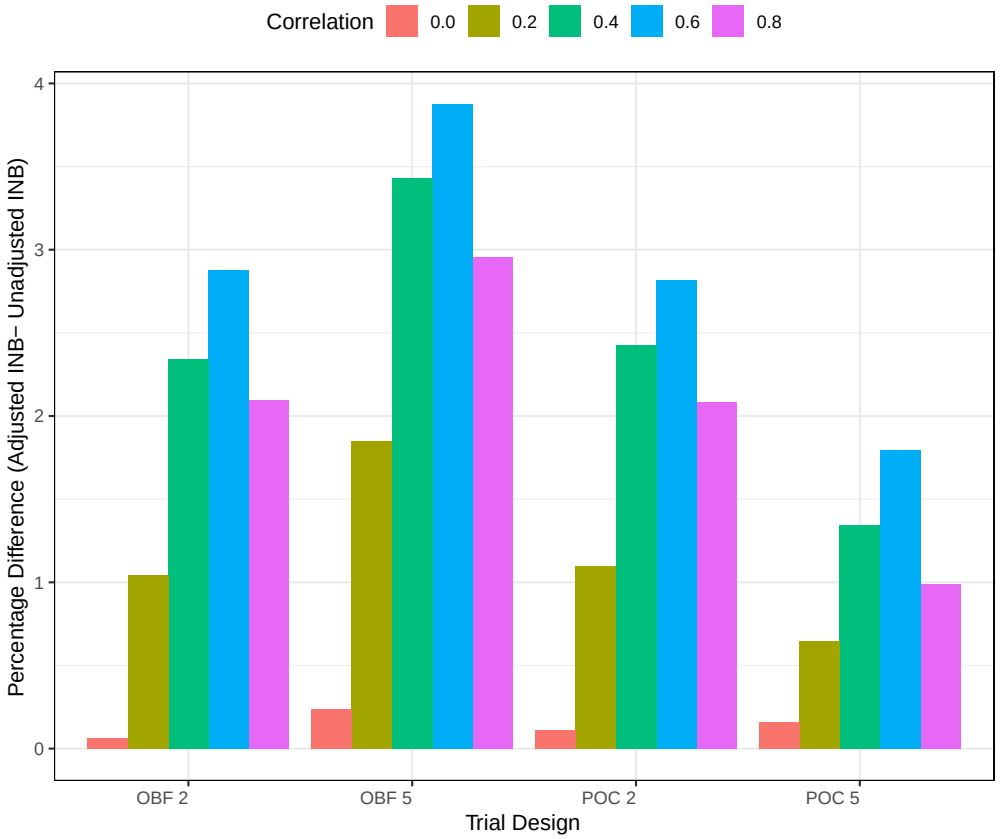


FIGURE 7.10: Difference between adjusted and unadjusted estimates of the INB estimated from the deterministic health economic model based on 2,000 simulated trials for Simulation Study Two. FIX; fixed sample size design, POC; Pocock stopping rule, OBF; O'Brien-Fleming stopping rule

differences between adjusted and unadjusted estimates, despite large differences in estimates of costs in Table 7.8. This suggests that the difference between the estimates due to the adaptive design is in the opposite direction to the differences due to the analysis methods, discussed in Section 7.7.4.

7.7.4 Impact of Analysis Methods on the Results

In the simulation studies, only unadjusted point estimates and confidence intervals have been calculated for FIX, as there are no early examinations of the data. In Chapter 6, a bootstrapping approach is described to calculate adjusted point estimates for the probability based health economic model parameters. This contrasts to the unadjusted method where the point estimate is calculated directly from the trial data. It would be possible to use the bootstrapping approach to calculate the unadjusted estimate; however, it was felt that this would not reflect how the unadjusted analysis would be conducted in practice. A researcher is unlikely to consider the

bootstrapping approach when they can calculate the estimate they require directly from the trial data.

A consequence of these different approaches is that some of the differences between adjusted and unadjusted estimates from the adaptive designs might be due to the bootstrapping method rather than the bias due to the adaptive design. However, what the analysis presented does so is how the adjusted analysis will contrast from the likely unadjusted analysis conducted in practice.

7.8 Discussion

7.8.1 Summary of Key Findings from this Chapter

In this chapter the extent to which adaptive designs affect a within trial and model-based health economic analysis following the trial has been explored. This was considered firstly in a simplified context assuming clinical and health economic outcomes follow a multivariate Normal distribution for a within trial analysis. This was then extended to a more realistic setting based on the CACTUS case study, described in Chapter 5, where the Normality assumption was relaxed and both a within trial and model-based analyses evaluated.

In the first simulation study, high levels of correlation between primary and health economic outcomes were shown to introduce bias in the point estimates, including the INB. The bias was greatest for POC with five analyses. For this design, when the correlation was high (equal to one) the sign of the INB changed in the unadjusted analysis. This could lead researchers to draw the wrong conclusion about the short-term cost-effectiveness of the intervention compared to the control. This highlights the potential for an adaptive design to impact a within trial health economic analysis and the importance of conducting an analysis that accounts for the adaptive nature of the trial.

The adjusted analysis used the BAMLE introduced in Section 2.5 and extended to within trial health economic outcomes in Section 6.4. In the first simulation study, the adjusted approach reduced some of the bias in the point estimates; however, it did not eradicate it completely. Adjusted confidence intervals based on the SMO approach, described in Section 2.5.4.2, provided

wider intervals reflecting the greater uncertainty when using an adaptive design. These intervals also have coverage close to 0.95, however, there is slight under coverage for trials with a Pocock stopping rule and high correlation between outcomes.

In the second simulation study, the unadjusted within trial primary outcome and costs were vulnerable to bias especially for correlations of 0.4 to 0.8. The bias was highest when the correlation level was 0.6 rather than 0.8 due to the structure of the underlying data. The QALY and INB were less affected by the design and the correlation value than in the simplistic scenario. The adjusted estimates provided a good reduction in the bias for the within trial health economic analysis. Despite high correlation between the variables relating to the primary outcome and utility at a given follow-up time point, this did not translate into a strong correlation between the primary outcome and the QALY. This suggests that the structure of the data may affect the levels of correlation between variables and hence the affect of an adaptive design on the health economic analysis.

Comparing the deterministic model results for the adaptive designs (POC and OBF) to FIX showed the unadjusted results for each adaptive design slightly under-estimated the INB of the computer-based intervention compared to control. However, these values were far from zero and so the same conclusion regarding cost-effectiveness were made in each case. In scenarios where the INB estimate is closer to zero it is possible that unadjusted and adjusted analyses could give conflicting results. This could result in resources being wasted as incorrect decisions about the allocation of resources are made based on inaccurate unadjusted analyses.

The small difference between the unadjusted fixed and adaptive designs in the CACTUS case study could be explained by:

- Low levels of correlation between the primary outcome and the health economic model parameters.
- The influence of the trial based model parameters on the model. If the model had used the primary outcome directly, for example to inform transition probabilities, there may be more bias in the health economic results reflecting bias levels in the primary outcome.
- The influence of model parameters from external sources might have diluted any effect of the adaptive design on the health economic results.

Based on the findings of this chapter alone it is difficult to prove or disprove these explanations in a single case study and so further work should explore the impact of the adaptive design on the health economic analysis in case studies with different characteristics.

7.8.2 How this fits with Existing Literature

The BAMLE was able to reduce considerably the bias in the outcomes considered, however it did not remove all bias. Similar results were found by Emerson *et al.*, 1990 and Todd *et al.*, 1996. Wassmer *et al.*, 2016 suggest this is due to the fact the MLE does not varying symmetrically around its mean as its (sequential) distribution is truncated by the stopping rule. The BALMEs also had smaller RMSE as was found by Emerson *et al.*, 1990 and Liu *et al.*, 2008 suggesting the adjustments gave point estimates with reduced bias and higher accuracy.

The results in this chapter reflect previous research that showed the bias in the point estimates of the primary outcome was greater for POC compared to OBF (Pinheiro, 1997; Li *et al.*, 1999). This occurs because it is difficult to stop at the earliest interim analyses for OBF, as discussed in Section 2.4. This means the point estimates observed under POC are likely to be extreme as they are based on less data. Likewise, the designs with a greater number of interim analyses resulted in greater bias as the early interim analyses took place after only a small number of participants had reached the outcome of interest. As shown in Figure 2.3, there is much greater variation in the estimate of the primary outcome when the sample size is small.

Skalland, 2015 compared the SMO approach for estimating confidence for secondary outcomes to the approach described by Whitehead *et al.*, 2000. The authors compared the width and coverage of 95% confidence intervals for a single arm trial with POC and OBF rules, three analyses of the data, a correlation of 0.2 and 0.8 between the primary and secondary outcome. As in Simulation Study One, they found that the width of the secondary intervals following an OBF stopping rule were largely unaffected by the correlation value. However, there were greater differences between the width and coverage depending on correlation for the POC stopping rule.

The simulation studies have focussed on showing how the choice of stopping rule and number of interim analyses can affect the health economic analysis following an adaptive design. The

impact of the primary effect size, health economic outcome effect sizes, the primary and secondary variances or the timing of the interim analyses on the analysis were not considered. The influence of the effect size of the primary outcome on the bias in the point estimate following a group sequential design has been well discussed in the literature. Wassmer *et al.*, 2016 explain that the bias in unadjusted standardised mean is higher for moderate values of the parameter as only extreme values of the primary outcome will cross the stopping boundary at an interim analysis and cause the trial to stop early, thus resulting in an exaggerated point estimate. The bias is smaller for larger absolute values of the parameter as the point estimate required to stop the trial is now relatively less extreme. The bias is smallest when the true parameter value is close to zero as the over or under estimates of the parameter cancel each other out, as it is equally likely that the trial will cross the upper and lower boundaries. By similar reasoning, the bias in the secondary outcome will follow the same pattern, further explaining the small levels of bias observed in the QALY outcome for the CACTUS case study. This should be explored in further work.

7.8.3 Considerations for Practice

It is recommended that adjusted and unadjusted point estimates and confidence intervals are presented for primary and health economic (secondary) outcomes with a clear statement of the methods used. This reflects the guidelines recommended in the Adaptive Designs CONSORT extension (Dimairo *et al.*, 2019b). This will allow researchers to establish the impact of the design on their analyses as well as allow users of the research to choose the appropriate estimates for their own analyses.

High levels of correlation between the primary and health economic outcomes have the largest impact on the results. It could be argued these high correlations are unlikely to be observed in practice. However, the estimated correlation between the QALY and net benefit was found to be is high regardless of the pairwise correlation between the primary outcome and QALY. This is not surprising when considering the formula for calculating the net benefit in Equation 2.16. A preference based primary outcome may have these high correlation values with QALY and INB. This suggests that point estimates and confidence intervals need to be carefully adjusted to account for any potential bias as the design of the trial has the potential to change the conclusions drawn about cost-effectiveness.

7.8.4 Strengths and Limitations

This is the first simulation study to consider the impact an adaptive design has on the estimation of a within trial and model-based health economic analysis. As described in Chapter 5, the within trial analysis of the CACTUS case study is typical of analyses conducted in the UK setting to provide evidence of the short-term cost-effectiveness of an intervention. It can be argued that the adjustment methods considered are generalisable to other within trial analyses collecting cost and outcome data. The impact of the adaptive design on the within trial analysis is likely to be context dependent and as shown in this chapter will depend on the levels of correlation between the health economic outcomes and the primary outcome in the trial.

The adjustments and results from the model-based analysis are less generalisable as they are specific CACTUS health economic model and trial context. However, a number of scenarios have been considered increasing the generalisability of the results and demonstrating the potential impact of an adaptive clinical trial design on a health economic analysis.

The health economic model used does not allow participants in the control of the trial to move from the aphasia state to the response state. This reflects the CACTUS pilot analysis where none of the control arm patients had a response to treatment. As such, there was no data to estimate the probability of a good response for participants in the control arm. The lack of data would have made it difficult to assess the bias in the estimate as was summarised for the probability of a good response for participants in the control arm. Further work could explore the affect including this parameter in the model has on the results but it is unlikely to change the conclusions regarding the impact of a group sequential design on a health economic analysis.

The simulation studies were limited to 2,000 iterations due to the computational intensity of the steps involved; applying the adaptive design to simulated data and calculating adjusted point estimates and confidence intervals. It was not possible to use the high performance computing at the University of Sheffield to increase the speed of the calculations as the `RCTdesign` package, that plays a fundamental role in the calculation of the stopping boundary and adjusted primary outcome, is not available on Linux based systems.

In the simulated trials, it has been assumed that all data are complete for all outcomes. This is a simplified scenario as data from trials is often incomplete, especially health economic outcomes (Faria *et al.*, 2014). The impact of missing data on the health economic analysis following an adaptive design should be explored in further work.

7.8.5 Recommendations for Further Methodological Research

As per the recommendations of Emerson *et al.*, 1990, a SMO approach was used to calculate a 95% confidence interval for the primary outcome following an adaptive trial. For consistency, the SMO approach was also used to calculate intervals for health economic outcomes. Skalland, 2015, however, suggest that the Whitehead *et al.*, 2000 approach could give shorter interval whilst still giving the desired coverage levels. Further work could calculate the Whitehead *et al.*, 2000 intervals for the primary and secondary outcomes and compare these to the SMO intervals considered in this thesis.

Further issues when an adaptive design stops early, not considered in this analysis, include less heterogeneity in the trial data. For example, participants may be from homogeneous groups, such as the same geographic area, if a trial stops before all centres are recruiting. This could lead to less data on participants from subgroups where the intervention may be more clinically- and cost-effective. In the Big CACTUS trial, the intervention was unlikely to be cost-effective in the whole trial population, however there was potentially for the computer-based intervention to be cost-effective in patients with mild to moderate word finding ability (Palmer *et al.*, 2019). If the trial used an adaptive design that had stopped at an early interim analysis it is possible the number of participants in this subgroup would be too small to explore this. Further work could assess other potential impacts that an adaptive design might have on a health economic analysis beyond the accuracy of the analysis.

As discussed, the generalisability of these findings is limited to some extent by the context and health economic model of the CACTUS case study. Further work should explore the impact of an adaptive design on a health economic analysis using different clinical trial settings and health economic models that use the trial data to varying degrees.

7.8.6 Considerations for the Calculation of Expected Value of Sample Information for Adaptive Trials

There is clear potential for the adaptive nature of a clinical trial to impact a health economic analysis. This chapter has considered the health economic analysis following an adaptive clinical trial; however, other types of health economic analyses may be affected when adaptive designs are used. Chapter 8 considers opportunities to increase the efficiency of adaptive clinical trials by extending the methods of EVSI to guide the design of adaptive clinical trials. It will be important in this chapter to understand the potential for the adaptive designs considered to impact the EVSI calculations and to explore the extent to which the adjustment methods described in Chapter 6 account for this.

7.9 Chapter Summary

This chapter builds on the theory development of Chapter 6 that discusses how existing bias adjustment methods can be operationalised in a health economic analysis following an adaptive clinical trial. Using simulations, the health economic analysis has been shown to be sensitive to the choice of stopping rule, number of interim analyses and correlation between primary and health economic outcomes. The risk of bias is greatest when there is high correlation between the primary and health economic outcomes and when a Pocock stopping rule is used with up to five analyses. In the real-world setting, the impact of an adaptive design might be reduced as observed correlations between the primary outcome and health economic outcomes may be low, however, it is not always possible to know this in advance of the data collection. Therefore, it is recommended that both adjusted and unadjusted analyses are presented when conducting a health economic analysis that uses data from a clinical trial with an adaptive design.

Chapter 8 uses these adjustments when calculating the EVSI to guide the design of an adaptive clinical trial.

Chapter 8

Using Health Economics to Guide the Design of a Proposed Adaptive Clinical Trial – Expected Value of Sample Information for Group Sequential Designs

8.1 Introduction

Value of information analysis (VOIA) methods, described in Section 2.8, allow researchers to quantify their decision uncertainty when making resource allocation decisions. Their use is increasing in the UK healthcare setting, to help guide and inform the allocation of scarce healthcare budgets (Steuten *et al.*, 2013; Mohiuddin *et al.*, 2014; Welton *et al.*, 2015).

Using VOIA in adaptive trials has been discussed by authors in the literature identified in Chapter 2. Despite these methodological developments, the use of methods in the design and monitoring of adaptive designs in practice has been limited as reported in Chapter 3. Possible explanations, based on the qualitative study in Chapter 4, are the lack of familiarity with value of information methods and the computational intensity and complexity of current approaches.

Additionally, qualitative study participants were unanimous that, while cost-effectiveness considerations are important, clinical effectiveness should remain the focus of an adaptive clinical trial.

In Chapter 2 the methods for adjusting the analysis following a group sequential design were outlined. In Chapter 6 these methods were considered in the context of a health economic analysis following a group sequential design. Based on the exploration of these methods under a range of scenarios, in Chapter 7, it is recommended that both adjusted and unadjusted health economic analyses are conducted. This is especially important when there is little prior information available to the research team about the potential correlation between primary and health economic outcomes when designing the trial.

It is currently unclear whether adjustments to the EVSI methods are needed when the trial design under evaluation is an adaptive design, specifically a group sequential design.

8.2 Chapter Aims

This chapter aims to address the fourth research aim to extend existing methods of health economics to guide the design of an adaptive design whilst appropriately accounting for the adaptive nature of the trial design. The aims of the chapter are achieved by

1. Summarising the EVSI approach for the design of fixed design clinical trials, specifically using the non-parametric regression approach.

2. Extending the existing theory to guide the design of adaptive clinical trials, outlining changes required to calculate the EVSI in the adaptive design setting.

3. Considering the costs of conducting the adaptive clinical trial, providing practical guidance on how this can be calculated by extending current cost calculations for fixed sample size designs.

4. Considering appropriate adjustments to account for the adaptive nature of the trial, using the theory developed in Chapter 6, and comparing these to unadjusted approaches to explore the impact on the calculation of EVSI.

Method	Description
Expected Value of Perfect Information (EVPI)	The EVPI considers the scenario where further research would eliminate all decision uncertainty,(Chilcott *et al.*, 2003) representing the most that can be gained from further research. Drummond *et al.*, 2015. The EVPI can take only positive values.(Griffin *et al.*, 2010) It is potentially worthwhile conducting further research if the associated costs are less than the EVPI.
Expected Value of Partially Perfect Information (EVPPI)	It may not be feasible to collect further information about all the parameter inputs in a health economic analysis. Instead, the EVPPI can be calculated using the same idea as the EVPI calculation (Welton *et al.*, 2008)
Expected Value of Sample Information (EVSI)	The methods of EVPI and EVPPI estimate the value of eliminating all or an element of uncertainty in a decision problem. It is not, however, always possible to do this. It may be more feasible to consider the value of reducing some of the uncertainty (Ades *et al.*, 2004), for example, by conducting another clinical trial or continuing with an adaptive design when presented with interim data. The EVSI can be used to determine the value of a specific research design that will be used to inform a decision (Strong *et al.*, 2015).
Non-Parametric Regression	Developed by Strong *et al.*, 2015 this methods facilitates the calculation of EVPI, EVPPI and EVSI using the output of the probabilistic sensitivity analysis and generalised additive models.

TABLE 8.1: Summary of existing expected value of information analysis methods

5. Applying the approach to the CACTUS case study , described in Chapter 5, considering five group sequential trial designs including an O'Brien-Fleming or Pocock stopping rule with up to five analyses, low, medium and high correlations between primary and health economic outcomes and intervention costs for the proposed trial increased by up to 15 times.

8.3 Expected Value of Sample Information for Fixed Sample Size Designs

8.3.1 Overview of Existing Value of Information Methods

The methods of expected value of information are summarised in Table 8.1. To conduct a VOIA

for a fixed sample size design Fenwick *et al.*, 2020 propose seven steps:

1. Conceptualise and construct a health economic model.
2. Parameterise with evidence.
3. Generate a probability sensitivity analysis sample.
4. Identify uncertainty.
5. Establish whether more research is worthwhile.
6. Estimate the value of specific research.
7. Iterate with new evidence.

The following sections describe how each of these steps are applied in the context of a fixed sample size design. These steps are illustrated using the non-parametric regression approach introduced in Section 8.3.9.1.

8.3.2 Conceptualise and Construct a Health Economic Model and Parametrise with Evidence

A health economic model is constructed for the population of interest. This may be an existing model that has been developed for the disease of interest or a model from previous work such as a pilot study. In the CACTUS case study, in Chapter 5, a health economic model was built as part of the CACTUS pilot trial. This model can be used to inform the design of the subsequent full-scale trial. Alternative forms of prior evidence might include an observational study or expert opinion.

8.3.3 Generate the Probabilistic Sensitivity Analysis Sample and Assess of Uncertainty

A probabilistic sensitivity analysis (PSA) is generated, as described in Section 2.7.8, that is based on available prior evidence for the model parameters and denoted by θ. Let the PSA sample be

denoted by

$$\left\{\theta^{(1)}, \ldots, \theta^{(N_{PSA})}\right\}, \tag{8.1}$$

where N_{PSA} is the number of PSA samples.

For each row of the PSA sample the health economic model is evaluated to give a per person expected net benefit for each intervention. This is denoted by

$$\left\{NB\left(d, \theta^{(1)}\right), \ldots, NB\left(d, \theta^{(N_{PSA})}\right)\right\}, \tag{8.2}$$

where d represents the interventions under consideration. The uncertainty in the decision can be assessed visually using a cost-effectiveness plane. This plots the difference in costs and difference in effects for each PSA row on the same plane. The chosen cost-effectiveness threshold can be included, for example the £20,000 per QALY threshold discussed in Section 2.7.5.

8.3.4 Establish whether Further Research is Worthwhile

As described in Table 8.1, the EVPI can be calculated to establish whether it is worthwhile conducting further research regardless of the proposed design. This is calculated using

$$\text{EVPI} = \mathbb{E}_\theta\left[\max_d \left[\text{NB}(d, \theta)\right]\right] - \max_d \left[\mathbb{E}_\theta\left[\text{NB}(d, \theta)\right]\right]. \tag{8.3}$$

8.3.5 Estimate the Value of Specific Research

If the EVPI calculation suggests further research is worthwhile, the value of different research designs can be estimated by calculating the EVSI and ENBS as described in Section 2.8.2. This is broken down into six steps, with each discussed in detail:

1. Identifying the trial designs for comparison.

2. Simulating the trial results and analysis data sets.

3. Calculating summary statistics.

4. Calculating the EVSI.

5. Calculating the cost of sampling.

6. Comparing the ENBS of the proposed trial designs.

8.3.6 Identifying the Trial Designs for Comparison

The first step for applying the EVSI approach to guide the design of a fixed sample size trial is to choose the trial designs under consideration. This might include comparing the design of trials with a range of sample sizes, a different number of treatment arms or different lengths of follow-up (Tuffaha *et al.*, 2016).

8.3.7 Simulating the Trial Results and Analysis Datasets

A trial result dataset is simulated for each row of the PSA sample. This is based on the likelihood function for the pilot trial (or existing information in another form such as an observational study) to give a dataset representative of the population to be randomised into the future trial. The data simulation is informed by the PSA parameter estimates, to give a dataset in each row denoted by

$$\left\{x^{(1)}, \ldots, x^{(N_{PSA})}\right\}. \tag{8.4}$$

The trial analysis for the design under consideration is applied to each trial result dataset. This gives a trial analysis dataset denoted by

$$\left\{y^{(1)}, \ldots, y^{(N_{PSA})}\right\}. \tag{8.5}$$

8.3.8 Calculating Summary Statistics

Summary statistics for primary and secondary outcomes informing the health economic model are estimated from the trial analysis dataset in each row of the PSA sample. This will include the primary and secondary clinical outcomes and health economic outcomes such as healthcare resource use and health related quality of life. These statistics are denoted by

$$\left\{\hat{T}\left(y^{(1)}\right), \ldots, \hat{T}\left(y^{(N_{PSA})}\right)\right\}. \tag{8.6}$$

8.3.9 Calculating the Expected Value of Sample Information

Adopting the notation of Strong *et al.*, 2015, the EVSI is the difference between the expected net benefit given sample information minus the expected net benefit given current information. The health economic model has input parameters (θ) to estimate the net benefit of each intervention ($d = 1, \ldots, D$) under consideration. For simplicity, in this chapter θ is used to denote all health economic model parameters, including those estimated from trial data in just one arm of the trial trial denoted by ν in previous chapters and parameters estimated outside the trial. The net benefit is denoted by $\mathrm{NB}(d, \theta)$ as given in Section 2.8 and Section 8.3.3. This gives a per person EVSI of

$$\mathrm{EVSI} = \mathbb{E}_Y \left[\max_d \left[\mathbb{E}_{\theta|Y} \left[\mathrm{NB}(d, \theta) \right] \right] \right] - \max_d \left[\mathbb{E}_\theta \left[\mathrm{NB}(d, \theta) \right] \right], \tag{8.7}$$

for data Y to be collected. A population level EVSI is estimated by multiplying the individual level EVSI by the time horizon and the annual prevalence for the population under consideration to give (Welton *et al.*, 2013)

$$\mathrm{popEVSI} = \mathrm{EVSI} \times T \times N_p, \tag{8.8}$$

where

N_p is the annual prevalence for the population under consideration,

T is the time horizon of the decision problem.

8.3.9.1 Non-parametric Regression

Strong *et al.*, 2015 have developed a method that is flexible and computationally efficient for EVSI calculations that is adopted in this thesis. It uses non-parametric regression and so does not require the existence of conjugate distributions or parametric assumptions to be made (Ades *et al.*, 2004; Brennan *et al.*, 2007; Strong *et al.*, 2015). Instead the output of the PSA, described in Section 2.7.8.1 is used. After conducting the steps in Sections 8.3.3, 8.3.7 and 8.3.8

the PSA sample will resemble Matrix 8.9

$$\begin{bmatrix} 1 & \theta^{(1)} & \mathrm{NB}_{d_1}^{(1)} & \mathrm{NB}_{d_2}^{(1)} & x^{(1)} & y^{(1)} & T(y^{(1)}) \\ 2 & \theta^{(2)} & \mathrm{NB}_{d_1}^{(2)} & \mathrm{NB}_{d_2}^{(2)} & x^{(2)} & y^{(2)} & T(y^{(2)}) \\ 3 & \theta^{(3)} & \mathrm{NB}_{d_1}^{(3)} & \mathrm{NB}_{d_2}^{(3)} & x^{(3)} & y^{(3)} & T(y^{(3)}) \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \\ N_{PSA} & \theta^{(N_{PSA})} & \mathrm{NB}_{d_1}^{(N_{PSA})} & \mathrm{NB}_{d_2}^{(N_{PSA})} & x^{(N_{PSA})} & y^{(N_{PSA})} & T(y^{(N_{PSA})}) \end{bmatrix} \tag{8.9}$$

One of the difficult and computationally expensive aspects of EVSI calculation is calculating

$$\mathbb{E}_{\theta|Y}\left[\mathrm{NB}(d,\theta)\right]. \tag{8.10}$$

$\mathrm{NB}(d,\theta)$ is unknown and data (Y) are collected to calculate the expected $\mathrm{NB}(d,\theta)$ given the data (Y). As this value will not be the true $\mathrm{NB}(d,\theta)$ an error term (ϵ) is required. Each time the conditional expectation of the NB is calculated for each value of the trial result datasets ($Y = Y^{(i)}$) a different value for $\mathbb{E}_{\theta|\mathbf{Y}}\left[\mathrm{NB}(d,\theta)\right]$ is obtained. Therefore the conditional expectation can be thought of as a function of θ, denoted by $g(d,\theta)$

$$\mathrm{NB}(d,\theta) = g(d,\theta) + \epsilon. \tag{8.11}$$

As the data might be highly dimensional, in the case of a trial with multiple outputs, a summary statistic $T(Y)$ can be used giving

$$\mathrm{NB}(d,\theta) = g\left(d, T(Y)\right) + \epsilon. \tag{8.12}$$

This can be treated as a non-parametric regression problem. The results of the PSA can be thought of as 'noisy' data about the model inputs (θ) that can be used to estimate the dependent variable $\mathrm{NB}(d,\theta)$. Plotting the sample values of $\mathrm{NB}(d,\theta)$ against the sampled values for the parameters (θ) a regression equation can be fitted. The equation of the regression line is the expected $\mathrm{NB}(d,\theta)$ given the data Y, hence the conditional expectation $E_{\theta|Y}\left[\mathrm{NB}(d,\theta)\right]$. Equation

8.7 can now be written as

$$\text{EVSI} = \mathbb{E}_Y \left[\max_d \{g(d, T(Y))\} \right] - \max_d \{\mathbb{E}_Y [g(d, T(Y))]\} . \tag{8.13}$$

As the form of $g(\cdot)$ is not directly of interest, just the value it takes for the sample of parameter values, a generalised additive model (a nonparametric regression approach) can be used (Hastie *et al.*, 1986). This fits the model using smoother functions such as splines (Strong *et al.*, 2015). By estimating the form of the function $g(\cdot) = \hat{g}(\cdot)$ the conditional expectation can be estimated from these fitted values. Therefore in each row of the PSA fitted values can be calculated $\hat{g}\left(d, T\left(y^{(i)}\right)\right)$ that estimate $\mathbb{E}_{\theta|Y^{(i)}}[\text{NB}(d, \theta)]$ as illustrated in Matrix 8.14.

$$\begin{bmatrix} \theta^{(1)} & \text{NB}_{d_1}^{(1)} & \text{NB}_{d_2}^{(1)} & x^{(1)} & y^{(1)} & T(y^{(1)}) & \hat{g}\left(d, T(y^{(1)})\right) \\ \theta^{(2)} & \text{NB}_{d_1}^{(2)} & \text{NB}_{d_2}^{(2)} & x^{(2)} & y^{(2)} & T(y^{(3)}) & \hat{g}\left(d, T(y^{(2)})\right) \\ \theta^{(3)} & \text{NB}_{d_1}^{(3)} & \text{NB}_{d_2}^{(3)} & x^{(3)} & y^{(3)} & T(y^{(3)}) & \hat{g}\left(d, T(y^{(3)})\right) \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ \theta^{(N_{PSA}} & \text{NB}_{d_1}^{(N_{PSA})} & \text{NB}_{d_2}^{(N_{PSA})} & x^{(N_{PSA})} & y^{(N_{PSA})} & T(y^{(N_{PSA})}) & \hat{g}\left(d, T(y^{(N_{PSA})})\right) \end{bmatrix} \tag{8.14}$$

The EVSI equation can now be written as

$$\widehat{\text{EVSI}} = \mathbb{E}_Y \left[\max_d \{\hat{g}(d, T(Y))\} \right] - \max_d \{\mathbb{E}_Y [\{\hat{g}(d, T(Y))\}] \tag{8.15}$$

$$= \frac{1}{N_{PSA}} \sum_{i=1}^{N_{PSA}} \max_d \left\{ \hat{g}^{(i)}\left(d, T(Y)\right) \right\} - \max_d \left\{ \frac{1}{N_{PSA}} \sum_{i=1}^{N_{PSA}} \hat{g}^{(i)}\left(d, T(Y)\right) \right\} . \tag{8.16}$$

8.3.10 Calculating the Cost of Sampling for Fixed Sample Size Design

An important step in choosing the most cost-effective research design is understanding the costs associated with conducting the research, known as the cost of sampling (Eckermann *et al.*, 2009). For fixed sample size designs, the total cost of sampling is composed of fixed, variable

and opportunity costs. The following notation is adopted

TC_{FD} is the total cost of sampling,

C_f is the fixed cost,

C_v is the variable cost per patient incurred by every participant in the trial,

$C_{v,I}$ is the variable cost per patient incurred by participants in the intervention arm only,

$C_{v,I}$ is the variable cost per patient incurred by participants in the control arm only,

C_o is the opportunity cost per patient.

Each cost is estimated based on information available before the trial is conducted.

8.3.10.1 Fixed Costs

Fixed costs (C_f) include any costs incurred in the trial set up, before any participants are randomised, and any costs at the end of the trial incurred, regardless of when the trial finishes. Typical fixed costs include; staff and meeting costs during the study and recruitment, approval from ethics committee, site recruitment and training, archiving costs and dissemination. Fixed costs are independent of the number of participants.

8.3.10.2 Variable Costs

Variable costs include costs associated with identifying, randomising, delivering the intervention to and following-up participants. Additional tasks fundamental to the successful conduct of the trial will continue throughout these periods and include meetings for the trial steering committee, data monitoring committee and public involvement groups. Costs incurred during this time include staff costs and database management. Variable costs depend on

n the expected number of participants in the trial for the given design,

n_I the expected number of participants in the intervention arm,

n_C the expected number of participants in the control arm.

8.3.10.3 Opportunity Costs

The opportunity cost can be thought of as the financial cost of delaying a decision to obtain more information (Zhao *et al.*, 2009). This is an important consideration as during a trial some participants will be randomised to an inferior treatment (Willan, 2008). Additionally, patients outside of the trial will also not be able to receive the superior treatment until the trial ends and it is made available on the NHS.

Willan, 2008 suggest, for a two-arm trial, the opportunity cost is equal to the INB of new intervention compared to the control based on information available before the trial begins. This quantifies the opportunity cost on a monetary scale, which is consistent with the other components of the cost of sampling.

The opportunity cost depends on

n_o the expected number of participants not randomised to the superior intervention,

Assuming an equal allocation ratio this is given by

$$n_o = n\left(\frac{n_{arms}-1}{n_{arms}}\right), \tag{8.17}$$

where

$$\begin{aligned} &n \text{ is the total number of particpiants,} \\ &n_{arms} \text{ is the number of arms in the trial.} \end{aligned} \tag{8.18}$$

8.3.10.4 Total Cost of Sampling for a Fixed Sample Size Design

The cost of sampling for the fixed sample size design can therefore be calculated using

$$TC_{FD} = C_f + nC_v + n_I C_{v,I} + n_C C_{v,C} + n\left(\frac{n_{arms}-1}{n_{arms}}\right) C_o. \tag{8.19}$$

8.3.11 Comparing the Expected Net Benefit of Sampling of Trial Designs

Once the EVSI has been calculated for a fixed sample size design and the cost of sampling estimated it is possible to calculate the ENBS. The ENBS is the difference between the population level EVSI and the cost of sampling (TC_{FD}) for a specific design (Chilcott *et al.*, 2003), given by

$$\text{ENBS} = \text{popEVSI} - TC_{FD}. \tag{8.20}$$

The ENBS can be calculated at the design stage of a clinical trial to compare different research designs, for example, different sample sizes to identify the most cost-effective design. The optimal design will have the highest ENBS. This can then be used in conjunction with other factors that influence the design of a clinical trial, such as the frequentist sample size calculation, known recruitment issues for the target population and available resources to determine the design for the future research.

8.4 Extending the EVSI Approach to Adaptive Clinical Trials

8.4.1 Overview

In this section, the methods of EVSI are extended to guide the design of group sequential clinical trials, highlighting the additional considerations to ensure the adaptive characteristics of the trial are fully captured. The aim is to develop this approach so that researchers can determine the cost-effective design for a trial comparing designs with different stopping rules and number of interim analyses and to compare adaptive designs with fixed sample size designs.

Figure 8.1 provides an overview of the approach I have developed. The same steps proposed by Fenwick *et al.*, 2020 and described in Section 8.3 are followed for adaptive designs, however, additional considerations are required in Step 6: *Estimate the value of specific research*. The first is the requirement to adjust the point estimates and confidence intervals of primary and secondary endpoints to allow for the adaptive nature of the trial design, as discussed in Chapter 6. The second is ensuring the cost of sampling accounts for the additional costs and any potential costs savings from an adaptive design.

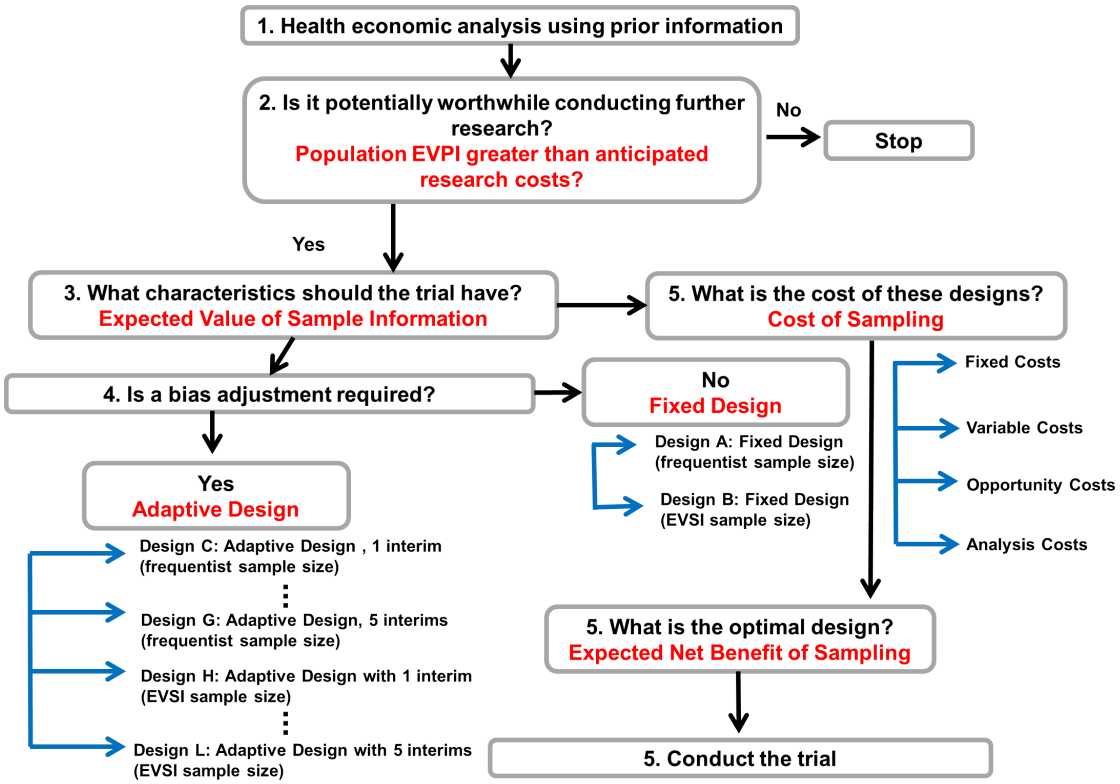


FIGURE 8.1: Diagram of how the value of information analysis approach can be extended to guide the design of adaptive clinical trials. EVPI; expected value of perfect information

The steps required for calculating the EVSI for an adaptive design are summarised in Figure 8.2, building on the steps discussed for the fixed sample size design in Section 8.3. The following sections discuss each of these steps in more detail, highlighting how they differ from the fixed sample size design case.

8.4.2 Identifying the Trial Designs for Comparison

The first step in applying the EVSI approach to guide the design of adaptive clinical trial is to choose the trial designs under consideration. As with the fixed sample size design this will require choosing an estimate of the clinically important difference for the primary outcome, an estimate of the population variance and type I and type II errors (Flight *et al.*, 2016). These choices are the same regardless of the adaptive nature of the trial and are usually informed by prior information or discussions with the clinical research team.

When deciding whether to use an adaptive design additional considerations include the choice of stopping rule and the number of interim analyses. In this chapter, a group sequential design is considered comparing the Pocock stopping rule and the O'Brien-Fleming stopping rule (each

FIGURE 8.2: Summary of the steps taken to calculate the expected net benefit of sampling for an adaptive design

1. Identifying the trial designs for comparison including:
   (a) determine the clinically important difference for the primary outcome
   (b) estimate of the population variance for the primary outcome
   (c) type I and type II error rates
   (d) stopping rule and number of interim analyses
   (e) sample size
2. Develop a health economic model for the population of interest (or use an existing model).
3. Generate a probabilistic sensitivity analysis (PSA) sample using prior evidence for parameters (θ)

$$\{\theta^{(1)}, \ldots, \theta^{(N_{PSA})}\}. \tag{8.21}$$

4. For each row of the PSA sample evaluate the health economic model calculating a per person expected net benefit for each treatment arm

$$\left\{NB(d, \theta^{(1)}), \ldots, NB(d, \theta^{(N_{PSA})})\right\}. \tag{8.22}$$

5. For each row of the PSA sample simulate a trial result dataset, using the likelihood function for the trial data

$$\left\{x^{(1)}, \ldots, x^{(N_{PSA})}\right\}. \tag{8.23}$$

6. Apply the trial analysis for the group sequential trial design chosen in step 1 to each trial result dataset. This establishes whether the trial would have stopped early at any of the interim analyses and gives a trial analysis dataset

$$\left\{y^{(1)}, \ldots, y^{(N_{PSA})}\right\}. \tag{8.24}$$

7. Summary statistics for primary and secondary outcomes informing the health economic model are estimated from the trial analysis dataset in each row of PSA sample. These statistics are adjusted for bias to account for the adaptive nature of the trial using the BAMLE method,

$$\left\{\tilde{T}\left(y^{(1)}\right), \ldots, \tilde{T}\left(y^{(N_{PSA})}\right)\right\}. \tag{8.25}$$

8. Using a generalised additive model (Strong *et al.*, 2015) regress the net benefits from step 4 with the estimated health economics model parameters in step 7 to calculate adjusted ($\widetilde{\text{EVSI}}$).
9. The population EVSI is calculated by multiplying the per patient EVSI by the time horizon and prevalence.
10. The ENBS is the costs of the proposed design minus the population EVSI,

$$\widetilde{\text{ENBS}} = \widetilde{\text{popEVSI}} - TC_{AD} \tag{8.26}$$

11. This process is repeated for another possible trial design option by returning to Step 1 for as many different GSDs or fixed design designs under consideration as the analyst wishes.

described in Section 2.4) with up to five equally spaced analyses of the data. The sample size is then informed by these choices.

8.4.3 Generating a Probabilistic Sensitivity Analysis Sample

As for the fixed sample size design approach, described in Section 8.3, a health economic model is required for the population of interest. Once the design under consideration has been selected it is necessary to create a PSA sample based on available prior evidence for parameters θ. For each row of the PSA sample the health economic model is evaluated to give a per person expected net benefit for each intervention under consideration.

8.4.4 Simulating Trial Result and Analysis Datasets

As in the fixed sample size design described in Section 8.3.7, a trial result dataset is simulated for each row of the PSA sample. Rothery *et al.*, 2020 suggest that data sets should be simulated taking into account how the data from the trial would be analysed and account for any potential bias in the data. In Chapter 7 it was shown that the adaptive nature of the trial could impact the health economic analysis following the trial. It is important, therefore, to consider this bias in the EVSI calculation for adaptive designs. Failing to adjust for this bias could result in a spurious estimate of the EVSI and the wrong design choice being made, potentially wasting limited resources.

The data simulation is informed by the PSA parameter estimates, to give a dataset in each row. The trial analysis for the trial design under consideration is applied to each trial result dataset. Where an adaptive design is considered, this establishes whether the trial would have stopped early at any of the interim analyses and gives a trial analysis dataset. For example, if a trial with a Pocock stopping rule with five equally spaced analyses is being evaluated, the first group of simulated patient in $x^{(1)}$ form the analysis set for the first interim analysis. The primary outcome is calculated and compared to the stopping boundary for the rule. If the estimate crosses the boundary, the trial stops and the trial results dataset is formed from the patients randomised into the trial up to that point. If the boundary is not crossed, the trial continues to the next interim analysis until the trial crosses the boundary or the final analysis is reached. This is repeated for each PSA sample.

8.4.5 Calculating Summary Statistics

As in Section 8.3.8, summary statistics for primary and secondary outcomes informing the health economic model are estimated from the trial analysis dataset in each row of the PSA sample. The Bias Adjusted Maximum Likelihood Estimate (BAMLE) (described in Section 2.5) is used to give adjusted summary statistics denoted by

$$\left\{\tilde{T}(y^{(1)}), \ldots, \tilde{T}\left(y^{(N_{PSA})}\right)\right\}. \tag{8.27}$$

Unadjusted summary statistics using the MLE are calculated to allow a comparison of adjusted and unadjusted approaches. These statistics are denoted by

$$\left\{\hat{T}(y^{(1)}), \ldots, \hat{T}\left(y^{(N_{PSA})}\right)\right\}. \tag{8.28}$$

8.4.6 Calculating the Expected Value of Sample Information

Using a generalised additive model (Strong *et al.*, 2015) the net benefits in the PSA sample are regressed with the estimated summary statistics to calculate an adjusted ($\widetilde{\text{EVSI}}$) and unadjusted ($\widehat{\text{EVSI}}$) EVSI. The population EVSI is calculated by multiplying the per patient EVSI by the time horizon and prevalence.

8.4.7 Calculating the Cost of Sampling for Adaptive Designs

The next step is to estimate the cost of sampling for the design. In addition to costs for the fixed sample design outlined in Section 8.3.10, the costs associated with conducting an interim analysis are required (Willan, 2008), where

C_a is the cost of analysis.

For the fixed design, the cost of analysis is included in the fixed costs. For the adaptive design, however, this is now separated as multiple analyses may take place depending on the design chosen.

The analysis costs should include any costs associated with conducting an analysis of the endpoints used to inform interim decision making (typically clinical endpoints). Analysis costs will include data validation, cleaning and statistical analyses each of which incurs a staff cost. The expected analysis costs depend on

$$N_a \text{ the expected number of analyses.}$$

The cost of sampling equation now becomes

$$TC_{AD} = C_f + N_a C_a + nC_v + n_I C_{v,I} + n_c C_{v,C} + n\left(\frac{n_{arms}-1}{n_{arms}}\right) C_o. \tag{8.29}$$

8.4.7.1 Practical Steps for Calculating the Cost of Sampling for an Adaptive Design

The calculation of the cost of sampling can be more challenging than for the fixed sample size design as the progress of the trial is less predictable, and there is currently limited guidance on how to do this. I propose the following approach which is then applied to the CACTUS case study in Section 8.7 and summarised in Figure 8.3. This approach is based on the maximum required sample size for the chosen design and the proposed recruitment rate for this design.

Firstly, the study period can be considered in four distinct phases and allocated a specific number of months:

1. Set-up - period before first participant is recruited (e.g. 6-months)
2. Recruitment – period where participants are recruited into the trial (e.g. 12-months)
3. Follow-up – period after last participant recruited but follow-up data collection on-going (e.g. 12-months)
4. Trial end – period once all data collected, then analysed and disseminated (e.g. 6-months)

Each cost incurred during the study can be broken down into a monthly cost and allocated to a month in the given phase of the trial. Some costs, such as dissemination costs, may fall into a single month in a single study phase (trial end). Other costs may cover multiple phases, for

example, site training might take place predominantly during the trial set-up period but some sites might be trained during the trial recruitment period if the trial continues through interim analyses.

Using the proposed monthly recruitment rate, costs per participant can be converted to a monthly cost. For example, if the anticipated recruitment rate is one participant per centre per month, in a month where six centres are expected to be recruiting the monthly recruitment rate will be six participants. Costs associated with randomising a participant can be multiplied by six to give the monthly cost. All costs are given as a monthly cost so they can be easily combined.

The fixed costs (C_f) are the sum of the monthly costs in the trial set-up and trial-end phases. The analysis cost (C_a) is the sum of the analysis costs associated with the analysis of the primary endpoint used for interim decision making, as discussed in Section 8.4.7.

To calculate the variable costs incurred by all participants (C_v) the costs incurred by each trial participant each month during the trial recruitment and follow-up periods are added together. This is then added to all other costs incurred during these phases such as staff and meeting costs that continue throughout this period. The sum of these monthly costs is then divided by the maximum number participants to be recruited in the trial to give a cost per participant in each month.

To calculate the variable costs incurred by participants in the intervention arm only ($C_{v,I}$) a similar process is followed. The costs incurred by each trial participant in the intervention arm in each month during the trial recruitment and follow-up periods are added together. These costs are then divided by the maximum number participants to be recruited to the intervention arm. This gives a cost per participant in each month. The costs in the recruitment and follow-up periods are added together to give a variable cost for each participant in the intervention arm during the trial. This is repeated to calculate the variable costs incurred by participants in the control arm ($C_{v,C}$) including the costs in the recruitment and follow-up phases that are incurred by participants receiving the control arm treatment.

FIGURE 8.3: Practical steps for calculating the cost of sampling for a clinical trial with an adaptive design

1. Divide the study into four phases:
   - Set-up
   - Recruitment
   - Follow-up
   - Trial end

2. Calculate the monthly cost for each cost incurred and allocate to a month in a given phase.

3. Calculate the fixed cost as the sum of monthly costs in the set-up and trial end phases.

4. Calculate the analysis costs as the cost of conducting the primary outcome analysis.

5. Calculate the variable cost per participant as the sum of all costs incurred by all participants in each month of the recruitment and follow-up periods divided by the number of participants to be recruited.

6. Calculate the variable cost per participant in the intervention arm as the sum of costs incurred in the intervention arm in each month of the recruitment and follow-up periods divided by the number of participants to be recruited.

7. Repeat step 5 for each arm of the trial.

8. Calculate the cost of sampling using,

$$TC_{AD} = C_f + N_a C_a + nC_v + n_I C_{v,I} + n_c C_{v,C} + n\left(\frac{n_{arms} - 1}{n_{arms}}\right) C_o.$$

8.4.8 Comparing the Expected Net Benefit of Trial Designs

The ENBS is the population EVSI minus the costs of the proposed design . This can be calculated using the adjusted approach, denoted by $\widetilde{\text{ENBS}}$ or an unadjusted approach denoted by $\widehat{\text{ENBS}}$,

$$\widetilde{\text{ENBS}} = \text{pop}\widetilde{\text{EVSI}} - TC_{AD} \tag{8.30}$$

$$\widehat{\text{ENBS}} = \text{pop}\widehat{\text{EVSI}} - TC_{AD} \tag{8.31}$$

This process is repeated to give the ENBS for a range of designs under consideration. The optimal design from a health economic perspective has the highest ENBS. This information can be used to guide the design of a clinical trial alongside discussions with clinical teams, including the use of adaptive clinical trials as well as traditional fixed sample size designs.

8.5 Comparing EVSI for Five Possible Designs for the CACTUS Case Study

Section 8.4 gave a general approach for extending existing EVSI methods to guide the design of adaptive clinical trials. The following section applies this approach to the CACTUS case study (described in Chapter 5) to illustrate how the approach can be operationalised in a real-world example and to compare adjusted and unadjusted results that allow for the adaptive nature of the trial.

For context, it is assumed that the CACTUS pilot trial has ended. Based on the pilot results, the decision has been made to conduct a full scale randomised controlled trial. The proposed full scale trial will be a two arm trial comparing the computer-based intervention and usual care control, as described in Chapter 6. The primary outcome will be the improvement in percentage words named correctly from baseline to 6-months. Recruitment is planned to take place over an 18-month period. The approach outlined in Figure 8.2 is used to guide the design of the hypothetical future trial.

Rule	Number of Analyses	Correlation between Primary and Secondary Outcomes	Prior Uncertainty from Pilot Trial	Intervention Costs
Fixed sample size	1	(0.0, 0.4, 0.8)	High, Low	£769.25 to £11,538.75
O'Brien Fleming	2	(0.0, 0.4, 0.8)	High, Low	£769.25 to £11,538.75
O'Brien Fleming	5	(0.0, 0.4, 0.8)	High, Low	£769.25 to £11,538.75
Pocock	2	(0.0, 0.4, 0.8)	High, Low	£769.25 to £11,538.75
Pocock	5	(0.0, 0.4, 0.8)	High, Low	£769.25 to £11,538.75

TABLE 8.2: Summary of the 90 scenarios considered for each of the five designs for the hypothetical trial following the CACTUS pilot trial.

8.6 Methods

8.6.1 Trial Design and Data Characteristics

The five designs in Table 8.2 are assessed. The Pocock and O'Brien-Fleming stopping rules have differing characteristics (as discussed in Chapter 2) and so provide an interesting comparison between available adaptive designs. Pocock, 1983 suggest there is little statistical benefit to conducting more than five interim analyses and so two and five analysis are considered. This reflects the five design options explored in the simulation studies of Chapter 7 (FIX, OBF2, OBF5, POC2 and POC5). Each design is applied in `R` using the `RCTdesign` package. To explore the impact of different characteristics on the results, 90 scenarios are considered for each design varying the correlation between primary and health economic outcomes, the level of prior uncertainty from the pilot trial and the intervention costs.

A range of correlation values is explored to assess their impact on the analysis. It has been assumed there is a negative correlation between the primary outcome (improvement in percentage words named correctly from baseline to 6-months) and costs; as a participant's word naming ability improves, they incur fewer costs to the NHS for their on-going care. A positive correlation has been assumed for the primary outcome and utilities, as a participant's word naming ability improves their quality of life improves. Absolute correlations of $0.0, 0.4, 0.8$ are explored covering a range of no, medium and high correlation for this case study.

It is anticipated that the level of prior uncertainty when designing the future trial will influence the optimal design. When there is less prior uncertainty the expected value of conducting further research is reduced. As such a cheaper trial design is likely to be advantageous. To explore this further, the CACTUS pilot data is replicated twice, to give a larger sample size and hence reduced prior uncertainty, whilst maintaining the same mean and standard deviations for the data to allow comparisons with the CACTUS pilot analysis.

The sensitivity of the results to the cost of the intervention is explored by varying the intervention cost in the estimate of the INB from the pilot trial and the subsequent EVSI and ENBS calculations. The costs are varied over 15 values and summarised graphically to identify the optimal design.

8.6.2 Data Generating Mechanism

To generate the PSA sample the bootstrapping methods described in Section 2.7.8 are used. The CACTUS pilot data are bootstrapped and 6,000 samples are drawn giving $N_{PSA} = 6,000$.

To simulate a trial result dataset (step 4 of Figure 8.2), data are simulated using the approach outlined in Simulation Study Two in Chapter 7. This gives a trial dataset for each of the PSA rows. Parameters for the data simulation are based on the bootstrapped CACTUS pilot data forming PSA samples. A willingness to pay threshold of £20,000 per QALY is used throughout as per NICE guidance with a discount rate of 3.5% applied to costs and benefits (National Institute for Health and Care Excellence, 2013b). The time horizon and prevalent population are taken from the Latimer *et al.*, 2013 health economic analysis, described in Chapter 5, giving an average of 27,616 patients expected to eligible for and compliant with the computer-based intervention per year over a ten year period.

8.6.3 Trial Results Estimates

Adjusted and unadjusted estimates of the health economic model parameters are calculated using the steps outlined in Chapter 6. Adjusted and unadjusted estimates of EVSI are then calculated using the steps of Section 8.4.5. This utilises `R` code provided by Prof Mark Strong as part of the SAVI platform to implement the generalised additive model (Strong *et al.*, 2020). Only unadjusted estimates are presented for the fixed sample size design.

The percentage differences between $\widehat{\text{EVSI}}$ and $\widetilde{\text{EVSI}}$ are calculated and compared. Adjusted and unadjusted point estimates of the primary clinical outcome are reported with the width of the 95% confidence interval. This summarises how each design helps to reduce uncertainty in the clinical outcome. The ENBS is then calculated for each of the scenarios considered using Equation 8.20 and compared to determine the optimal trial design from a health economic perspective.

8.7 Results

In this section the steps outlined in Section 8.4.7 are applied to the CACTUS case study to estimate the components of the cost of sampling in Equation 8.29. The EVSI approach extended to consider adaptive designs is then used to design a hypothetical future trial based on the CACTUS pilot data and the CACTUS pilot data with reduced uncertainty. Each design is considered with correlation of 0.0, 0.4, 0.8 between the primary and health economic outcomes. Sensitivity analyses then explore the impact of varying the intervention costs on these results.

For the CACTUS pilot data 5,996 of the 6,000 PSA samples converged. The four of the bootstrap samples did not result in values of the primary outcome that could be converted to the parameters of the Beta distribution for the data simulation. These values were likely to be in the extreme tails of the distribution. As this was only 0.067% of the results it is not felt that this will greatly impact the results.

8.7.1 Sample Sizes

A frequentist sample size for a fixed sample size design trial, using the formula for a superiority trial (Flight *et al.*, 2016) is calculated based on the CACTUS pilot data. It is assumed that the type I and type II error rates are 0.05 and 0.1 respectively. The clinical important difference in the improvement in proportion of words named correctly between the intervention and control arm taken from the pilot trial data is 0.127 with an estimated standard deviation of 0.3338. This gives an overall sample size of 292 participants assuming an equal allocation between arms. All group sequential designs are based on these sample size parameters.

Based on the CACTUS pilot data with reduced prior uncertainty the same parameters are used. The estimated standard deviation is 0.3274 because of the differences in the pooled estimate of

the standard deviation in the two datasets. This gives an overall sample size of 280 participants assuming an equal allocation between arms.

8.7.1.1 Maximum and Average Sample Size and Proportion of Trials Stopping at each Analysis

Table 8.3 summarises the expected sample size, number of analyses and distribution of the sample size for each design considered over the interim and final analyses.

The behaviour of the expected number of analyses is similar for the CACTUS pilot and the scenario with reduced prior uncertainty. The OBF design with five analyses has the highest expected number of analyses (3.60 for the CACTUS pilot). The POC design with two analyses has the fewest expected number of analyses (1.43 for the CACTUS pilot). The FIX design has one analysis as there is no opportunity to examine the data during the trial.

POC applies a large penalty for early examination of the data; the design with five interim analyses has the highest maximum sample size, 352 for the pilot data analysis and 338 for the reduced prior uncertainty analysis. The penalty is much smaller for OBF. The design with two analyses has a maximum sample size close to FIX.

Although POC designs have a high maximum sample size they have small expected sample sizes compared to FIX as only a small number of trials reach the final analyses. The OBF designs also have a smaller expected sample size than the FIX, as FIX has no opportunity to stop the trial early.

The POC designs have a smaller proportion of trials reaching the final analysis when compared to OBF. The OBF designs make it difficult to stop early-on in the trial, however the POC designs require the same level of difficulty regardless of when the interim analysis is conducted. Only a small number of trials stopped at the first interim analysis with OBF with five interim analyses, where the total number of participants was 60 and 58 for the pilot analysis and the reduced prior uncertainty analysis respectively. In contrast, almost 30% of trials stopped at the first analysis of POC with five analyses where the total sample size was 72 and 68 respectively.

		CACTUS Pilot Data				
design		**FIX**	**OBF 2**	**OBF 5**	**POC 2**	**POC 5**
Correlation				0.0		
Expected number of analyses		1.00	1.59	3.60	1.43	2.90
Maximum sample size		292	294	300	320	352
Expected sample size		292.00	233.87	215.83	228.30	205.34
Proportion stopping at analysis (number of participants at each analysis)	1	1.00 (292)	0.41 (148)	0.01 (60)	0.57 (160)	0.29 (72)
	2	-	0.59 (294)	0.24 (120)	0.43 (320)	0.20 (142)
	3	-	-	0.24 (180)	-	0.11 (212)
	4	-	-	0.15 (240)	-	0.07 (282)
	5	-	-	0.36 (300)	-	0.32 (352)
Cost sampling (£million)		2.13	1.84	1.75	1.81	1.70
Cost of Sampling for a trial stopping at analysis (£million)	1	2.13	1.42	0.98	1.48	1.04
	2	-	2.14	1.28	2.27	1.39
	3	-	-	1.57		1.73
	4	-	-	1.87	-	2.08
	5	-	-	2.17	-	2.42
		Reduced Prior Uncertainty Data				
Expected number of analyses		1.00	1.63	3.64	1.43	2.93
Maximum sample size		280	282	288	208	338
Expected sample size		280.00	230.48	209.93	220.22	198.44
Proportion stopping at analysis (number of participants at each analysis)	1	1.00 (280)	0.37 (142)	0.01 (58)	0.57 (154)	0.25 (68)
	2	-	0.63 (282)	0.20 (116)	0.43 (308)	0.24 (136)
	3	-	-	0.29 (174)	-	0.14 (204)
	4	-	-	0.17 (230)	-	0.08 (272)
	5	-	-	0.34 (288)	-	0.29 (338)
Cost sampling (£million)		2.07	1.82	1.72	1.77	1.67
Cost of Sampling for a trial stopping at analysis (£million)	1	2.07	1.39	0.97	1.45	1.02
	2	-	2.08	1.26	1.71	1.36
	3	-	-	1.54	-	1.69
	4	-	-	1.82	-	2.03
	5	-	-	2.11	-	2.36

TABLE 8.3: Simulated trial results for expected number of analyses, expected sample size, proportion of trials stopping at each analysis and expected cost of sampling, assuming zero correlation between the primary and health economic outcomes, for pilot and reduced uncertainty based on 5,996 possible trial results for the pilot and 6,000 possible trial results for the pilot with reduced prior uncertainty

8.7.2 Calculating the Cost of Sampling for the CACTUS Trial

Financial information from the Big CACTUS grant application was used to inform the cost of sampling for the hypothetical CACTUS trial. The approach outlined in Section 8.4.7 was applied. In contrast to the original Big CACTUS trial, it was assumed that all costs associated with conducting the trial were of interest regardless of who incurred these costs. This meant that research costs were considered at 100% of their value, although in practice (and in the Big CACTUS trial) funders such as the NIHR might provide 80% of these costs. Research support costs were also included.

It was assumed that the costs of usual care provided in the control arm and the intervention arm were zero. This is because they are not a cost of the trial and would be provided to a participant whether or not they were taking part in the study. The cost of the computer-based intervention was considered a trial cost and was included in the cost of sampling. The intervention cost of £769.25 was taken from the CACTUS pilot trial. These choices are discussed in more detail in Section 8.8.

The cost of sampling for each design considered is given in Table 8.3. The cost of sampling for the fixed sample size design is £2,068,177.56. The components of the cost of sampling are given in Table 8.4. The fixed costs are £682,414.83 and account for 33% of the cost of sampling. These costs are incurred regardless of the design chosen to conduct the trial. These costs differ from the original Big CACTUS trial as they are based on two-arm scenarios, consider all costs associated with conducting the trial and are estimated based on the Big CACTUS grant application not the actual spend of the Big CACTUS trial.

The FIX has the highest cost of sampling of the five designs considered. The adaptive designs with two analyses have similar costs of sampling regardless of the stopping rule used, as do the adaptive designs with five analyses. POC with five analyses has the smallest cost of sampling reflecting the smaller expected number of analyses and smaller expected sample size.

The expected cost of sampling for trials stopping at each analysis for each of the designs are given in Table 8.3. This shows that even when a trial can stop at the first analysis large costs are incurred, especially when the first analysis is conducted halfway through the trial (when there are two planned analyses). The trials stopping at the first analysis of five have the smallest

	Description	Price
Fixed	All costs incurred which do not depend on the number of participants and the design of the trial including trial set up costs, dissemination costs and secondary analysis costs	£682,414.83
Variable	Research costs for all study participants including the intervention costs, data collection costs, recruitment costs	£3,371.19 per participant
	Costs of delivering the computer-based intervention to the intervention arm participants	£769.25 per participant
	Costs of delivering usual care to the control arm participants	£0.00 per participant
Analysis	The cost of one month of a junior statistician's time to conduct the analysis required at an interim analysis of the primary, clinical endpoint	£874.33 per analysis
Opportunity	The incremental net benefit taken from R model analysis of the CACTUS pilot from Chapter 5	£2,380.44 per participant

TABLE 8.4: Summary of cost of sampling for simulated trial scenarios based on the Big CACTUS trial grant application

cost of sampling as they have one fifth of the maximum number of participants. This is slightly smaller for OBF as the overall maximum sample size is smaller so the first analysis is conducted on the fewest number of participants when compared to the other designs.

8.7.3 Expected Value of Sample Information and Expected Net Benefit of Sampling

FIX has the highest EVSI for all correlations considered in Table 8.5 for the pilot data analysis and the pilot data analysis with reduced prior uncertainty in Table 8.6. Consequently, the fixed sample size design has the highest ENBS, despite having the highest cost of sampling. From a cost-effectiveness perspective is the optimal trial design for the future clinical trial in each case.

The EVSI is higher for the adaptive designs with two analyses compared to five analyses, perhaps reflecting the larger expected sample size for these designs in Table 8.3. The EVSI is smallest for POC with five analyses for both the pilot data analysis and the reduced prior uncertainty scenario for all correlation values explored.

The EVSI is much less for the pilot with reduced prior uncertainty compared to the pilot analysis. This reflects less uncertainty before the start of the trial and hence less information to be gained by conducting further research. The ENBS for the reduced uncertainty scenario is less

	Unadjusted					Adjusted				
Design	**FIX**	**OBF 2**	**OBF 5**	**POC 2**	**POC 5**	**FIX**	**OBF 2**	**OBF 5**	**POC 2**	**POC 5**
Correlation	0.0									
EVSI	32.91	23.81	22.26	24.02	17.97	-	26.64	21.61	25.62	17.99
EVSI (SE)	5.64	4.67	4.47	4.69	3.92	-	5.05	4.43	4.79	4.04
Pop EVSI (millions)	9.09	6.58	6.15	6.63	4.96	-	7.36	5.97	7.07	4.97
ENBS (millions)	6.96	4.74	4.39	4.82	3.26	-	5.52	4.21	5.26	3.27
Width primary CI	0.15	0.18	0.19	0.18	0.21	-	0.18	0.19	0.18	0.22
Correlation	0.4									
EVSI	35.57	28.44	27.51	28.40	18.73	-	31.32	26.75	30.37	20.59
EVSI (SE)	5.95	5.19	5.01	5.19	4.11	-	5.29	4.78	5.42	4.14
Pop EVSI (millions)	9.82	7.85	7.60	7.84	5.17	-	8.65	7.39	8.39	5.69
ENBS (millions)	7.70	6.00	5.84	6.01	3.47	-	6.80	5.63	6.56	3.98
Width primary CI	0.15	0.18	0.19	0.18	0.21	-	0.18	0.19	0.18	0.22
Correlation	0.8									
EVSI	36.95	26.82	26.23	26.69	21.17	-	27.96	24.67	29.09	22.52
EVSI (SE)	5.74	4.90	5.17	5.27	4.28	-	4.99	4.87	5.02	4.71
Pop EVSI (millions)	10.20	7.41	7.24	7.37	5.85	-	7.72	6.81	8.03	6.22
ENBS (millions)	8.08	5.57	5.51	5.56	4.16	-	5.88	5.08	6.22	4.53
Width primary CI	0.15	0.18	0.19	0.18	0.21	-	0.18	0.19	0.18	0.22

TABLE 8.5: EVSI results for five different proposed trial designs comparing adjusted and unadjusted under three different scenarios for the extent of correlation between primary and health economic outcomes (a) uncorrelated, with correlation 0.0, 0.4, 0.8. EVSI estimated using 5,996 PSA samples

than zero in some scenarios suggesting it is not worthwhile conducting the further research from a health economic perspective. When the correlation is 0.4 the adjusted and unadjusted ENBS have a different sign for the OBF designs and the POC design with two analyses. This suggests different conclusions would be drawn about the cost-effectiveness of conducting this research.

The width of the 95% confidence interval for the primary outcome is higher for all adaptive designs compared to FIX and widest for POC with five analyses. This indicates greater uncertainty in the trial results when an adaptive design is used, mirroring the behaviour of the EVSI and the results in Chapter 7. The intervals are wider for adaptive designs with five analyses compared to two analyses, a likely consequence of the smaller expected sample size for these designs.

Figure 8.4a shows the point estimate for the EVSI for the pilot data with its 95% confidence interval and the equivalent plot for the pilot data with reduced prior uncertainty in Figure

	Unadjusted					Adjusted				
Design	**FIX**	**OBF 2**	**OBF 5**	**POC 2**	**POC 5**	**FIX**	**OBF 2**	**OBF 5**	**POC 2**	**POC 5**
Correlation	0.0									
EVSI	6.89	4.64	3.52	4.29	3.46	-	5.07	3.85	4.87	4.32
EVSI (SE)	1.79	1.36	1.10	1.26	0.97	-	1.50	1.23	1.48	1.16
Pop EVSI (millions)	1.90	1.28	0.97	1.18	0.96	-	1.40	1.06	1.34	1.19
ENBS (millions)	-0.17	-0.54	-0.75	-0.59	-0.71	-	-0.42	-0.66	-0.43	-0.47
Width primary CI	0.16	0.18	0.19	0.19	0.21	-	0.18	0.19	0.18	0.22
Correlation	0.4									
EVSI	9.07	5.29	4.36	5.28	2.77	-	6.97	6.29	6.61	4.20
EVSI (SE)	2.19	1.55	1.18	1.41	0.89	-	1.60	1.50	1.57	1.05
Pop EVSI (millions)	2.50	1.46	1.20	1.46	0.76	-	1.92	1.74	1.83	1.16
ENBS (millions)	0.44	-0.35	-0.52	-0.31	-0.89	-	0.11	0.01	0.06	-0.50
Width primary CI	0.16	0.18	0.19	0.19	0.22	-	0.18	0.19	0.18	0.22
Correlation	0.8									
EVSI	7.90	5.32	4.78	4.39	2.18	-	6.65	5.15	6.13	2.56
EVSI (SE)	1.84	1.47	1.31	1.36	0.78	-	1.54	1.28	1.53	0.94
Pop EVSI (millions)	2.18	1.47	1.32	1.21	0.60	-	1.84	1.42	1.69	0.71
ENBS (millions)	0.11	-0.35	-0.40	-0.56	-1.05	-	0.02	-0.29	-0.08	-0.94
Width primary CI	0.16	0.18	0.19	0.19	0.21	-	0.18	0.19	0.18	0.22

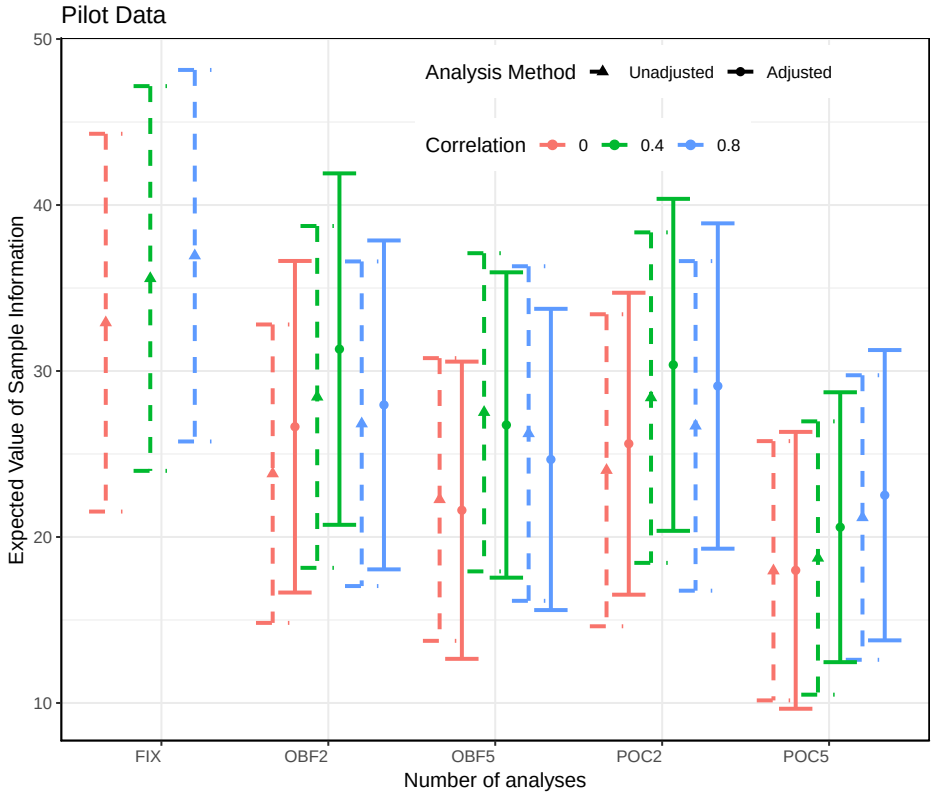
TABLE 8.6: EVSI results for five different proposed trial designs comparing adjusted and unadjusted under three different scenarios for the extent of correlation between primary and health economic outcomes with correlation 0.0, 0.4, 0.8 and with reduced uncertainty. EVSI estimated using 6,000 PSA sample

8.4b. The 95% intervals are wide and overlapping for all scenarios considered, suggesting the choice of design for the trial in uncertain if considering the EVSI alone.

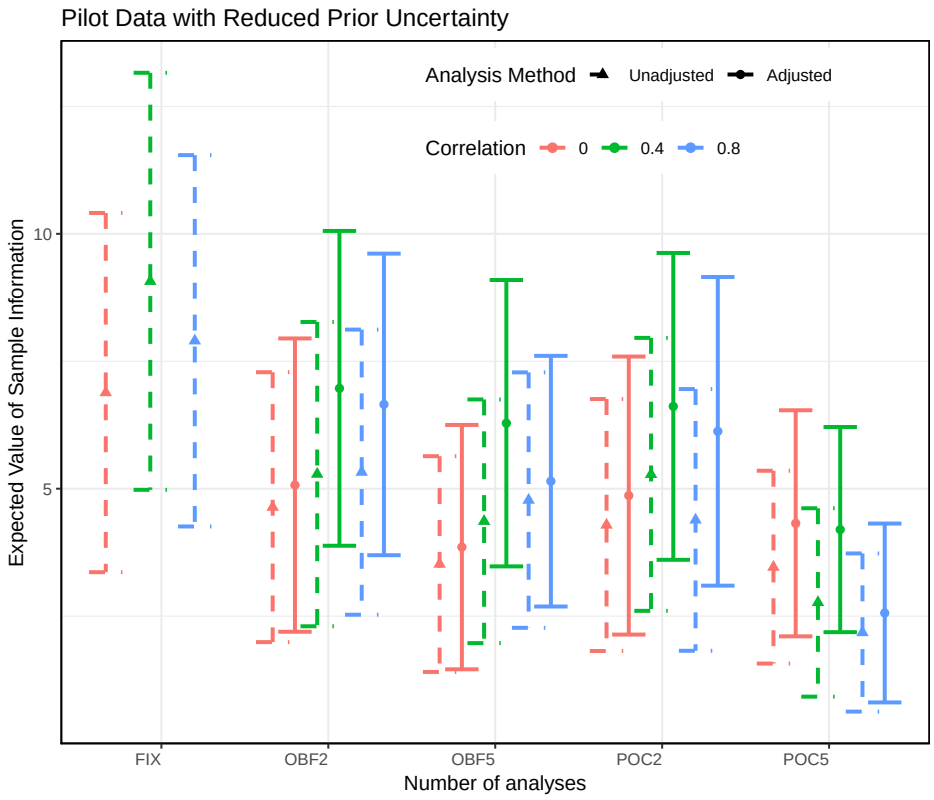
8.7.4 Varying Intervention Costs in the CACTUS Case Study

To explore the sensitivity of the results to the cost of the intervention, the computer-based intervention in the CACTUS case study, the cost (£769.25) is multiplied by a factor of two to 15 and the EVSI and cost of sampling calculations repeated.

Figure 8.5a illustrates the designs under consideration for each intervention costs ranked by ENBS for the pilot data. The unadjusted ENBS is used for FIX and the adjusted ENBS for the adaptive designs. The first point gives the results summarised in Table 8.5 for an intervention cost of £769.25. It is clear that the adaptive designs with two analyses perform better than the five analyses designs when the intervention costs are low, as they have a higher EVSI that outweighs the increased cost of sampling for these designs. However, once the intervention costs are higher than approximately £8,000 the advantage of the higher EVSI is outweighed by



(A) Pilot Data



(B) Pilot with Reduced Prior Uncertainty

FIGURE 8.4: Adjusted and unadjusted expected value of sample information and its 95% confidence interval for five designs given the pilot data (5,996 PSA samples) and the pilot data with reduced prior uncertainty (6,000 PSA samples) for an intervention cost of £769.25 and correlation values 0.0, 0.4 and 0.8 between the primary and health economic outcomes. FIX; fixed sample size design, POC; Pocock stopping rule, OBF; O'Brien-Fleming stopping rule

the increased cost of sampling due to their higher expected sample sizes. Similar results are observed across the correlations.

FIX has the highest ENBS until the intervention costs are eight to ten times higher than in the original CACTUS pilot (£769.25). Once the intervention costs reach these higher levels, the adaptive designs become the optimal design reflecting the smaller expected sample size and hence lower cost of sampling. It is noted in Figure 8.6a that the adaptive design becomes optimal once the ENBS of sampling is close to zero, and from a health economic perspective it is not worthwhile conducting further research.

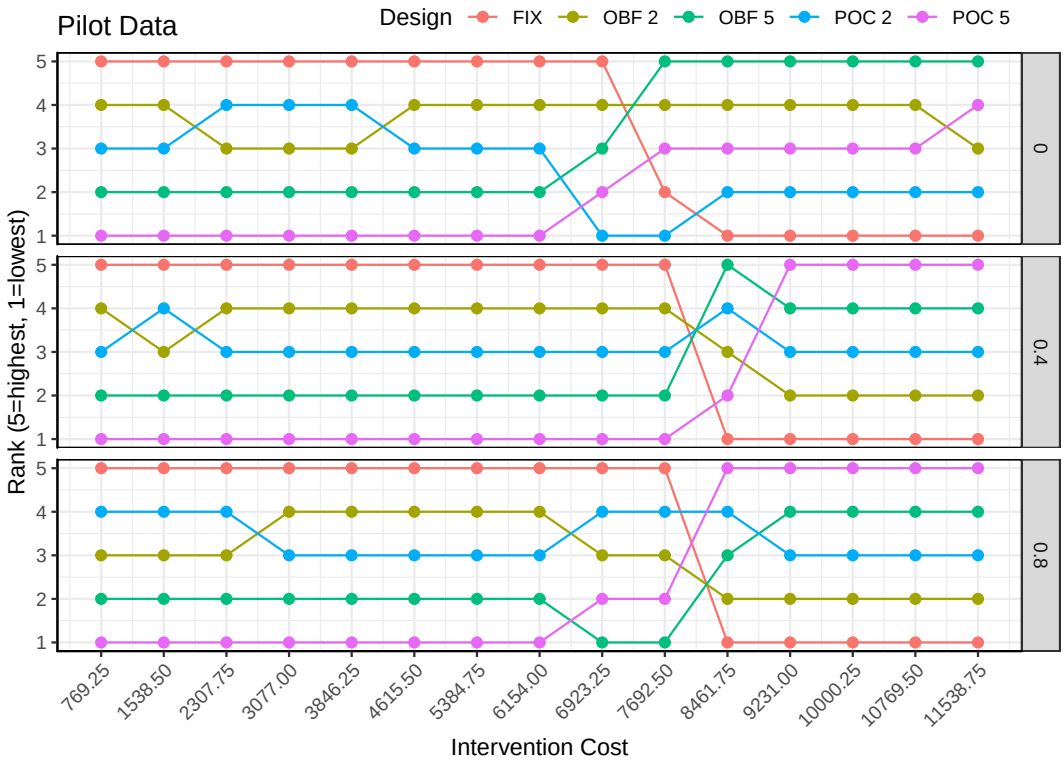
There is an increase in the ENBS as the intervention cost increases as the uncertainty in the cost-effectiveness decision increases. Once the intervention costs reach £3,846.25 this uncertainty begins to decrease as it becomes clearer that the computer-based intervention is not going to be cost-effective when compared to usual care.

For the scenario where there is reduced prior uncertainty there is a much quicker decline to an ENBS below zero, as illustrated in Figure 8.6b. This reflects the much smaller EVSI for the design. FIX is the optimal design for intervention costs up to approximately £4,646, when the correlation is 0.4, after which the adaptive designs perform better and the ENBS is below zero. POC with five analyses is optimal when intervention costs are high and as illustrated in Figure 8.5b Similar results are observed across the correlations.

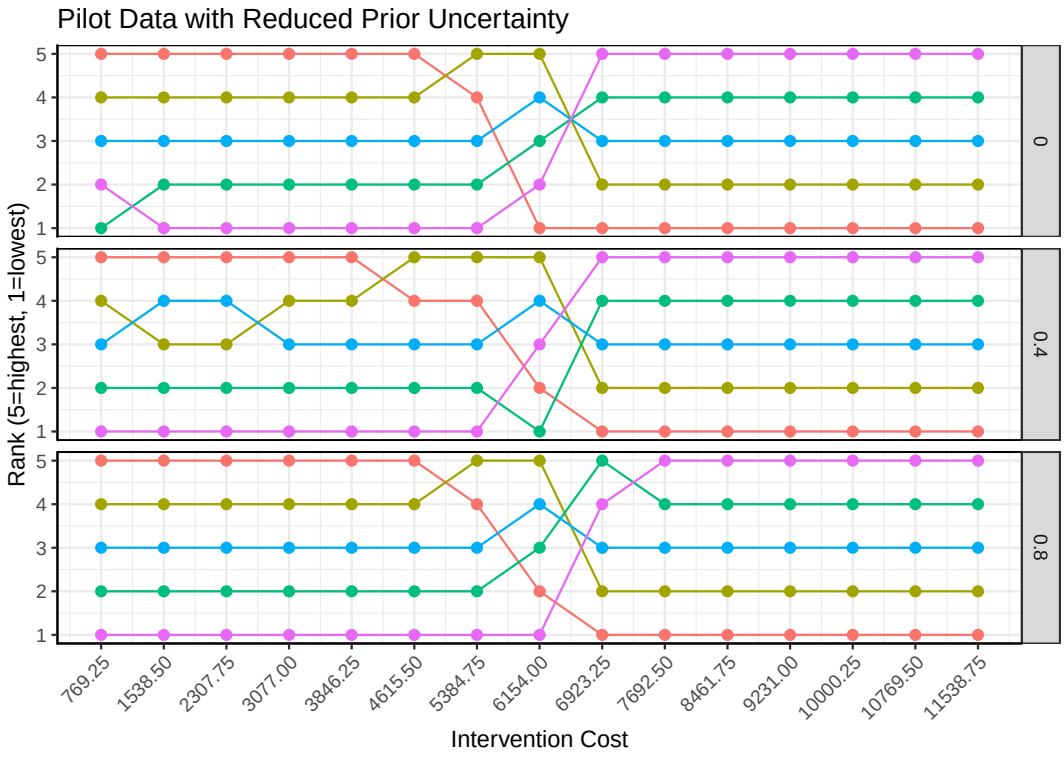
8.7.5 Comparison of Adjusted and Unadjusted Expected Value of Sample Information

Table E.1 in Appendix E summarises the adjusted and unadjusted estimates of the EVSI, ENBS, health economic model parameters and the primary outcome for the pilot data from the PSA samples. The model parameters and primary outcome give similar results for the pilot data with reduced prior uncertainty. There is also no difference between estimates for the baseline utility in the control arm as this is not affected by the design of the trial and so set to be equal.

As expected from the results in Chapter 7, the difference between the adjusted and unadjusted estimates of the primary outcome are greatest for the adaptive designs with five interims compared to two interims and highest for POC compared to OBF. The differences between the

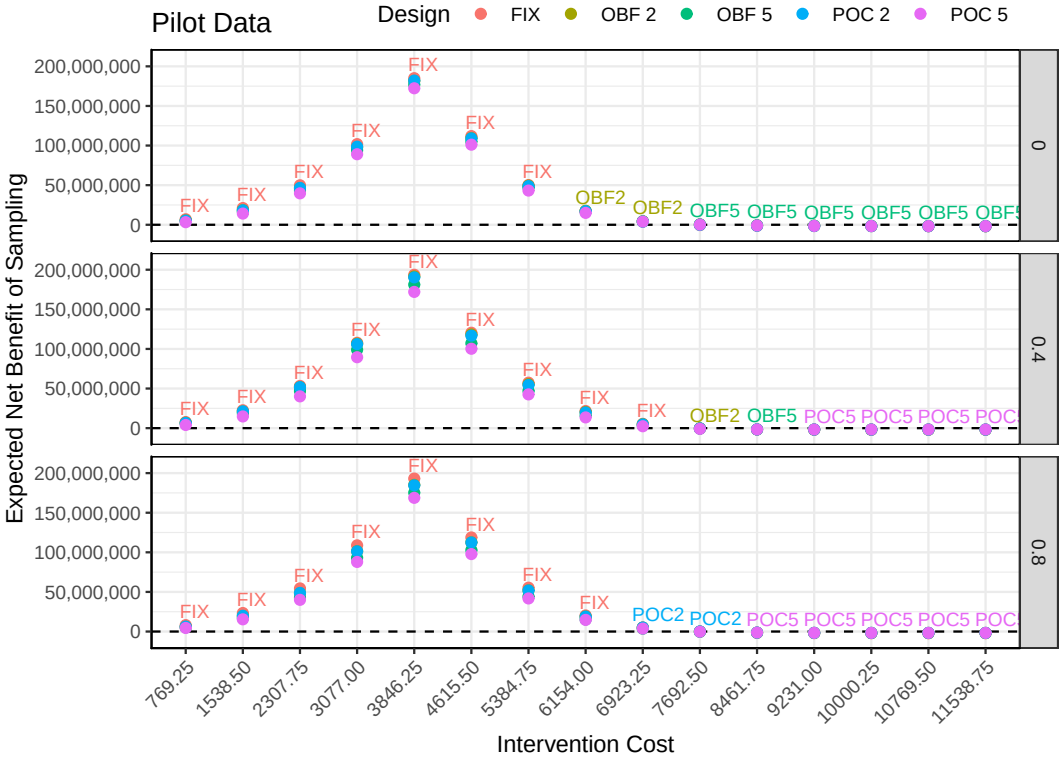


(A) Pilot Data

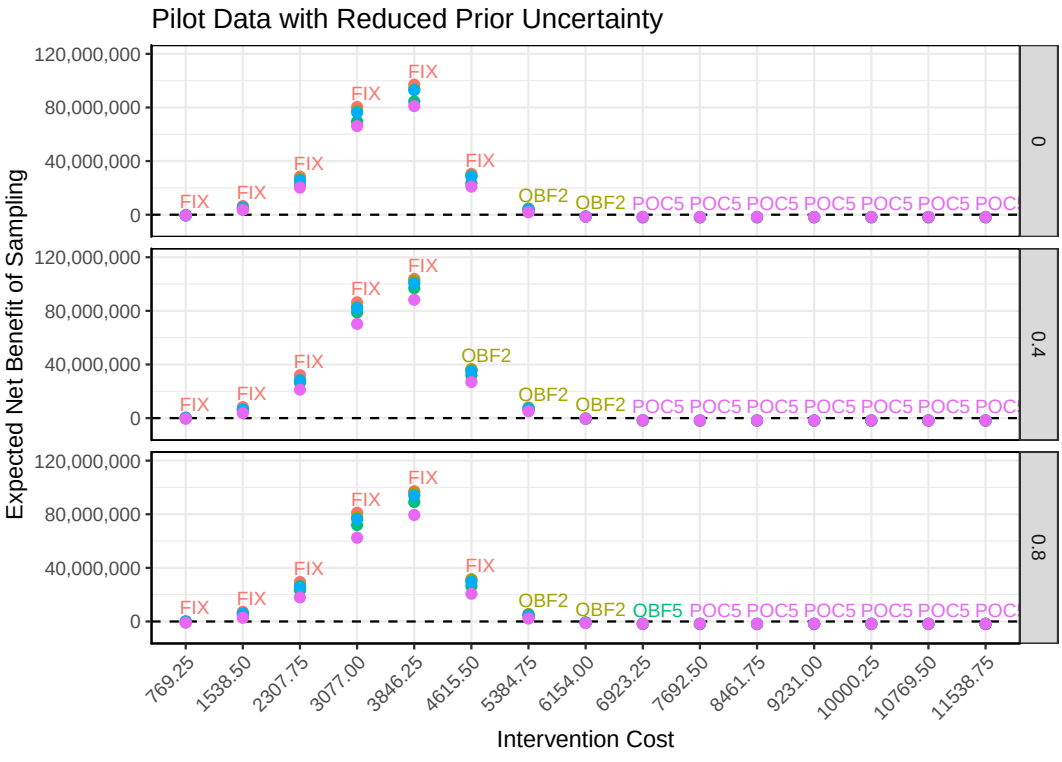


(B) Pilot with Reduced Prior Uncertainty

FIGURE 8.5: Trial designs ranked by their adjusted expected net benefit of sampling for five designs given the pilot data (5,996 PSA samples) and the pilot data with reduced prior uncertainty (6,000 PSA samples) for 15 different intervention costs. FIX; fixed sample size design, POC; Pocock stopping rule, OBF; O'Brien-Fleming stopping rule



(A) Pilot Data



(B) Pilot with Reduced Prior Uncertainty

FIGURE 8.6: Summary of the adjusted expected net benefit of sampling for five designs given the pilot data (5,996 PSA samples) and the pilot data with reduced prior uncertainty (6,000 PSA samples) for 15 different intervention costs. FIX; fixed sample size design, POC; Pocock stopping rule, OBF; O'Brien-Fleming stopping rule

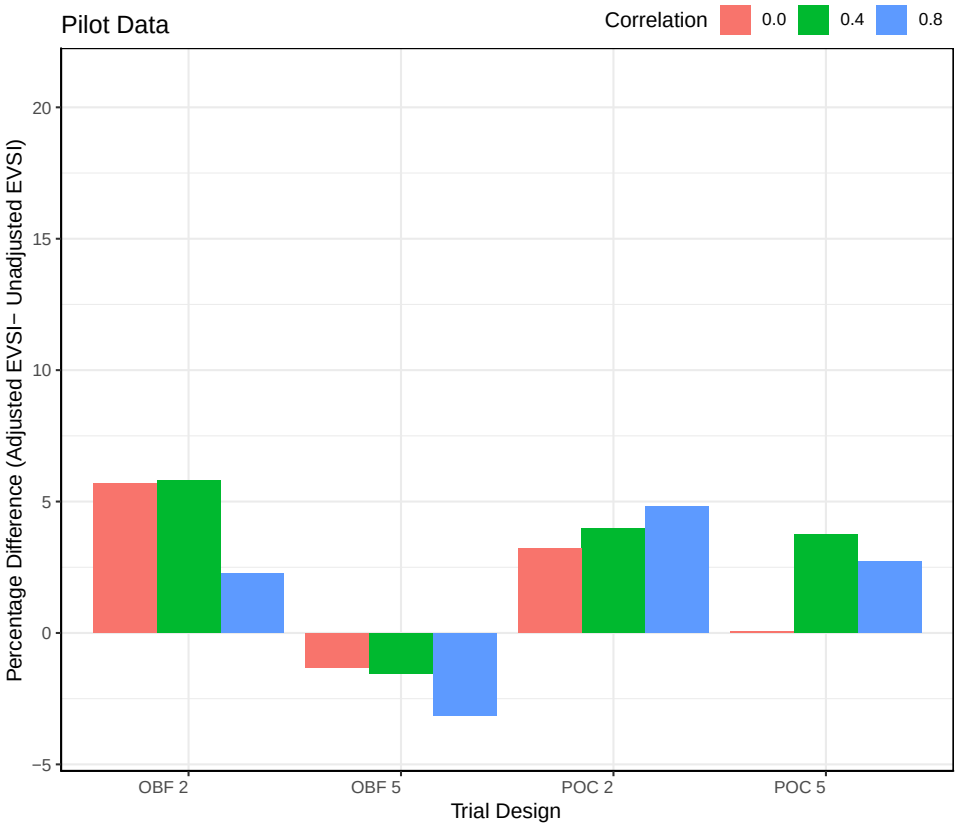
model parameters are small and close to zero for the costs parameters and the utility improvement. The percentage differences are higher for the probability of good response and probability of relapse, reaching 1.7% and 12.1% respectively. The primary outcome is used to calculate these model parameters and this is biased even when there is no correlation between primary and health economic outcomes. The difference between adjusted and unadjusted estimates of the relapse rate follows the same patterns as for the primary outcome.

As shown in Figure 8.7a, the percentage differences are small for the EVSI for the pilot data. There is no obvious pattern across the designs and correlations considered. The percentage difference is small (less than 6%) in all scenarios considered. Figure 8.7b gives the percentage difference for the pilot data with reduced prior uncertainty. The percentages are much greater in this setting, reaching over 20% for POC with five analyses and correlation 0.4. The differences are highest for all correlations for POC with five analyses reflecting the results seen in Chapter 7.

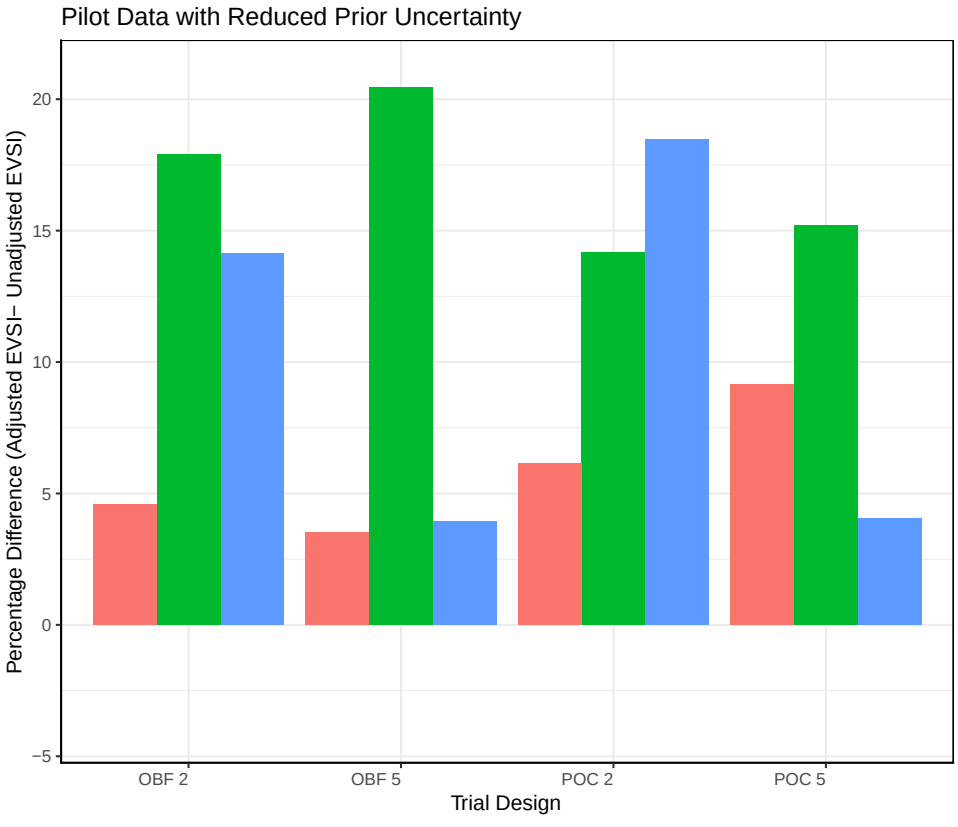
In some cases when the EVSI is small, the adjustment can give an ENBS below zero when the unadjusted estimate is greater than zero and vice versa. When the ENBS is greater than zero there is value in conducting the research from a health economic perspective, however a result below zero suggests the research is not worthwhile. Hence, the adjusted and unadjusted estimates for the same design can give contradicting results.

8.7.6 Impact of Analysis Methods on the Expected Value of Sample Information

In this chapter, the unadjusted ENBS for the fixed sample size design has been compared to the adjusted ENBS for the adaptive designs. There is no requirement to adjust the analysis following a fixed sample size design, as there are no early examinations of the data. As discussed in Section 7.7.4, different analysis methods are required to estimate the adjusted and unadjusted health economic model parameters, as the adjusted parameters cannot be directly estimated from the trial data. As such, the difference between adjusted and unadjusted estimates in EVSI may be a consequence of the difference analysis methods, as well as any biases introduced by the adaptive designs. This approach was adopted, however, as it felt that this best reflected the analysis approach that would be undertaken by a researcher conducting the analysis of a fixed sample size design.



(A) Pilot Data



(B) Pilot with Reduced Prior Uncertainty

FIGURE 8.7: Percentage difference between the adjusted and unadjusted estimates of expected value of sample information for the five designs and correlations 0.0, 0.4, 0.8. FIX; fixed sample size design, POC; Pocock stopping rule, OBF; O'Brien-Fleming stopping rule

8.8 Discussion

8.8.1 Summary of Key Findings from this Chapter

In this chapter the existing methods of EVSI to guide the design of fixed sample size designs have been extended to adaptive clinical trials. These methods build on the work of Chapters 6 and 7 to appropriately adjust for the adaptive nature of the trial design. The approach was operationalised in the context of the CACTUS case study, described in Chapter 5, to guide the design of a full scale randomised controlled trial following the CACTUS pilot. Reflecting the recommendations from the qualitative study that clinical effectiveness should remain the focus of an adaptive design, the O'Brien-Fleming and Pocock stopping rules were designed to demonstrate clinical effectiveness.

Based on the CACTUS pilot data and the reduced prior uncertainty scenario the fixed sample size design had the highest EVSI and ENBS compared to the adaptive designs. As the intervention costs were increased and ENBS re-calculated the potential savings in expected sample size offered by the adaptive designs gave them a higher ENBS. The adaptive designs with five analyses were preferred when variable costs were high as they offered early interim analyses with a small number of participants.

The fixed costs of the trial, incurred regardless of the design used were approximately 30% of the cost of sampling. This reflects the large upfront effort and resource expenditure when conducting a trial. The financial benefits of stopping a trial early are likely to be small when the fixed costs are high relative to the variable costs. For example, in a trial testing an expensive intervention such as a bone marrow transplant for the treatment of cancer (each transplant costing approximately £80,000 (Department of Health, 2019)). Likewise, when the variable costs associated with assessing the trial outcomes in all patients may be high, such as requiring an Octreotide scan to detect cancerous tumours, costing approximately £1,435 per scan (Department of Health, 2019).

In the CACTUS case study, the bias adjustments made a small impact on the estimates of EVSI and ENBS, even when there was high correlation between primary and secondary outcomes. However, there were greater differences between the adjusted and unadjusted estimates the reduced prior uncertainty scenario. In some cases adjusted and unadjusted estimates gave

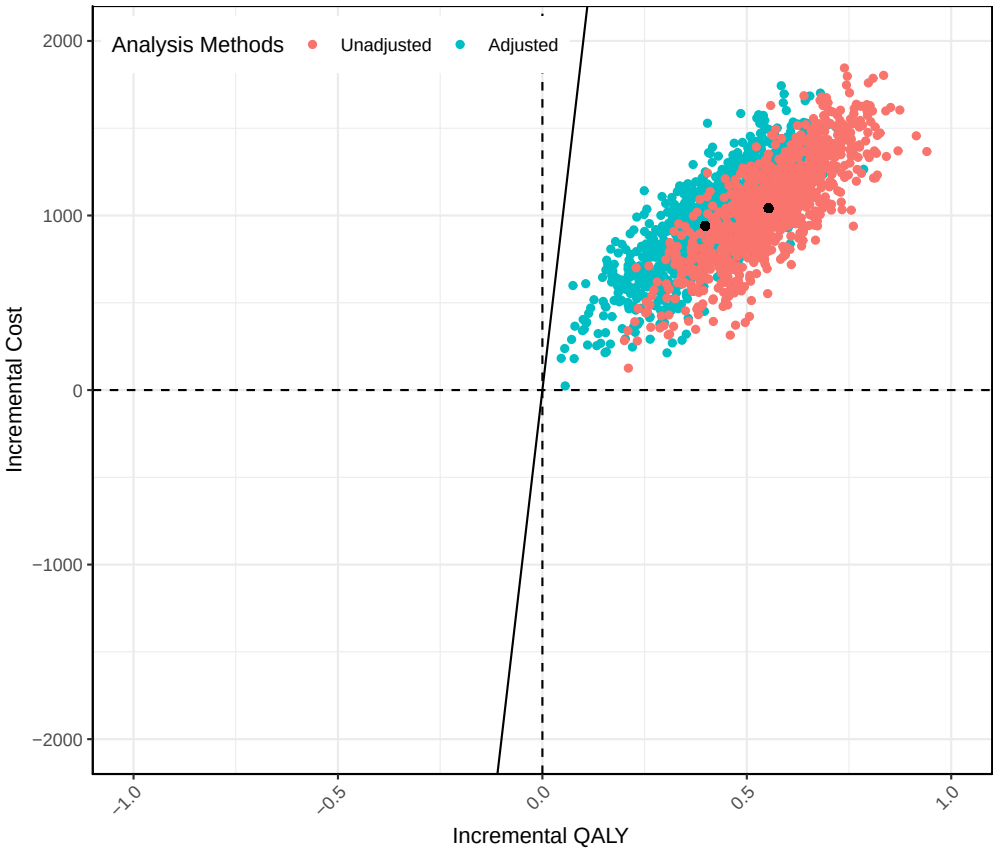
contradicting results. EVSI values change when there is a change in the decision uncertainty. If the bias adjustments have little impact on the decision uncertainty there will only be small differences between the adjusted and unadjusted EVSI estimates even if there are large differences between the adjusted and unadjusted model parameters estimates.

Figure 8.8 shows two hypothetical scenarios to illustrate when adjustments for the adaptive nature of the trial design are likely to affect EVSI estimates. In Figure 8.8a, the incremental cost and QALY are plotted in a cost-effectiveness plane, with a willingness to pay threshold of £20,000 per QALY represented by the black diagonal line. The red dots represent the PSA results for an adaptive design that have not been adjusted for the adaptive nature of the trial and the blue dots the adjusted PSA results . The black dot in the centre of each cloud of points is the mean ICER. In this case, the adjustments result in a small reduction in incremental costs and incremental QALY. However, all points lie below the willingness to pay threshold. This suggests there is little decision uncertainty that the new intervention is cost-effective and little value in conducting further research. The adjusted and unadjusted EVSI estimates are unlikely to be affected by the bias adjustments.

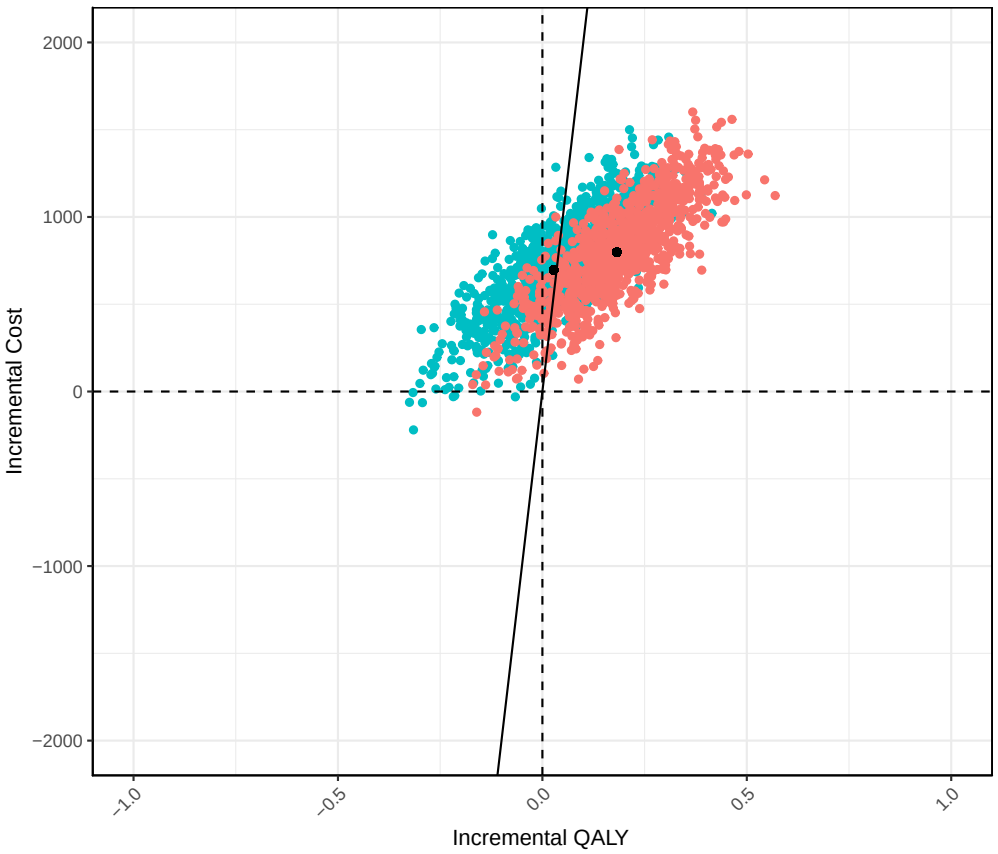
In contrast, the scenario in Figure 8.8b has much greater decision uncertainty. The unadjusted PSA points lie across the quadrants of the cost-effectiveness plane and either side of the willingness to pay threshold. In the unadjusted case, the new intervention is likely to be more costly and potentially less effective. The adjusted analysis, on the other hand, suggests that the new intervention may be more costly but is likely to provide more QALY. While there is still some decision uncertainty the majority of the adjusted PSA points lie below the willingness to pay threshold. The unadjusted and adjusted EVSI estimates will differ, with a larger EVSI expected for the unadjusted analysis compared to the adjusted analysis. This highlights the importance of adjusting the EVSI analysis to account for the adaptive nature of any trial designs considered. Plotting the results of the adjusted and unadjusted PSA analyses on the cost-effectiveness plane will help researchers to understand the impact that failing to adjust appropriately their analyses will have on their EVSI estimates.

8.8.2 How this fits with Existing Literature

As highlighted by Asselt *et al.*, 2018 the costing of various types of research is an under studied area of VOIA research. I have therefore outlined steps for researchers to calculate accurately



(A) Low Decision Uncertainty



(B) High Decision Uncertainty

FIGURE 8.8: illustration of two contrasting scenarios where adjustments for the adaptive nature of a clinical trial design are likely to affect the EVSI estimate

the cost of sampling for their own adaptive designs. This builds on the simplistic approach described in the application of the Chick *et al.*, 2017 model to the ProFHER clinical trial described by Forster *et al.*, 2019 and reflects the advice of Fenwick *et al.*, 2020. Asselt *et al.*, 2018 consider the costs of performing the research regardless of the payer in the Netherlands. A similar approach has been adopted in this chapter. It is, however, common for research costs funded by the NIHR to be costed at 80% of the full economic cost and for excess treatment costs and service support costs associated with conducting the research to be funded as part of the commissioning process and the Clinical Research Network respectively (NHS England, 2017). The appropriate perspective will depend on the target audience for the EVSI analysis and it is important to capture relevant costs for the audience.

The adaptive designs may have a lower ENBS compared to the fixed sample size design because their benefits are not adequately captured in the existing VOIA methods. Koffijberg *et al.*, 2018 suggest that additional constraints reflecting the priorities of the decision maker can be added to VOIA calculations. They consider a maximum acceptable rate of complication and a maximum acceptable budget. Alternatives relevant to adaptive clinical trials could favour designs that reduce the number of required patients favouring this ethical and/or practical advantage when there are a limited number of available participants. Further work should explore this with all stakeholders including members of the public.

As discussed in Chapter 2, there are many adaptive designs that exist and are being used in practice (Hatfield *et al.*, 2016). In this chapter, the focus has been on the group sequential design, however, the methods described in Section 8.4 could be considered for other adaptive designs. For example, the approach could be used to determine the optimal number of arms to start with in a multi-arm multi-stage (MAMS) design. It will be important to consider the practicalities of this approach in these designs using a case study to explore potential statistical and operational issues.

This is the first extension of non-parametric regression methods for the calculation of EVSI to the design of adaptive clinical trials. Alternative methods for the calculation of EVSI such as the importance sampling method (Menzies, 2016); Gaussian approximation method (Jalal *et al.*, 2015; Jalal *et al.*, 2018) and moment matching method (Heath *et al.*, 2018) could be used in Step 7

of the methods summarised in Box 8.2. These methods may provide advantages over the non-parametric approach, such as handling a large number of health economic model parameters and shorter computation times.

8.8.3 Considerations for Practice

VOIA should be used in practice to guide the design of clinical trials as they offer a formal way to quantify the value of fixed sample size designs compared to alternative adaptive designs. This will enable researchers to provide a quantified justification for their choice of adaptive design as per the recent guidance from the Food and Drugs Administration in the United States (U.S. Food and Drug Administration, 2019).

The adaptive designs performed best when the ENBS of the designs were below zero. This suggests that it is not worthwhile from a cost-effectiveness perspective conducting these trial designs; however, this may be common in practice where the motivation for conducting the research is to demonstrate the clinical effectiveness of an intervention. In these cases, the EVSI of the research may be low and so a design that answers the clinical question but with a small costs is preferable. As discussed by Willan *et al.*, 2012; Andronis *et al.*, 2016; Fenwick *et al.*, 2020 the methods of VOIA assume that clinicians will implement the treatment shown to be optimal even if there is a lack of evidence of a statistically significant clinical effect size. The findings of the qualitative study in Chapter 4 found this is not likely to be popular in practice. In reality, the value of information calculation may form part of many pieces of evidence used to design an adaptive clinical trial. It is anticipated these methods will be used by research teams (including clinical experts, statisticians, health economists and public advisors) to inform a wider discussion on the best choice of trial design.

These methods can be used by funding committees to inform the allocation of research funding. Participants in the qualitative study in Chapter 4 highlighted that formal VOIA calculations are not currently routine practice in funding committee decisions making. Instead, informal rules of thumb are used to compare the cost-effectiveness of proposals. The use of EVSI to guide the design of adaptive trials can formalise these cost-effectiveness considerations, providing a quantitative justification for the chosen design and an ENBS that can be compared with other study designs regardless of disease area or proposed design. This builds on the work of Glynn *et al.*, 2020, who have been working with the NIHR Health Technology Assessment programme

directors to trial the use of the Rapid Assessment of Need for Evidence (RANE) tool in their funding panel reviews. As discussed in Section 2.8.3, the RANE tool is an `R Shiny` application that calculates the value of further research with a small number of input parameters rather than requiring a full economic model.

As highlighted in Chapter 4, under current funding structures, these methods will require work to be carried out before the trial is funded. In situations like the CACTUS case study where pilot data and a health economic model were available the time burden of this work will be substantially reduced, however, in cases where this prior work is not available the resources required may be substantial and unobtainable without first securing funding. This is potentially one explanation for the lack of health economics being used in the design of adaptive clinical trials highlighted in Chapter 3. For these methods to be used to their full potential funding bodies, need to consider alternative ways to fund this work. An upfront investment will ensure study teams are able to propose cost-effective research designs based on quantifiable methods.

8.8.4 Limitations

As with other EVSI methods, the computation time for carrying out the methods is high (Strong *et al.*, 2015; Heath *et al.*, 2018). In the application to the CACTUS case study, 6,000 PSA samples were used, with each design taking approximately three hours to complete. Ideally, a full range of trial designs would be considered and compared when choosing the appropriate design for a future trial (Rothery *et al.*, 2020). However, this may not be viable given the time constraints associated with designing clinical trials for a grant application. One potential solution to facilitate the comparison of more potential design is to use high performance computing. This was not possible in this thesis due to the use of the `RCTdesign` package that currently does not run on Linux based platforms. Further work will explore how this method can be used without the `RCTdesign` to allow use of the high performance computers in Sheffield. The number of simulations in the CACTUS case study is sufficient to demonstrate the first application of the methods.

In the case study, the trial datasets in each row of the PSA have been generated assuming there is no missing data. Missing data are common in health economic data collected as part of trial, with participants incorrectly completing quality of life and resource use questionnaires (Faria *et al.*, 2014). Assuming there is no missing data is an idealistic scenario and highlighted by

Rothery *et al.*, 2020 the datasets should reflect the data collected. The impact this might have on the optimal design choice especially in the adaptive design context is unknown. This should be explored in further work.

8.8.5 Recommendations for Further Methodological Research

In the operationalisation of the methods to the CACTUS case study, adaptive designs with a clinical effectiveness rules based on the findings of the qualitative study (Chapter 4) and suggestions from the public advisory group. However, these methods could be extended and applied at the interim analysis of an adaptive clinical trial. This would allow research teams to assess the cost-effectiveness of continuing the research at that stage and even inform the conduct of the future trial such as the number and timing of further interim analyses. A simple approach would be to update the EVSI calculation with the available data at the interim; however, this would not be an optimal approach, as this does not take account of all possible future interim analyses. As discussed in Chapter 2, Pertile *et al.*, 2014; Chick *et al.*, 2017 consider an optimal approach using a Bayesian sequential economic evaluation model to approximate an optimal stopping rule based on cost-effectiveness. Care would be needed when applying this approach in practice to ensure the needs and preferences of stakeholders are met, as discussed in Chapter 4.

Further guidance is required on the costs associated with an interim analysis. Although in most cases these costs are likely to be a small proportion of the overall cost of sampling for a design it is important to capture accurately the costs associated with each design considered. The Adaptive Designs Working Group has acknowledged the lack of resources in this area. I am now part of a team working on the Costing Adaptive Trials (CAT) project to provide practical advice for research teams planning adaptive trial designs.

8.9 Chapter Summary

The literature review in Chapter 3 highlighted how health economics is rarely used the design and analysis of adaptive clinical trials in practice. The qualitative study in Chapter 4 discussed how stakeholders thought that clinical effectiveness should be the main focus of an adaptive design and were reluctant for health economics to be used to inform interim adaptations and monitor the trial. In this chapter the methods of non-parametric regression for the calculation

of EVSI were used to determine the optimal design for a group sequential design based on the number of analyses and clinical effectiveness stopping rule.

EVSI can be used to compare possible adaptive and non-adaptive trial designs when planning a clinical trial. This can guide and justify the choice of characteristics and prevent limited research budgets being wasted. It is recommended that both adjusted and unadjusted analyses are presented to control for the potential impact of the adaptive designs to maintain the accuracy of the calculations.

Chapter 9

Discussion

9.1 Introduction

Adaptive clinical trials use data collected as a trial progresses to inform modifications to the trial, without compromising validity or integrity. These trials are commonly designed such that interim decisions and modifications are based on clinical effectiveness. Despite its importance in the allocation of health research and healthcare resources cost-effectiveness is often secondary to clinical outcomes. It is currently unclear what impact using an adaptive design can have on a subsequent health economic analysis. Additionally, opportunities are potentially being missed to incorporate health economics into the adaptive trial at the design and analysis stages.

The overarching aim of the thesis was to answer the question:

How can health economics be used in the design and analysis of adaptive clinical trials to increase the efficiency of healthcare decision making?

A multi-disciplinary approach has been adopted to:

1. Review the current use of health economics in the design and analysis of adaptive clinical trials in the research literature and in practice.

2. Understand stakeholder views towards the use of health economics in the design and analysis of adaptive clinical trials.

3. Explore the potential for an adaptive design to impact the health economic analysis following a clinical trial.

4. Extend existing health economics methods to guide the design of an adaptive design whilst appropriately accounting for the adaptive nature of the trial design.

9.2 Chapter Aims

This chapter summarises the findings from each chapter of the thesis, demonstrates how they answer each of the research aims and evaluates the strengths and limitations of the approach taken. The findings are summarised to provide a picture of the current landscape of the use of health economics in the design and analysis of adaptive clinical trials. Recommendations for stakeholders in health technology assessments are formed, based on these findings, to guide the further development of these methods and their translation into practice.

9.3 Review of Health Economics in the Design and Analysis of Adaptive Clinical Trials in the Research Literature and in Practice

The first aim of the thesis was to review the use of health economics in the design and analysis of adaptive clinical trials in the research literature and in practice. This was important to understand the current research landscape and identify areas for further work. An investigation of the research literature and a review of current practice were reported in Chapter 2 and Chapter 3 respectively, to achieve this.

The novel review of current practice identified 37 clinical trials with an adaptive design and a health economic analysis. These trials were identified from a range of sources including `clinicaltrials.gov` and the NIHR Health Technology Assessment Journal. Data were extracted for each trial relating to the use of health economics in their design, analysis and reporting to establish how health economic outcomes were used.

Only $^{3}/_{37}$ trials considered health economics in their design and none of the trials seemed to adjust the health economic analysis to account for the adaptive nature of the trial. One of the 19 trials with results reported using health economics outcomes at an interim analysis. The reporting of these 19 trials was also felt to be suboptimal with little discussion of how the

adaptive design might impact the results and few trials made interim results available. Three exemplars were identified (OPTIMA trial, GDHT trial and PRESSURE-2 trial) that illustrated how health economics and adaptive designs could be used together in clinical trials. While none of the trials used these methods to their full potential, they serve as strong examples as to what can be achieved.

Despite existing methods literature that combined health economics and adaptive designs (described in Chapter 2) it was clear these methods were not being used in practice. To understand the barriers to this approach a qualitative study (Chapter 4) was undertaken to identify these issues before progressing with further methodological work.

As far as I am aware, this is the first work to investigate the application of health economic evaluations in trials with adaptive designs. The work reviewing adaptive clinical trials with health economics has been published in Value in Health (Flight *et al.*, 2019a).

To identify additional trials with an adaptive design and a health economic analysis published since the original review was undertaken the Health Technology Assessment journal was searched again on the 20.07.20. The source was chosen as in Chapter 3 the majority of trials meeting the inclusion criteria were UK, publicly funded clinical trials and the NIHR are the largest funder of health and care research in the UK (National Institute for Health Research, 2020). The same search strategy, outlined in Chapter 3, was used to identify articles published between 18.08.16 and 20.07.20. Twenty-six articles were identified. Eleven trials did not meet the inclusion criteria described in Section 3.4.2 and ten were already included in the previous analysis. Four new articles met the inclusion criteria and one article reported results for a trial that was included in the review of Chapter 3 but before the results were published. None of the trials adjusted their health economic analysis for the adaptive nature of the trial, although two of the trials considered appropriate adjustments for their primary outcome analysis. The DeCoDeR trial (Gabbay *et al.*, 2017) discussed using the internal pilot of the trial to consider a sample size re-estimation, contamination from individual randomisation and an evaluation of two recruitment strategies that would consider the cost-effectiveness of the approaches. Unfortunately the trial was stopped due to poor recruitment at the pilot stage so it is not possible to assess the role that cost-effectiveness played in this interim decision making. These results are consistent and complementary with the findings in Chapter 3: few trials are considering

health economics to its full extent in the design and analysis of adaptive trials and the bias adjustments methods are not being applied to health economic analyses.

9.4 Understanding Stakeholder Views towards the Use of Health Economics in the Design and Analysis of Adaptive Clinical Trials

Chapter 3 highlighted that there was limited use of health economics in the design and analysis of adaptive clinical trials in practice. The qualitative study, in Chapter 4, aimed to understand the key issues preventing these methods being used in practice. These findings were then used to inform the areas for further consideration in the thesis as well as recommendations made for the research community.

Twenty-nine participants from key stakeholder groups in health technology assessments were identified. This included researchers, decision makers and members of the public. Data were collected using interviews and focus groups. A framework analysis was used to identify themes in the data such as ethical, methodological and practical issues associated with the use of health economics in adaptive design clinical trials.

It was clear that researchers were not aware of the potential for an adaptive design to affect their health economic analysis. This supported findings from the review in Chapter 3. It was felt this lack of awareness has the potential to waste limited resources as inaccurate data may be used in healthcare decision making. As the use of adaptive design is increasing this was identified as an important area for further consideration in Chapter 6 and Chapter 7.

Additionally, there was a clear steer from the qualitative study participants that the aim of a clinical trial should be to demonstrate the clinical effectiveness of an intervention. This was supported by the public advisory panel. However, the study participants did acknowledge the importance of cost-effectiveness in healthcare decision making where budgets are limited. On this basis, the use of health economics to guide the design of adaptive clinical trials with adaptations informed by clinical effectiveness was explored in Chapter 8.

Researcher participants suggested that further case studies and training would be required for research teams to implement these approaches in practice. Therefore the methodological developments are operationalised in the CACTUS case study described in Chapter 5.

To the best of my knowledge, this is the first qualitative study to investigate the role of health economics in the design and analysis of adaptive clinical trials with stakeholders in health technology assessments. The finding that the primary aim of an adaptive clinical trial should be to concentrate on clinical effectiveness has the potential to impact all adaptive designs with health economic evaluations. The qualitative study has been published in Trials (Flight *et al.*, 2020).

9.5 Exploring the Potential Impact of an Adaptive Design on a Health Economic Analysis Following the Trial

Chapter 3 and Chapter 4 highlighted a lack of awareness of and guidance for the potential for an adaptive design to impact a health economic analysis following an adaptive clinical trial. Methods for adjusting the point estimates and confidence intervals for primary and secondary outcomes following adaptive trials are available but have not be considered in the health economic context.

This thesis focusses on the group sequential design as it is the most common adaptive design used in practice (Hatfield *et al.*, 2016; Hartford *et al.*, 2018; Bothwell *et al.*, 2018). It is well documented that the analysis following these trials requires adjustment (Whitehead, 1997; Jennison *et al.*, 2000). As these designs are increasingly used in practice there is an urgent need to understand their impact on a health economic analysis. Failing to do this quickly could result in inaccurate and incorrect healthcare decision making that undermines the potential benefits these innovative designs offer.

Emerson *et al.*, 1990 recommended the Bias Adjusted Maximum Likelihood estimate (BAMLE) and the Sample Mean Ordering (SMO) approach for adjusting point estimates and confidence intervals of a primary outcome, respectively, following a group sequential design. In Chapter 6, the work of Whitehead, 1986a; Whitehead, 1986b for BAMLE for primary and secondary outcomes was extended to the within trial health economic outcomes; costs, QALY and INB. The theory for calculating adjusted confidence intervals of Emerson *et al.*, 1990 and Skalland, 2015 was extended for these health economic outcomes. The theory was extended to health economic model parameters estimated from the adaptive design in the context of the CACTUS health economic model, introduced as the case study for the thesis in Chapter 5.

In Chapter 7, a simulation study assessed the extent to which the stopping rule, number of interim analyses and correlation between primary and health economics outcomes affects the within trial and model-based health economic analysis following a group sequential design. BAMLE and SMO approaches were compared to the unadjusted approach, that do not account for the adaptive nature of the trial.

The simulation study showed how the bias in the point estimate of within trial health economic outcomes increases with the correlation between the primary and health economic outcomes, as anticipated from the literature (Whitehead, 1997). The bias was greater for the Pocock stopping rule compared to the O'Brien-Fleming stopping rule and increased as the number of interim analyses increased, consistent with previous work (Pinheiro, 1997). Adjusted confidence intervals provided coverage close to the nominal level and were wider than the unadjusted intervals, reflecting the greater uncertainty from the smaller sample size of the adaptive designs. The overall impact of the adaptive design on the health economic analysis was reduced in a more realistic simulation of trial data where correlations between primary and health economic outcomes were small, resulting in little bias in the health economic analysis.

It is anticipated that the impact could be greater when the primary outcome plays an influential role in the health economic model. The correlations between outcomes can be difficult to anticipate before a trial, especially if there is limited prior information. The consequences of failing to adjust the analysis are large, therefore, it is recommended that all health economic analyses that use data from an adaptive design report both adjusted and unadjusted analyses.

As far as I am aware, this is the first extension and operationalisation of the adjustment methods following a group sequential design to the health economic context. The work has potential to impact all health economic analyses that follow an adaptive clinical trial to ensure accurate healthcare decision making even when these innovative designs are used.

9.6 Extending Expected Value of Sample Information to Guide the Design of an Adaptive Clinical Trial

Based on the findings of Chapter 3 it was clear opportunities were being missed to incorporate health economics into the design and analysis of adaptive clinical trials. The qualitative

study in Chapter 4 found that stakeholders in health technology assessments considered the aim of a clinical trial as being to demonstrate clinical effectiveness before demonstrating cost-effectiveness. On this basis, a sensible compromise is to use health economics to guide the design of an adaptive clinical trial. However, once the trial is running the primary focus is to demonstrate clinical effectiveness. For example, using VOIA to guide the design of a group sequential design that uses a stopping rule based on clinical effectiveness to determine whether the trial can stop early.

Chapter 8 aimed to strike this balance. The existing theory of expected value of sample information (EVSI) (introduced in Chapter 2) was extended to guide the design of an adaptive clinical trial. The CACTUS case study was used to illustrate how the method would work in the real-world setting, comparing adaptive and fixed sample size designs.

To extend the existing EVSI methods two key additional considerations were required:

1. The appropriate adjustment of analyses to account for the adaptive nature of the trial, shown to be important in Chapter 6 and Chapter 7;

2. The calculation of the cost of conducting the adaptive research design to accurately capture any potential cost savings from this approach.

Applying this approach to the CACTUS case study showed that the fixed sample size design was optimal, from a health economic perspective, as it gave the highest expected net benefit of sampling (ENBS) for little additional cost compared to the O'Brien-Fleming and Pocock stopping rules considered. However, in a sensitivity analysis where the intervention costs for the CACTUS trial were varied, the cost savings of the adaptive design became clear for scenarios where there was a high intervention cost as, on average, these trials required fewer participants.

These methods should be used in practice to guide the design of clinical trials as they offer a formal way to quantify the value of fixed sample size designs compared to alternative adaptive designs. This will enable researchers to provide a quantified justification for their choice of adaptive design as required in the recent guidance from the Food and Drugs Administration (U.S. Food and Drug Administration, 2019).

This work has the potential to impact the design of all adaptive clinical trials. The design of a trial is influenced by many factors, such as the availability of the patient population, willingness of participants to be randomised and practicalities of delivering the intervention and conducting the trial. It is anticipated that research teams (that include clinical experts, statisticians, health economists and public advisors) will use these methods to inform a wider discussion on the best choice of trial design. This complements the work of Glynn *et al.*, 2020 who have developed the RANE tool for VOIA and are in discussion with the NIHR HTA funding panel to trial its using in their funding panels.

The work in this chapter also complements the work of Pertile *et al.*, 2014; Chick *et al.*, 2017; Forster *et al.*, 2019 who have developed a stopping rule for a trial based on the cost-effectiveness of the research. Their work considers evaluating the cost-effectiveness of the research as the trial progress, however, it can be used at the initial design stage to determine whether it is optimal to use a fixed or adaptive design. I am now collaborating with Chick, Pertile and Forster to extend the work of this thesis further considering the practical steps required to use these approaches in NIHR funded research.

9.7 Public Involvement in this Thesis

A key component to the success of the work in this thesis was the involvement of members of the public. Public involvement in research is defined as 'research carried out with or by members of the public' (INVOLVE, 2016). It ensures that is research relevant and useful to the patient and the public (Boote *et al.*, 2011), as well as increasing transparency and accountability.

9.7.1 Public Involvement Methods

As discussed in Chapter 1, an advisory panel of members of the public was formed following an information session held in April 2016. The aim was to provide more information about the research planned and ideas for how the advisory panel would support this work. Attendees were able to ask questions as well as find out more about how the group would be involved.

Seven members of the public expressed an interest in joining the advisory panel. Five meetings were held with the advisory group over the duration of the PhD. At each meeting, the group were provided with training in the methods of adaptive clinical trials and health economics required to contribute to the discussions of the work. The remainder of the meetings were spent discussing the progress of the research and specific areas that required their input. This included the design of the qualitative study in Chapter 4 and drafting the plain English summary. Annual reports were sent to the group between meetings to keep them up to date with progress and the group were invited to share any thoughts between meetings via email.

9.7.2 Public Involvement Reflections

The public advisory panel has directly influenced the research reported in this thesis, ensuring it is relevant and appropriate from the public perspective. The group emphasised throughout the project the importance of building on existing methods and approaches rather than trying to radically change practice. This was reflected in the work of Chapter 6 and Chapter 8 where the methodological developments focussed on improving the accuracy of health economics analyses following adaptive designs as they are commonly conducted and using health economics to design a clinical trial with a clinical effectiveness stopping rule.

The group regularly discussed issues important to the public such as quality of life, safety and the impact of the approach on trial participants (including follow-up) if a study is stopped early. The group were highly influential in the design of the qualitative study (Chapter 4) where changes were made to the information sheet, consent form and topic guide for members of the public. Additionally, the group suggested creating a video of the background information needed for the qualitative study.

The group have helped to develop dissemination materials for the thesis. A poster was presented at the NIHR Trainee Meeting Leeds 2016 outlining the fellowship in plain English. The abstract for the poster was written in collaboration with the group and the group helped to design the poster produced. This poster was also presented at the ScHARR Post Graduate Research conference 2018 and was awarded the prize for best poster. The group also co-wrote the plain English summary at the start of the thesis.

9.7.3 Recommendations for Public Involvement in Methods Research

An 'information session' can be useful for members of the public to find out about the project and how they can get involved. Even for experienced public advisors, methods projects such as this can be outside their clinical area of interest or current knowledge base. Writing a role description and terms of reference for the group can help to ensure time with members of the public is used efficiently and meets all expectations. It was helpful to have a large number of advisors as the number engaged in the project at one time might fluctuate. This may be more prevalent in methods research projects where advisors may not have a personal interest in the topic, as it does not relate to a clinical area they are interested in. Meeting more than once a year could facilitate greater input into the project; however, needs to be balanced against the time it takes to plan these meetings and the additional burden this places on the advisors.

Each member of the advisory panel had previous experience of public involvement in health research. This was useful as the group were familiar with clinical trials and research. The work in this project was just an extension of their existing knowledge base. However, their views and contributions may differ from members of the public with no previous experience, from different backgrounds. By recruiting from existing groups, this perpetuates issues around under-represented groups in public involvement. One suggestion from the advisory group was to have 'bring a friend' meetings so a more experienced public involvement advisor can bring a less experienced friend. This will hopefully create a less intimidating experience for new advisors whilst increasing the pool of people involved in research.

Before starting the fellowship, I had no experience of involving the public in my research. There are not always clear opportunities for public involvement in methods research compared to research that directly involves members of the public as participants (such as a clinical trial). However, realistic aims can be set that reflect attainable levels of involvement. Over the last four years, I have developed my communication skills, being able to explain complex statistical approaches in plain English. When planning public involvement it is important to allow sufficient time for the organisation of meetings and preparation of materials so the public can contribute in a meaningful way.

The use of public involvement in methodological research in my thesis has been pioneering. I have looked to share the best practice that I have developed from my own research and research with other colleagues (Flight *et al.*, 2019b). I have been contacted by other researchers planning to involve the public in their methods based research. I was invited to talk about my experiences at the NIHR Statistics Group, Statisticians as Principal Investigators meeting (March 2018) and at an MRC TRMP meeting in November 2019. I have since been invited to join a steering group organised by the Health Research Board Trials Methodology Research Network (HRB-TMRN) to develop plain English resources explaining trial methodology.

9.8 Strengths and Contributions of this Thesis

A multidisciplinary approach has been adopted to achieve the aims of this thesis, using both qualitative and quantitative approaches to ensure that the recommendations made are focussed on the long-term goal of applying these methods into the design and conduct of research in the NIHR setting and beyond.

Robust methodology was used throughout the thesis, starting with a comprehensive review of current practice in Chapter 3 that used a wide range of sources to identify clinical trials with an adaptive design and health economic analysis beyond the UK setting. The qualitative study in Chapter 4 included a range of stakeholders, including health economists that had been hard to reach in previous research on adaptive designs (Dimairo *et al.*, 2015). The qualitative study included members of the public, a stakeholder group that could easily be overlooked in the development of statistical methods, however with an informative and valuable perspective. The simulation study of Chapter 7 and analysis in Chapter 8 were anchored in the CACTUS case study, providing a real-world perspective to the development of the methods. A range of scenarios was also used in each chapter to understand the behaviour of the methods beyond the context of the case study and increase the generalisability of the results.

A further strength of the thesis is the embedded public involvement from application to dissemination. When starting this research project, there were few resources available for researchers and the public about involvement in methods research projects. With the support of the advisory panel, we have navigated this process successfully illustrating how the public can

be involved in this work and have been able to share our experiences to encourage others to do the same.

Results from the thesis have been actively disseminated to a wide range of audiences and in a number of formats for frequent feedback. The background and research aims from Chapters 1 and 2 were presented as oral conference presentations at Society for Clinical Trials 2016, Royal Statistical Society 2018, Research Student Conference in Probability and Statistics 2018 (awarded prize) and as a poster at the NIHR Trainee Meeting 2016. The review of current practice in Chapter 3 was published in Value in Health. This work identified by the editors as of high interest and so a press release was written. This was also identified by Cytel of interest and became a topic on their blog. This demonstrates the wide reaching impact of this work outside of the UK public funding sector. This work was presented at the joint Society of Clinical Trials and International Clinical Trials Methodology Conference 2017 in an invited session I organised, presented at and chaired.

The findings of the qualitative study of Chapter 4 have been published in Trials and have been presented as an oral presentation at the International Clinical Trials Methodology Conference 2019. Early findings from Chapters 6 and 7 were discussed at the Health Economic Study Group 2018 and results from Chapter 8 were presented as a poster at Society for Medical Decision Making 2019. I was also invited to give external seminars at the University of York and University of Bristol on these chapters.

9.9 Limitations of this Thesis

The focus of the thesis has been the use of health economics in adaptive designs for definitive trials. This has been considered without restriction in the review of current practice in Chapter 3 and in the qualitative study in Chapter 4. The work of Chapters 6 and 7 focusses on the group sequential design. This design was chosen as it is one of the most commonly used adaptive designs and has been well discussed in the analysis adjustment literature. It is important to note that other adaptive designs are used and are growing in popularity. These designs may face similar and contrasting issues relating to the adjustments of health economic analyses and opportunities for using health economics during the design and analysis of the trial. For example, a multi-arm multi-stage trial where interim analyses are used to determine which of

multiple treatment arms are carried forward to the next stage of the trial could consider the cost-effectiveness of each intervention at the interim decision making.

In the adjustment of point estimates and confidence intervals throughout the thesis the BAMLE (Whitehead, 1986a; Whitehead, 1986b) have been used with the SMO approach for confidence intervals (Emerson *et al.*, 1990; Skalland, 2015). This choice was justified based on the recommendations of Emerson *et al.*, 1990 and the availability of these adjustments for primary outcomes in the `RCTdesign` package. This code was then extended to the health economic outcomes in Chapter 6 using existing code provided by Skalland, 2015 and by creating new user defined functions. Other approaches are available such as the median unbiased estimate (Whitehead, 1997) and the uniform minimum variance unbiased estimator (Emerson *et al.*, 1997). These methods are likely to have different properties compared to the approaches used, but are still likely to provide a more accurate analysis, that reflects the adaptive nature of the trial than the unadjusted health economic analysis.

The `RCTdesign` package provided existing `R` code for the implementation and analysis of group sequential designs. This was package was chosen as it was thought it would be easy to extend the analysis to include adjusted point estimates and confidence intervals for secondary outcomes. However, the package is computationally intensive and is not currently able to run on Linux based systems; despite confirmation from the developer this would be available in 2018. This means it has not been possible to utilise the high performance computer at the University of Sheffield. Despite efforts to parallelise the code to run efficiently on a Windows machine, the number of simulations considered for the simulations studies of Chapter 7 and the number of bootstrap samples used in Chapter 8 were limited to 2,000 and 6,000 respectively. While this number of simulations is acceptable to demonstrate the theory developed in the thesis, the ability to conduct a higher number of samples will be desirable especially when comparing designs for a future clinical trial.

The qualitative study participants, while including a range of stakeholders, did not include many female researchers or policy makers such as NIHR research programme managers or NICE technical advisors. These groups may have important views about the use of health economics in the design and analysis of adaptive clinical trials. The findings from Chapter 4 could be used to open this dialogue with these groups.

The generalisability of the results in Chapter 7 and Chapter 8 are limited to some extent by the CACTUS case study used to illustrate the theory development relating to the adjustment of model parameters and the comparison of the cost-effectiveness of research designs. While there was little bias in the within trial and deterministic model results in the CACTUS setting, it is not possible to draw conclusions about the level of bias in all health economic analyses. This result is likely to depend on the correlation structure of the CACTUS case study and the role of the trial based parameters in the health economic model. Likewise it is not possible to draw conclusions that the fixed sample size design will always be the most cost-effective design option as was found in Chapter 8 as again this will depend heavily on the context such as prior information available, amount of information gained from the trial and the cost of conducting the research. Instead, these chapters have shown how the methods developed can be used in the real-world setting, facilitating the adoption in other trial scenarios.

9.10 Summary of Recommendations

Based on the work conducted in the thesis the following recommendations are made relating to the use of health economics in the design, analysis, practical implementation and reporting of adaptive clinical trials. Key points are summarised in Figure 9.1.

9.10.1 Design of Clinical Trials with an Adaptive Design and Health Economic Analysis

Reflecting the perspectives of stakeholders in the qualitative study of Chapter 4 it is recommended that the importance of clinical effectiveness is reflected in the development of methods for using health economics in adaptive trials. Possible approaches suggested by qualitative study participants include

- Using health economics to guide the design of an adaptive design where adaptations to the trial are informed by clinical effectiveness, as described in Chapter 8.
- Using early examinations of the trial to check all health economic data are being collected as required.
- Using early trial data to update the health economic model.

- Using a hierarchy of interim decision rules where any decisions made based on cost-effectiveness depend on decisions made about clinical outcomes.
- Only considering health economic outcomes at later examinations of the data.
- Using health economic data to make modifications to the trial such as increasing the sample size but not major changes such as to stop the trial early.

Following a lack of pre-specification of health economic analysis in the review of Chapter 3 and work by Thorn *et al.*, 2017, it is recommended that proposed health economic analyses are outlined in a Health Economic and Decision Modelling Analysis Plan (HEDMAP) before the start of an adaptive design. This will be crucial in maintaining the validity and integrity of adaptive designs that use health economics, with analysis plans including a description of the monitoring and adaptation plan, as well as pre-specification of methods used at interim analyses (Thorn *et al.*, 2017; U.S. Food and Drug Administration, 2019)

Finally, it is recommended that members of the public should be fully involved in the design of a trial; advising on the role of health economics; developing materials presented to potential participants about the trial design and the impact it might have on them; and developing plain English summaries of the results. This key role of the public was highlighted by the public advisory panel supporting this research as well as public participants in the qualitative study of Chapter 4.

9.10.2 Analysis of Clinical Trials with an Adaptive Design and Health Economic Analysis

It is recommended that both a health economic analysis that uses data from an adaptive design considers both an unadjusted and adjusted analysis accounting for the adaptive nature of the trial. As shown in Chapter 7 there is great potential for the adaptive nature of the trial to impact results. This will prevent the unintentional introduction of bias that could compromise healthcare decision making and potentially prevent patients receiving the treatment they need as resources are being wasted on research and treatments that are not cost-effective.

9.10.3 Implementation of Clinical Trials with an Adaptive Design and Health Economic Analysis

To facilitate the implementation of health economics in the design and analysis of adaptive clinical trials in the real-world setting it is recommended that software and tutorial style case studies are developed. More training resources were identified by qualitative study participants in Chapter 4 as important if health economics is to be successfully implemented in trials and a lack of current resources is a possible explanation for the lack of trials using health economics to its full potential in adaptive trials in the review of Chapter 3. This will help research teams to understand the methods and allow them to interpret the results of trials using this approach or use these methods in their own research. A Practical Adaptive and Novel Designs Toolkit (PANDA) is under development that aims to provide researchers with training materials on adaptive design clinical trials (Dimairo *et al.*, 2019a). I am co-applicant on the project and will provide materials on considerations for health economic analyses when an adaptive design is used.

Despite mixed opinions in the qualitative study in Chapter 4 it is recommended that health economists are included on all data monitoring and ethics committees (DMECs) where health economics is used as part of the design and analysis of adaptive trials. This is felt to be important if health economic data inform interim adaptations to a trial. Existing resources that help research teams identify DMEC statisticians, such as StatLink (NIHR Statistics Group, 2018), could be extended to identify health economists. All DMEC members could be paid for their contribution and time. Mock DMECs can be used to train members where they can review the health economic and clinical data and see where issues with using the health economic data arise.

Chapter 3 identified limited guidance provided to researchers on the potential for an adaptive design to impact a health economic analysis. To ensure that health economic analyses following an adaptive design are not compromised it is important that guidance documents for the economic evaluation of clinical trials, such as the NICE Guide to Technology appraisals (National Institute for Health and Care Excellence, 2013a) are updated to highlight the potential impact of the design on analyses. Further support could be provided to research teams though

technical support documents provided by the Decision Support Unit (National Institute for Health and Care Excellence, 2020).

It is recommended that funding bodies such as the NIHR, provide alternative funding options that allow researchers to develop new trial designs. Possible options suggested by qualitative study participants in Chapter 4 include researchers being given time at the start of a study to fully develop an adaptive trial design that uses health economics. Researchers could also look for methodology grants to fund the development of designs. To facilitate this change, it is suggested that groups representing statisticians and health economists (such as the MRC Adaptive Designs Working Group and ISPOR— The Professional Society for Health Economics and Outcomes Research) work together to persuade funders and regulators to fund adaptive clinical trials.

Qualitative study participants in Chapter 4 suggested that statisticians and health economists frequently work independently. It is recommended that they should be encouraged to work together and to increase communication to facilitate the implementation of health economics in adaptive trials by sharing expertise. Locally, health economists and statisticians working on the same clinical trial should have regular meetings throughout the study. Nationally, joint events between groups such as the NIHR Statistics Group, NIHR Economics Group and the Health Economic Study Group should be arranged to discuss common issues and encourage training in statistics for health economists and health economics for statisticians.

It is important to understand the costs of conducting an adaptive trial such as the costs of finding patients, training staff and analysing data so that you can compare adaptive and non-adaptive trial designs and inform stopping rules based on health economics, as demonstrated in Chapter 8. It is important to develop a standardised approach for calculating the costs of an adaptive clinical trial. This work is on-going in the Cost of Adaptive Trials study on which I am a co-applicant.

9.10.4 Reporting of a Clinical Trial with an Adaptive Design and Health Economic Analysis

Chapter 3 found the reporting of clinical trials with an adaptive design and health economic analysis to be suboptimal. To improve the reporting of the health economic analysis of adaptive designs the CHEERS checklist should be extended to this setting to improve the reporting of clinical trials with an adaptive design and health economic analysis.

As highlighted in the Adaptive Design CONSORT extension and in the simulation study in Chapter 7 it is important to report the methods used for analysis, including any adjustments made to account for the adaptive nature of the trial. It is recommended that both adjusted and unadjusted analyses are presented (Dimairo *et al.*, 2019b). This recommendation should be applied to health economic outcomes and analyses, so a reader can make a judgement about the most appropriate estimates for their needs. It is also recommended that appropriate points from these guidelines should be applied to clinical trial registries and in the trial protocol.

9.11 Recommendations for Further Methodological Research

There are a number of opportunities for further work stemming from this thesis. As a starting point the methods developed should be explored in the context of case studies with different characteristics to the CACTUS case study described in Chapter 5. This could include a health economic model that uses the primary outcome from the trial analysis directly.

The methods of Chapter 8 that extend EVSI to guide the design of adaptive clinical trials could consider alternative methods for calculating EVSI. Additionally, these calculations could be updated using data collected at an interim analysis to determine the optimal design of the remainder of the trial, such as the number or timing of future interim analyses. However, this approach would need to consider the concerns raised in Chapter 4 regarding the clinical effectiveness focus of a trial, perhaps being used as a supplementary interim analysis once clinical effectiveness has been assessed.

The within trial health economic analysis methods used in the thesis have used mean estimates of the costs and quality adjusted life years in the interventions arms being compared. More sophisticated methods such as seemingly unrelated regression can be used in a within

FIGURE 9.1: Summary of recommendations for design, analysis, implementation and reporting of clinical trials with an adaptive design and health economic analysis

1. **Design**
   - Reflect the importance of clinical effectiveness to stakeholders in the design of the trial.
   - Pre-specify analyses in a health economic analysis plan especially if health economics is to inform interim decision making.
   - Include members of the public when designing a trial.
   - Develop accessible information about the trial design and the impact it might have on the for potential trial participants.
2. **Analysis**
   - Calculate adjusted and unadjusted point estimates and confidence intervals for primary and health economic outcomes.
3. **Implementation**
   - Develop software and tutorial style case studies.
   - Include a health economist on data monitoring and ethics committees.
   - Updated guidance documents for researchers.
   - Develop standard approach for calculating the costs of an adaptive clinical trial
   - Funders should provide adequate resources to fund the development of these research designs.
   - Encourage statisticians and health economists to work together.
4. **Reporting**
   - Extend the CHEERS checklist to include points specific to adaptive designs
   - Include a description of adjusted and unadjusted analysis methods.

trial analysis that accounts for the relationship between costs and QALYs and also allows for the adjustment of covariates such as baseline utility (Willan *et al.*, 2004; Manca *et al.*, 2005). Regardless of the approach used, if the analysis follows an adaptive design adjusted point estimates and confidence intervals should be presented. Further work could extend the theory of this thesis to the seemingly unrelated regression approach.

To encourage the use of these approaches in practice, as highlighted in Chapter 4, software, case studies and training resources need to be developed. The `R` code developed during the thesis could be adapted to work outside of the `RCTdesign` package because of its noted computation limitations and developed into a standalone package or Shiny application such as the RANE tool or the SAVI tool (Strong *et al.*, 2020). This will make the methods accessible to research teams as well as funding panel members and policy makers.

To facilitate the promotion of the use of health economics in the design and analysis of adaptive clinical trials the EcoNomics of Adaptive Clinical Trials (ENACT) project has been set up. ENACT aims to explore the use of value-based adaptive clinical trial designs for efficient delivery of NIHR research. This project is funded by the NIHR CTU Support Fund and is in collaboration with researchers at the University of York, INSEAD and the University of Verona. This collaborative group was formed following a weeklong research visit with Dr Martin Forster during the PhD research. The project brings the theoretical work of Pertile *et al.*, 2014; Chick *et al.*, 2017 and the theoretical and practical work of this thesis into one project to explore how value-based adaptive designs can be used in practice in the NIHR.

9.12 Conclusions

This thesis has considered how health economics can be used in the design and analysis of adaptive clinical trials. The recommendations made will aid researchers and decision makers who want to increase the efficiency of their health research and to embed health economics into the design and analysis of their adaptive trials. The methods have been developed to ensure they appropriately adjust analyses to maintain accuracy of decision making.

Cost-effectiveness considerations are unavoidable in a health care system with limited funding for health research and health care interventions. Adaptive designs provide an appealing alternative to large and cost fixed sample size designs in appropriate scenarios. To date the

use of health economics and cost-effectiveness considerations in adaptive designs had received limited attention in the research literature and in practice.

The research in this thesis has identified potential barriers to the implementation of health economics in the design and analysis of adaptive clinical trials whilst also making practical recommendations to maximise the opportunities that using health economics and adaptive designs together can bring. It is hoped the innovative research from this thesis will impact all adaptive trial designs with a health economic analysis.

Bibliography

[1] National Institute of Health Research. (2020). Annual Efficient Studies funding calls for CTU projects, [Online]. Available: https://www.nihr.ac.uk/documents/ad-hoc-funding-calls-for-ctu-projects/20141 (visited on 02/26/2020).

[2] P. Gallo, C. Chuang-Stein, V. Dragalin, B. Gaydos, M. Krams, *et al.*, "Adaptive designs in clinical drug development—an executive summary of the phrma working group", *Journal of Biopharmaceutical Statistics*, vol. 16, no. 3, pp. 275–283, 2006.

[3] F. Bretz, F. Koenig, W. Brannath, E. Glimm, and M. Posch, "Adaptive designs for confirmatory clinical trials", *Statistics in Medicine*, vol. 28, no. 8, pp. 1181–1217, 2009.

[4] M. F. Drummond, M. J. Sculpher, K. Claxton, G. L. Stoddart, G. W. Torrance, *et al.*, *Methods for the economic evaluation of health care programmes*. Oxford University Press, 2015.

[5] National Health Service. (2016). About the NHS: Principles and values that guide the NHS, [Online]. Available: http://www.nhs.uk/NHSEngland/thenhs/about/Pages/nhscoreprinciples.aspx (visited on 12/22/2016).

[6] National Institute for Health and Care Excellence. (2014). What We Do, [Online]. Available: https://www.nice.org.uk/ (visited on 09/12/2019).

[7] ——, (2013). Guide to the Methods of Technology Appraisal, [Online]. Available: http://www.nice.org.uk/article/pmg9/chapter/foreword (visited on 07/12/2017).

[8] National Institute for Health Research. (2017). Funding opportunities for research and career development, [Online]. Available: https://www.nihr.ac.uk/funding-and-support/documents/NIHR\%20Funding\%20Opportunities\%20Booklet\%202017.pdf (visited on 08/07/2018).

[9] ——, (2017). General assessment criteria, [Online]. Available: https://www.nihr.ac.uk/funding-and-support/documents/funding-for-research-studies/how-to-apply/general-assessment-criteria.pdf (visited on 08/07/2018).

[10] G. H. Guyatt, D. L. Sackett, J. C. Sinclair, R. Hayward, D. J. Cook, *et al.*, "Users' guides to the medical literature: Ix. a method for grading health care recommendations", *JAMA*, vol. 274, no. 22, pp. 1800–1804, 1995.

[11] Committee for Medicinal Products for Human Use. (2006). ICH Topic E 9: Statistical Principles for Clinical Trials, [Online]. Available: https://www.ema.europa.eu/en/ich-e9-statistical-principles-clinical-trials (visited on 03/26/2020).

[12] S. Staniszewska, J. Brett, I. Simera, K. Seers, C. Mockford, *et al.*, "Gripp2 reporting checklists: Tools to improve reporting of patient and public involvement in research", *Research involvement and engagement*, vol. 3, no. 1, p. 13, 2017.

[13] S. C. Chow and R. Corey, "Benefits, challenges and obstacles of adaptive clinical trial designs", *Orphanet Journal of Rare Diseases*, vol. 6, no. 1, p. 79, 2011.

[14] P. Pallmann, A. W. Bedding, B. Choodari-Oskooei, M. Dimairo, L. Flight, *et al.*, "Adaptive designs in clinical trials: Why use them, and how to run and report them", *BMC Medicine*, vol. 16, no. 1, p. 29, 2018.

[15] S. C. Chow and M. Chang, *Adaptive design methods in clinical trials. Second Edition*. CRC Press, 2012.

[16] M. Dimairo, P. Pallmann, J. Wason, S. Todd, T. Jaki, *et al.*, "The adaptive designs consort extension (ace) statement: A checklist with explanation and elaboration guideline for reporting randomised trials that use an adaptive design", *BMJ*, 2019.

[17] S. C. Chow, R. Corey, and M. Lin, "On the independence of data monitoring committee in adaptive design clinical trials.", *Journal of Biopharmaceutical Statistics*, vol. 22, no. 4, pp. 853–67, 2012.

[18] S. Goodacre, M. Bradburn, P Fitzgerald, E. Cross, P.Collinson, *et al.*, "The RATPAC Randomised Assessment of Treatment using Panel Assay of Cardiac markers Trial: A Randomised Controlled Trial of Point-of-care Cardiac Markers in the Emergency Department", *Heart*, vol. 15, no. 3, pp. 190–196, 2011.

[19] P. Fitzgerald, S. W. Goodacre, E. Cross, and S. Dixon, "Cost-effectiveness of point-of-care biomarker assessment for suspected myocardial infarction: The randomized assessment of treatment using panel assay of cardiac markers (RATPAC) trial", *Academic Emergency Medicine*, vol. 18, no. 5, pp. 488–495, 2011.

[20] L Sutton, S. A. Julious, and S. W. Goodacre, "Influence of Adaptive Analysis on Unnecessary Patient Recruitment: Reanalysis of the RATPAC Trial", *Annals of Emergency Medicine*, vol. 60, pp. 442–448, 2012.

[21] A. Stevely, M. Dimairo, S. Todd, S. A. Julious, J. Nicholl, *et al.*, "An investigation of the shortcomings of the consort 2010 statement for the reporting of group sequential randomised controlled trials: A methodological systematic review", *PloS One*, vol. 10, no. 11, 2015.

[22] J. M. Wason, P. Brocklehurst, and C. Yap, "When to keep it simple–adaptive designs are not always useful", *BMC Medicine*, vol. 17, no. 1, pp. 1–7, 2019.

[23] S. J. Pocock, *Clinical trials: a practical approach*. Wiley, 1983.

[24] A. Hartford, M. Thomann, X. Chen, E. Miller, A. Bedding, *et al.*, "Adaptive designs: Results of 2016 survey on perception and use", *Therapeutic Innovation & Regulatory Science*, pp. 1–13, 2018.

[25] M. Dimairo, J. Boote, S. A. Julious, J. P. Nicholl, and S. Todd, "Missing steps in a staircase: a qualitative study of the perspectives of key stakeholders on the use of adaptive designs in confirmatory trials", *Trials*, vol. 16, 2015.

[26] I. Hatfield, A. Allison, L. Flight, S. A. Julious, and M. Dimairo, "Adaptive designs undertaken in clinical research: A review of registered clinical trials", *Trials*, vol. 17, no. 1, p. 150, 2016.

[27] P. Mistry, J. A. Dunn, and A. Marshall, "A literature review of applied adaptive design methodology within the field of oncology in randomised controlled trials and a proposed extension to the consort guidelines", *BMC Medical Research Methodology*, vol. 17, no. 1, p. 108, 2017.

[28] L. E. Bothwell, J. Avorn, N. F. Khan, and A. S. Kesselheim, "Adaptive design clinical trials: A review of the literature and clinicaltrials. gov", *BMJ Open*, vol. 8, no. 2, e018320, 2018.

[29] W. J. Meurer, L Legocki, S Mawocha, S. M. Frederiksen, T. C. Guetterman, *et al.*, "Attitudes and opinions regarding confirmatory adaptive clinical trials: a mixed methods analysis from the Adaptive Designs Accelerating Promising Trials into Treatments (ADAPT-IT) project", *Trials*, vol. 17, no. 1, p. 373, 2016.

[30] J. Quinlan, B. Gaydos, J. Maca, and M. Krams, "Barriers and opportunities for implementation of adaptive designs in pharmaceutical product development", *Clinical Trials*, vol. 7, no. 2, pp. 167–173, 2010.

[31] C. S. Coffey, B Levin, C Clark, C Timmerman, J Wittes, *et al.*, "Overview, hurdles, and future work in adaptive designs: perspectives from a National Institutes of Health-funded workshop", *Clinical Trials*, vol. 9, no. 6, pp. 671–680, 2012.

[32] J. Kairalla, C Coffey, M. Thomann, A. M. Thomann, and K. E. Muller, "Adaptive Trial Designs: A Review of Barriers and Opportunities", *Trials*, vol. 13, p. 145, 2012.

[33] J. M. Morgan, B. D. Dimitrov, J. Gill, S. Kitt, G. Andre Ng, *et al.*, "Rationale and study design of the REM-HF study: Remote management of heart failure using implanted devices and formalized follow-up procedures", *European Journal of Heart Failure*, vol. 16, no. 9, pp. 1039–1045, 2014.

[34] U.S. Food and Drug Administration. (2019). Guidance for Industry: Adaptive Design Clinical Trials for Drugs and Biologics., [Online]. Available: http://www.fda.gov/downloads/Drugs/Guidances/ucm201790.pdf (visited on 03/26/2020).

[35] R. B. North, K. Kumar, M. S. Wallace, J. M. Henderson, J. Shipley, *et al.*, "Spinal cord stimulation versus re-operation in patients with failed Back Surgery Syndrome: An international multicenter randomized controlled trial (EVIDENCE Study)", *Neuromodulation*, vol. 14, no. 4, pp. 330–335, 2011.

[36] P Warner, C. Weir, C. Hansen, A Douglas, M Madhra, *et al.*, "Low-dose dexamethasone as a treatment for women with heavy menstrual bleeding: Protocol for response-adaptive randomised placebo-controlled dose-finding parallel group trial (dexfem)", *BMJ Open*, vol. 5, no. 1, e006837, 2015.

[37] M. R. Sydes, M. K. Parmar, M. D. Mason, N. W. Clarke, C. Amos, *et al.*, "Flexible trial design in practice-stopping arms for lack-of-benefit and adding research arms mid-trial

in stampede: A multi-arm multi-stage randomized controlled trial", *Trials*, vol. 13, no. 1, p. 168, 2012.

[38] C. Jennison and B. W. Turnbull, *Group sequential methods with applications to clinical trials*. Chapman and Hall/CRC, 2000.

[39] S. S. Emerson. (2012). Boundary Scales in RCTdesign, (visited on 03/26/2020).

[40] W Tröger, D Galun, M Reif, A Schumann, N Stanković, *et al.*, "Viscum album [l.] extract therapy in patients with locally advanced or metastatic pancreatic cancer: A randomised clinical trial on overall survival", *European Journal of Cancer*, vol. 49, no. 18, pp. 3788–3797, 2013.

[41] S. J. Pocock, "Group sequential methods in the design and analysis of clinical trials", *Biometrika*, vol. 64, pp. 191–199, 1977.

[42] P. C. O'Brien and T. R. Fleming, "A multiple testing procedure for clinical trials", *biometrics*, vol. 35, pp. 549–556, 1979.

[43] J. Whitehead, *The design and analysis of sequential clinical trials*. Wiley, 1997.

[44] J. J. Zhang, G. Blumenthal, K. He, S. Tang, P. Cortazar, *et al.*, "Overestimation of the effect size in group sequential trials", *Clinical Cancer Research*, clincanres–3118, 2012.

[45] S. S. Ellenberg, D. L. DeMets, and T. R. Fleming, "Bias and trials stopped early for benefit", *JAMA*, vol. 304, no. 2, pp. 156–159, 2010.

[46] T. D. V. Swinscow, *Statistics at Square One*. John Wiley and Sons, 1997, pp. 1–158.

[47] J. Pinheiro, "Estimating and reducing bias in group sequential designs with Gaussian independent increment structure", *Biometrika*, vol. 84, no. 4, pp. 831–845, 1997.

[48] S. S. Emerson and T. R. Fleming, "Parameter estimation following group sequential hypothesis testing", *Biometrika*, vol. 77, no. 4, pp. 875–892, 1990.

[49] J. Whitehead, "On the bias of maximum likelihood estimation following a sequential test", *Biometrika*, vol. 73, no. 3, pp. 573–581, 1986.

[50] ——, "Supplementary analysis at the conclusion of a sequential clinical trial", *Biometrics*, pp. 461–471, 1986.

[51] M. D. Smith. (2020). Newton-raphson technique, [Online]. Available: https://web.mit.edu/10.001/Web/Course_Notes/NLAE/node6.html (visited on 04/14/2020).

[52] P. Armitage, C. McPherson, and B. Rowe, "Repeated significance tests on accumulating data", *Journal of the Royal Statistical Society. Series A (General)*, pp. 235–244, 1969.

[53] Y. Yan, and S. Coad. (2009). Estimation of secondary parameters following sequential tests for multivariate data, [Online]. Available: http://mat.izt.uam.mx/profs/anovikov/data/IWSM2009/contributed\%20papers/IWSM83.pdf (visited on 05/08/2018).

[54] C. Jennison and B. Turnbull. (2015). The design of group sequential clinical trials that test multiple endpoints, [Online]. Available: http://people.bath.ac.uk/mascj/talks_2015/cj_takeda_gsts_and_multiple_testing.pdf (visited on 05/08/2018).

[55] A Briggs, K Claxton, and M Schulpher, *Decision Modelling for Health Economic Evaluation*. Oxford: Oxford Univ. Press, 2006.

[56] M Chang, *Adaptive design theory and implementation using SAS and R*. CRC Press, 2014.

[57] G. Wassmer and W. Brannath, *Group sequential and confirmatory adaptive designs in clinical trials*. Springer, 2016.

[58] A. A. Tsiatis, G. L. Rosner, and C. R. Mehta, "Exact confidence intervals following a group sequential test", *Biometrics*, pp. 797–803, 1984.

[59] M. N. Chang and P. C. O'Brien, "Confidence intervals following group sequential tests", *Controlled Clinical Trials*, vol. 7, no. 1, pp. 18–26, 1986.

[60] T. M. Skalland, "An evaluation of design and inference in special topics of group sequential procedures", PhD thesis, Oregon State University, 2015.

[61] A. M. Jones, Ed., *The Elgar Companion to Health Economics*. Edward Elgar, 2006.

[62] Pharmaceutical Benefits Advisory Committee, *Guidelines for preparing submissions to the Pharmaceutical Benefits Advisory Committee. Version 4.3*. June. 2008, p. 350.

[63] Canadian Agency for Drugs and Technologies in Health. (2017). Guidelines for the Economic Evaluation of Health Technologies: Canada, [Online]. Available: https://www.cadth.ca/about-cadth/how-we-do-it/methods-and-guidelines/guidelines-for-the-economic-evaluation-of-health-technologies-canada.

[64] M. Drummond, A. Brandt, B. Luce, and J. Rovira, "Standardizing methodologies for economic evaluation in health care: Practice, problems, and potential", *International Journal of Technology Assessment in Health Care*, vol. 9, no. 1, pp. 26–36, 1993.

[65] L. Curtis and A. Burns. (2015). Unit Costs of Health & Social Care 2015, [Online]. Available: https://www.pssru.ac.uk/project-pages/unit-costs/unit-costs-2015/ (visited on 03/26/2020).

[66] Joint Formulary Committee. (2016). British national formulary, [Online]. Available: http://www.medicinescomplete.com (visited on 12/26/2016).

[67] S. J. Whitehead and S. Ali, "Health outcomes in economic evaluation: The QALY and utilities", *British Medical Bulletin*, vol. 96, no. 1, pp. 5–21, 2010.

[68] D. L. Patrick, H. E. Starks, K. C. Cain, R. F. Uhlmann, and R. A. Pearlman, "Measuring preferences for health states worse than death", *Medical Decision Making*, vol. 14, no. 1, pp. 9–18, 1994.

[69] M. Herdman, C. Gudex, A. Lloyd, M. Janssen, P. Kind, *et al.*, "Development and preliminary testing of the new five-level version of eq-5d (eq-5d-5l)", *Quality of Life Research*, vol. 20, no. 10, pp. 1727–1736, 2011.

[70] B. Van Hout, M. Janssen, Y.-S. Feng, T. Kohlmann, J. Busschbach, *et al.*, "Interim scoring for the EQ-5D-5L: mapping the EQ-5D-5L to EQ-5D-3L value sets", *Value in Health*, vol. 15, no. 5, pp. 708–715, 2012.

[71] National Institute for Health and Care Excellence. (2013). Position statement on use of the EQ-5D-5L valuation set for england, [Online]. Available: https://www.nice.org.uk/about/what-we-do/our-programmes/nice-guidance/technology-appraisal-guidance/eq-5d-5l (visited on 09/03/2019).

[72] A. H. Briggs, B. J. O'Brien, and G. Blackhouse, "Thinking outside the box: recent advances in the analysis and presentation of uncertainty in cost-effectiveness studies.", *Annual Review of Public Health*, vol. 23, pp. 377–401, 2002.

[73] E Fenwick, K Claxton, M Sculpher, and H. E. Letters, "Representing uncertainty: the role of cost-effectiveness acceptability curves", *Health Economics*, vol. 787, no. May, pp. 779–787, 2001.

[74] B O'Brien, "Economic Evaluation of Pharmaceuticals: Frankenstein's Monster or Vampire of Trials?", *Medical Care*, vol. 34, no. 12, 1996.

[75] S. Ramsey, R. Willke, H. Glick, S. D. Reed, A. Federico, *et al.*, "Cost-Effectiveness Analysis Alongside Clinical Trials II - An ISPOR Good Research Practives Task Force Report", *Value in Health*, vol. 18, pp. 161–172, 2015.

[76] M. J. Buxton, M. F. Drummond, B. a. Van Hout, R. L. Prince, T. a. Sheldon, *et al.*, "Modelling in economic evaluation: An unavoidable fact of life", *Health Economics*, vol. 6, no. 3, pp. 217–227, 1997.

[77] A. H. Briggs and A. M. Gray, "Power and sample size calculations for stochastic cost-effectiveness analysis", *Medical Decision Making*, vol. 18, no. 2, S81–S92, 1998.

[78] A. Brennan and R. Akehurst, "Modelling in health economic evaluation", *PharmacoEconomics*, vol. 17, no. 5, pp. 445–459, 2000.

[79] N. J. Welton, A. J. Sutton, N. Cooper, K. R. Abrams, and A. E. Ades, *Evidence Synthesis for Decision Making in Healthcare*. Wiley, 2012.

[80] H Raiffa and R Schlaiffer, *Applied Statistical Decision Theory*. John Wiley & Sons, Ltd, 2000.

[81] M. J. Sculpher, K. Claxton, M. Drummond, and C. McCabe, "Whither trial-based economic evaluation for health care decision making?", *Health Economics*, vol. 15, no. 7, pp. 677–687, 2006.

[82] A. Brennan, S. E. Chick, and R. Davies, "A taxonomy of model structures for economic evaluation of health technologies", *Health Economics*, vol. 15, no. 12, pp. 1295–1310, 2006.

[83] N. J. Welton, A. E. Ades, D. M. Caldwell, and T. J. Peters, "Research prioritization based on expected value of partial perfect information: A case-study on interventions to increase uptake of breast cancer screening", *Journal of the Royal Statistical Society. Series A: Statistics in Society*, vol. 171, no. 4, pp. 807–834, 2008.

[84] K Claxton, "Exploring Uncertainty in Cost-effectiveness Analysis", *PharmacoEconomics*, vol. 26, no. 9, pp. 781–798, 2008.

[85] A Saltelli, K Chan, and E. M. Scott, *Sensitivity Analysis*. Wiley, 2000.

[86] B. Efron and R. Tibshirani, "Bootstrap methods for standard errors, confidence intervals, and other measures of statistical accuracy", *Statistical Science*, pp. 54–75, 1986.

[87] B. A. Van Hout, M. J. Al, G. S. Gordon, and F. F. Rutten, "Costs, effects and ce-ratios alongside a clinical trial", *Health Economics*, vol. 3, no. 5, pp. 309–319, 1994.

[88] E. Fenwick, L. Steuten, S. Knies, S. Ghabri, A. Basu, *et al.*, "Value of Information Analysis for Research Decisions—An Introduction: Report 1 of the ISPOR Value of Information Analysis Emerging Good Practices Task Force", *Value in Health*, vol. 23, no. 2, pp. 139–150, 2020.

[89] K. Claxton and J. Posnett, "An economic approach to clinical trial design and research priority-setting", *Health economics*, vol. 5, no. 6, pp. 513–524, 1996.

[90] J. Chilcott, A. Brennan, A. Booth, J. Karnon, and P. Tappenden, "The role of modelling in prioritising and planning clinical trials", *Health Technology Assessment*, vol. 7, no. 23, 2003.

[91] K. Claxton and M. Sculpher, "Using value of information analysis to prioritise health research: some lessons from recent UK experience.", *PharmacoEconomics*, vol. 24, no. 11, pp. 1055–1068, 2006.

[92] S. W. Hee, T. Hamborg, S. Day, J. Madan, F. Miller, *et al.*, "Decision-theoretic designs for small trials and pilot studies: A review", *Statistical Methods in Medical Research*, vol. 25, no. 3, pp. 1022–1038, 2016.

[93] M. H. Chen and A. R. Willan, "Value of information methods for assessing a new diagnostic test", *Statistics in Medicine*, vol. 33, no. 11, pp. 1801–1815, 2014.

[94] C. Rothery, M. Strong, H. E. Koffijberg, A. Basu, S. Ghabri, *et al.*, "Value of information analytical methods: Report 2 of the ispor value of information analysis emerging good practices task force", *Value in Health*, vol. 23, no. 3, pp. 277–286, 2020.

[95] M. Strong, J. E. Oakley, A. Brennan, and P. Breeze, "Estimating the expected value of sample information using the probabilistic sensitivity analysis sample: A fast, nonparametric regression-based method.", *Medical Decision Making*, vol. 35, no. 5, pp. 570–583, 2015.

[96] A. E. Ades, G Lu, and K Claxton, "Expected Value of Sample Information Calculations in Medical Decision Modelling", *Medical Decision Making*, vol. 24, p. 207, 2004.

[97] S Griffin, N. J. Welton, and K Claxton, "Exploring the Research Design Space: The Expected Value of Information for Sequential Research Designs", *Medical Decision Making*, vol. 30, p. 155, 2010.

[98] K. A. Boyd, E. Fenwick, and A. Briggs, "Using an iterative approach to economic evaluation in the drug development process", *Drug Development Research*, vol. 71, no. 8, pp. 470–477, 2010.

[99] A. Brennan, S. Kharroubi, A. O. Hagan, and J. Chilcott, "Calculating Partial Expected Value of Perfect Information via Monte Carlo Sampling Algorithms", *Medical Decision Making*, no. 27, pp. 448–470, 2007.

[100] D. A. Berry, "Decision analysis and Bayesian methods in clinical trials", in *Recent Advances in Clinical Trial Design and Analysis*, Springer, 1995, pp. 125–154.

[101] N. J. Welton, J. J. Madan, D. M. Caldwell, T. J. Peters, and a. E. Ades, "Expected Value of Sample Information for Multi-Arm Cluster Randomized Trials with Binary Outcomes.", *Medical Decision Making [Online]*, vol. 34, no. April, pp. 352–65, 2013.

[102] N. A. Menzies, "An efficient estimator for the expected value of sample information", *Medical Decision Making*, vol. 36, no. 3, pp. 308–320, 2016.

[103] H. Jalal, J. D. Goldhaber-Fiebert, and K. M. Kuntz, "Computing expected value of partial sample information from probabilistic sensitivity analysis using linear regression metamodeling", *Medical Decision Making*, vol. 35, no. 5, pp. 584–595, 2015.

[104] H. Jalal and F. Alarid-Escudero, "A gaussian approximation approach for value of information analysis", *Medical Decision Making*, vol. 38, no. 2, pp. 174–188, 2018.

[105] A. Heath, I. Manolopoulou, and G. Baio, "Efficient monte carlo estimation of the expected value of sample information using moment matching", *Medical Decision Making*, vol. 38, no. 2, pp. 163–173, 2018.

[106] S. Mohiuddin, E. Fenwick, and K. Payne, "Use of value of information in UK health technology assessments", *International Journal of Technology Assessment in Health Care*, vol. 30, no. 6, pp. 553–570, 2014.

[107] N. J. Welton and H. H. Z. Thom, "Value of information: We've got speed, what more do we need?", *Medical Decision Making*, vol. 35, no. 5, pp. 564–566, 2015, PMID: 25840903.

[108] L. Steuten, G. van de Wetering, K. Groothuis-Oudshoorn, and V. Retèl, "A systematic and critical review of the evolving methods and applications of value of information in academia and practice", *PharmacoEconomics*, vol. 31, no. 1, pp. 25–48, 2013.

[109] A. Heath, N. R. Kunst, C. Jackson, M. Strong, F. Alarid-Escudero, *et al.*, "Calculating the expected value of sample information in practice: Considerations from three case studies", *arXiv preprint arXiv:1905.12013*, 2019.

[110] N. R. Kunst, E. Wilson, F. Alarid-Escudero, G. Baio, A. Brennan, *et al.*, "Computing the expected value of sample information efficiently: Expertise and skills required for four model-based methods", *arXiv preprint arXiv:1910.03368*, 2019.

[111] ISPOR. (2020). Value of information analysis for research decisions task force, [Online]. Available: https://www.ispor.org/member-groups/task-forces/value-of-information-analysis-for-research-decisions (visited on 03/05/2020).

[112] D. Glynn, C. Rothery, and K. Claxton. (2020). Rapid assessment of need for evidence, [Online]. Available: https://shiny.york.ac.uk/rane/ (visited on 02/24/2020).

[113] M. Strong, P. Breeze, C. Thomas, A. Brennan, and C. Stevens. (2020). Sheffield accelerated value of information, [Online]. Available: http://savi.shef.ac.uk/SAVI/ (visited on 02/24/2020).

[114] R. J. Lewis, A. M. Lipsky, and D. A. Berry, "Bayesian decision-theoretic group sequential clinical trial design based on a quadratic loss function: a frequentist evaluation.", *Clinical Trials (London, England)*, vol. 4, no. 1, pp. 5–14, 2007.

[115] D. A. Berry, N. Carolina, M. C. Wolff, and D. Sack, "Decision Making during a Phase III Randomized Controlled Trial", *Contemporary Clinical Trials*, vol. 15, no. 5, pp. 360–378, 1994.

[116] D. A. Berry and C.-H. Ho, "One-sided sequential stopping boundaries for clinical trials: A decision-theoretic approach", *Biometrics*, pp. 219–227, 1988.

[117] N. Cressie and P. B. Morgan, "The VPRT: A sequential testing procedure dominating the SPRT", *Econometric Theory*, vol. 9, no. 3, pp. 431–450, 1993.

[118] N. Cressie and J. Biele, "A sample-size-optimal bayesian procedure for sequential pharmaceutical trials", *Biometrics*, pp. 700–711, 1994.

[119] B. P. Carlin, J. B. Kadane, and A. E. Gelfand, "Approaches for optimal sequential decision analysis in clinical trials", *Biometrics*, pp. 964–975, 1998.

[120] J. B. Kadane and P. K. Vlachos, "Hybrid methods for calculating optimal few-stage sequential strategies: Data monitoring for a clinical trial", *Statistics and Computing*, vol. 12, no. 2, pp. 147–152, 2002.

[121] A. E. Brockwell and J. B. Kadane, "A gridding method for bayesian sequential decision problems", *Journal of Computational and Graphical Statistics*, vol. 12, no. 3, pp. 566–584, 2003.

[122] J. K. Wathen and J. A. Christen, "Implementation of backward induction for sequentially adaptive clinical trials", *Journal of Computational and Graphical Statistics*, vol. 15, no. 2, pp. 398–413, 2006.

[123] L. A. Orawo and J. A. Christen, "Bayesian sequential analysis for multiple-arm clinical trials", *Statistics and Computing*, vol. 19, no. 1, pp. 99–109, 2009.

[124] A. R. Willan and E. M. Pinto, "The value of information and optimal clinical trial design", *Statistics in Medicine*, vol. 24, no. 12, pp. 1791–1806, 2005.

[125] A. R. Willan and M Kowgier, "Determining optimal sample sizes for multi-stage randomized clinical trials using value of information methods", *Clinical Trials*, vol. 5, pp. 289–300, 2008.

[126] C. R. Mehta and N. R. Patel, "Adaptive, group sequential and decision theoretic approaches to sample size determination", *Statistics in Medicine*, vol. 25, no. 19, pp. 3250–3269, 2006.

[127] P. Pertile, M. Forster, and D. L. Torre, "Optimal Bayesian sequential sampling rules for the economic evaluation of health technologies", *Journal of the Royal Statistical Society. Series A: Statistics in Society*, vol. 177, no. 2, pp. 419–438, 2014.

[128] H. Chernoff and A. J. Petkau, "Sequential medical trials involving paired data", *Biometrika*, vol. 68, no. 1, pp. 119–132, 1981.

[129] S. Chick, M. Forster, and P. Pertile, "A bayesian decision theoretic model of sequential experimentation with delayed response", *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, vol. 79, no. 5, pp. 1439–1462, 2017.

[130] M Forster, S Brealey, S Chick, A Keding, B Corbacho, *et al.* (2019). Cost-Effective Clinical Trial Design: Application of a Bayesian Sequential Stopping Rule to the ProFHER Pragmatic Trial, [Online]. Available: https://ideas.repec.org/p/yor/yorken/19-01.html (visited on 03/26/2020).

[131] H. Handoll, S. Brealey, A. Rangan, A. Keding, B. Corbacho, *et al.*, "The ProFHER (PROximal Fracture of the Humerus: Evaluation by Randomisation) trial-a pragmatic multicentre randomised controlled trial evaluating the clinical effectiveness and cost-effectiveness of surgical compared with non-surgical treatment for proximal fracture of the humerus in adults.", *Health Technology Assessment*, vol. 19, no. 24, p. 1, 2015.

[132] P. Kouvelis, J. Milner, and Z. Tian, "Clinical trials for new drug development: Optimal investment and application", *Manufacturing & Service Operations Management*, vol. 19, no. 3, pp. 437–452, 2017.

[133] J. J. Thijssen and D. Bregantini, "Costly sequential experimentation and project valuation with an application to health technology assessment", *Journal of Economic Dynamics and Control*, vol. 77, pp. 202–229, 2017.

[134] I. C. Marschner and I. M. Schou, "Underestimation of treatment effects in sequentially monitored clinical trials that did not stop early for benefit", *Statistical Methods in Medical Research*, vol. 28, no. 10-11, pp. 3027–3041, 2019.

[135] Gusto Investigators, "An international randomized trial comparing four thrombolytic strategies for acute myocardial infarction", *New England Journal of Medicine*, vol. 329, no. 10, pp. 673–682, 1993.

[136] D. B. Mark, M. A. Hlatky, R. M. Califf, C. D. Naylor, K. L. Lee, *et al.*, "Cost effectiveness of thrombolytic therapy with tissue plasminogen activator as compared with streptokinase for acute myocardial infarction", *New England Journal of Medicine*, vol. 332, no. 21, pp. 1418–1424, 1995.

[137] C. C. Morgan, S. Huyck, M. Jenkins, L. Chen, A. Bedding, *et al.*, "Adaptive Design: Results of 2012 Survey on Perception and Use", *Therapeutic Innovation and Regulatory Science*, vol. 48, no. 4, pp. 473–481, 2014.

[138] L. Flight, F. Arshad, R. Barnsley, K. Patel, S. Julious, *et al.*, "A review of clinical trials with an adaptive design and health economic analysis", *Value in Health*, vol. 22, no. 4, pp. 391–398, 2019.

[139] clinicaltrials.gov. (2017). Background, [Online]. Available: https://clinicaltrials.gov/ct2/about-site/background (visited on 07/12/2017).

[140] D. Moher, K. F. Schulz, D. G. Altman, C. Group, and Others, "The CONSORT statement: revised recommendations for improving the quality of reports of parallel-group randomised trials", *The Lancet*, vol. 357, no. 9263, pp. 1191–1194, 2001.

[141] I. Shemilt, M. Mugford, M. Drummond, E. Eisenstein, J. Mallender, *et al.*, "Economics methods in Cochrane systematic reviews of health promotion and public health related interventions.", *BMC Medical Research Methodology*, vol. 6, p. 55, 2006.

[142] SAACTD Workshop Committee. (2009). Connecting Non-Profits to Adaptive Clinical Trial Designs: Themes and Recommendations from the Scientific Advances in Adaptive Clinical Trial Designs Workshop., [Online]. Available: https://custom.cvent.com/536726184EFD40129EF286585E55929F/files/2627e73646ce4733a2c03692fab26fff.pdf (visited on 02/21/2019).

[143] M. A. Detry, R. J. Lewis, K. R. Broglio, J. T. Connor, S. M. Berry, *et al.* (2012). Standards for the Design, Conduct, and Evaluation of Adaptive Randomized Clinical Trials, [Online]. Available: https://www.pcori.org/assets/Standards-for-the-Design-Conduct-and-Evaluation-of-Adaptive-Randomized-Clinical-Trials.pdf (visited on 03/26/2020).

[144] D. Husereau, M. Drummond, S. Petrou, C. Carswell, D. Moher, *et al.*, "Consolidated Health Economic Evaluation Reporting Standards (CHEERS) statement", *European Journal of Health Economics*, vol. 14, no. 3, pp. 367–372, 2013.

[145] C. Begg, M. Cho, S. Eastwood, R. Horton, D. Moher, *et al.*, "Improving the quality of reporting of randomized controlled trials: the CONSORT statement", *JAMA*, vol. 276, no. 8, pp. 637–639, 1996.

[146] K. F. Schulz, D. G. Altman, and D. Moher, "CONSORT 2010 statement: updated guidelines for reporting parallel group randomised trials", *BMC Medicine*, vol. 8, no. 1, p. 1, 2010.

[147] ISPOR. (2016). Pharmacoeconomic Guidelines Around The World, [Online]. Available: http://www.ispor.org/PEguidelines/index.asp (visited on 11/14/2016).

[148] A. Briggs, M. C. Weinstein, and E Fenwick, "Model parameter estimation and uncertainty analysis: A report of the ISPOR-SMDM modeling good research practices task force-6", *Value in Health*, vol. 15, pp. 835–842, 2012.

[149] European Medicines Agency. (2007). Reflection Paper on Methodological Issues in Confirmatory Clinical Trials Planned With an Adaptive Design, [Online]. Available: https://www.ema.europa.eu/en/documents/scientific-guideline/reflection-paper-methodological-issues-confirmatory-clinical-trials-planned-adaptive-design_en.pdf (visited on 03/26/2020).

[150] E. Bartha, T. Davidson, A. Hommel, K.-G. Thorngren, P. Carlsson, *et al.*, "Cost-effectiveness analysis of goal-directed hemodynamic treatment of elderly hip fracture patientsbefore-clinical research starts", *Anesthesiology: The Journal of the American Society of Anesthesiologists*, vol. 117, no. 3, pp. 519–530, 2012.

[151] E Bartha, C Arfwedson, A Imnell, M. Fernlund, L. Andersson, *et al.*, "Randomized controlled trial of goal-directed haemodynamic treatment in patients with proximal femoral fracture", *British Journal of Anaesthesia*, vol. 110, no. 4, pp. 545–553, 2013.

[152] E. Bartha, T. Davidson, T.-H. Brodtkorb, P. Carlsson, and S. Kalman, "Value of information: Interim analysis of a randomized, controlled trial of goal-directed hemodynamic treatment for aged patients", *Trials*, vol. 14, no. 1, pp. 1–10, 2013.

[153] S. Brown, I. L. Smith, J. M. Brown, C. Hulme, E. McGinnis, *et al.*, "Pressure relieving support surfaces: A randomised evaluation 2 (pressure 2): Study protocol for a randomised controlled trial", *Trials*, vol. 17, no. 1, p. 604, 2016.

[154] P. S. Hall, C. McCabe, R. C. Stein, and D. Cameron, "Economic evaluation of genomic test-directed chemotherapy for early-stage lymph node-positive breast cancer", *Journal of the National Cancer Institute*, vol. 104, no. 1, pp. 56–66, 2012.

[155] P. Hall, A. Smith, C. Hulme, A. Vargas-Palacios, J. Dunn, *et al.*, "The use of early decision modelling and value of information analysis in an adaptive trial design: results from the OPTIMA preliminary study", *Trials*, vol. 16, no. Suppl 2, O19, 2015.

[156] R. C. Stein, J. A. Dunn, J. M. S. Bartlett, A. F. Campbell, A. Marshall, *et al.*, "OPTIMA prelim: A randomised feasibility study of personalised care in the treatment of women with early breast cancer", *Health Technology Assessment*, vol. 20, no. 10, p. 201, 2016.

[157] R. C. Stein. (2016). OPTIMA Personalised Treatment of Breast Cancer Trial Protocol, [Online]. Available: https://www.journalslibrary.nihr.ac.uk/programmes/hta/1034501/#/ (visited on 03/26/2020).

[158] H. Earl, "Duration of Trastuzumab with Chemotherapy in Women with Early Breast Cancer: Six Months versus Twelve.", Tech. Rep., 2009.

[159] N. Tessitore, V. Bedogna, A. Poli, G. Lipari, P. Pertile, *et al.*, "Should current criteria for detecting and repairing arteriovenous fistula stenosis be reconsidered? Interim analysis of a randomized controlled trial", *Nephrology Dialysis Transplantation*, vol. 29, no. 1, pp. 179–187, 2014.

[160] Academy of Managed Care Pharmacy. (2016). A Format for Submission of Clinical and Economic Evidence of Pharmaceuticals in Support of Formulary Consideration, [Online]. Available: http://www.amcp.org/sites/default/files/2019-03/AMCP-Format-V4.pdf (visited on 03/26/2020).

[161] Australian Government, Department of Health. (2016). Guidelines for preparing a submission to the Pharmaceutical Benefits Advisory Committee, [Online]. Available: https://pbac.pbs.gov.au/ (visited on 03/26/2020).

[162] Scottish Medicines Consortium, "Guidance to submitting companies for completion of New Product Assessment Form (NPAF)", Tech. Rep. June, 2019.

[163] P. S. Hall, A. Smith, C. Hulme, A. Vargas-Palacios, A. Makris, *et al.*, "Value of information analysis of multiparameter tests for chemotherapy in early breast cancer: The optima prelim trial", *Value in Health*, vol. 20, no. 10, pp. 1311–1318, 2017.

[164] J. Nixon, S. Brown, I. L. Smith, E. McGinnis, A. Vargas-Palacios, *et al.*, "Comparing alternating pressure mattresses and high-specification foam mattresses to prevent pressure ulcers in high-risk patients: The PRESSURE 2 RCT", *Health Technology Assessment*, vol. 23, no. 52, 2019.

[165] A. R. Willan, "Optimal sample size determinations from an industry perspective based on the expected value of information.", *Clinical trials (London, England)*, vol. 5, no. 6, pp. 587–94, 2008.

[166] J. Thorn, C. Ridyard, D. Hughes, S. Wordsworth, B. Mihaylova, *et al.*, "Health economics analysis plans: The current state of play", *Trials*, vol. 18, 2017.

[167] National Institute for Health and Care Excellence. (2020). Nice decision support unit, [Online]. Available: http://nicedsu.org.uk/ (visited on 04/05/2020).

[168] L. Flight, S. Julious, A. Brennan, S. Todd, and D. Hind, "How can health economics be used in the design and analysis of adaptive clinical trials? a qualitative analysis", *Trials*, vol. 21, no. 1, pp. 1–12, 2020.

[169] A. Tong, P. Sainsbury, and J. Craig, "Consolidated criteria for reporting qualitative research (coreq): A 32-item checklist for interviews and focus groups", *International journal for quality in health care*, vol. 19, no. 6, pp. 349–357, 2007.

[170] J. W. Creswell, *Research design: Qualitative, quantitative, and mixed methods approaches*. Sage Publications, 2013.

[171] B. J. O'Brien, M. F. Drummond, R. J. Labelle, and A. Willan, "In search of power and significance: Issues in the design and analysis of stochastic cost-effectiveness studies in health care", *Medical care*, pp. 150–163, 1994.

[172] A Briggs, "Economic evaluation and clinical trials: size matters.", *BMJ*, vol. 321, no. 7273, pp. 1362–1363, 2000.

[173] J Cook, M Drummond, and Heyse, "Multiple Primary Endpoints in Clinical Trials", *Statistical Methods in Medical Research*, vol. 13, pp. 157–176, 2004.

[174] D. Ashby and S.-b. Tanb, "Where ' s the utility of clinical trials ? Bayesian data-monitoring", *Clinical Trials*, vol. 2, pp. 197–208, 2005.

[175] B. G. Koerkamp, J. J. Nikken, E. H. Oei, and T. Stijnen, "Value of Information Analysis Used to Determine the Necessity of Additional Research: MR Imaging in Acute Knee Trauma as an Example", *Radiology*, vol. 246, no. 2, pp. 420–425, 2008.

[176] F Trotta, G Apolone, S Garattini, and G Tafuri, "Stopping a trial early in oncology: For patients or for industry?", *Annals of Oncology*, vol. 19, no. 7, pp. 1347–1353, 2008.

[177] B. G. Koerkamp, S. Spronk, T. Stijnen, and M. M. Hunink, "Value of information analyses of economic randomized controlled trials: The treatment of intermittent claudication", *Value in Health*, vol. 13, no. 2, pp. 242–250, 2010.

[178] W. Hollingworth, D. McKell-Redwood, L. Hampson, and C. Metcalfe, "Cost–utility analysis conducted alongside randomized controlled trials: Are economic end points considered in sample size calculations and does it matter?", *Clinical Trials*, vol. 10, no. 1, pp. 43–53, 2013.

[179] C. D. Mullins, J. Vandigo, Z. Zheng, and P. Wicks, "Patient-centeredness in the design of clinical trials", *Value in Health*, vol. 17, no. 4, pp. 471–475, 2014.

[180] H. W. Tuffaha, H Reynolds, L. G. Gordon, C. M. Rickard, and P Scuffham, "Value of information analysis optimizing future trial design from a pilot study on catheter securement devices.", *Clinical trials (London, England)*, vol. 11, no. 6, pp. 648–56, 2014.

[181] L. J. Legocki, W. J. Meurer, S. Frederiksen, R. J. Lewis, V. L. Durkalski, *et al.*, "Clinical trialist perspectives on the ethics of adaptive clinical trials: a mixed-methods analysis", *BMC Medical Ethics*, vol. 16, no. 1, p. 27, 2015.

[182] S. K. Husbands, "PhD Thesis, Model-Building Process", *University of Birmingham*, no. February, 2016.

[183] M. Dimairo, E. Coates, P. Pallmann, S. Todd, S. A. Julious, *et al.*, "Development process of a consensus-driven consort extension for randomised trials using an adaptive design", *BMC Medicine*, vol. 16, no. 1, p. 210, 2018.

[184] C Pope and N Mays, *Qualitative Research in Health Care*, Second Edi. BMJ, 2000.

[185] M Sandelowski, "Sample size in qualitative research", *Research in nursing and health*, vol. 18, no. 2, pp. 179–183, 1995.

[186] V Braun and V Clarke, *Successful qualitative research: A practical guide for beginners*. SAGE Publications Inc., 2013.

[187] A. J. B. Fugard and H. W. W. Potts, "Supporting thinking on sample sizes for thematic analyses: a quantitative tool", *International Journal of Social Research Methodology*, vol. 5579, no. July 2015, pp. 1–16, 2015.

[188] G. Guest, A. Bunce, and L. Johnson, "How Many Interviews Are Enough ? An Experiment with Data Saturation and Variability", *Family Health International*, vol. 18, no. 1, pp. 59–82, 2006.

[189] J. W. Creswell, W. E. Hanson, V. L. Clark Plano, and A Morales, "Qualitative Research Designs: Selection and Implementation", *The Counseling Psychologist*, vol. 35, no. 2, pp. 236–264, 2007.

[190] B. Carlsen and C. Glenton, "What about N? A methodological study of sample-size reporting in focus group studies", *BMC Medical Research Methodology*, vol. 11, no. 1, p. 26, 2011.

[191] J. J. Francis, M. Johnston, C. Robertson, L. Glidewell, V. Entwistle, *et al.*, "What is an adequate sample size? operationalising data saturation for theory-based interview studies", *Psychology and Health*, vol. 25, no. 10, pp. 1229–1245, 2010.

[192] J Ritchie and L Spencer, "Qualitative data analysis for applied policy research", *Analysing qualitative data*, pp. 173–194, 1993.

[193] N. K. Gale, G. Heath, E. Cameron, S. Rashid, and S. Redwood, "Using the framework method for the analysis of qualitative data in multi-disciplinary health research.", *BMC Medical Research Methodology*, vol. 13, no. 1, p. 117, 2013.

[194] T. L. Beauchamp and J. F. Childress, *Principles of biomedical ethics*. USA: Oxford University Press, 2001.

[195] J. Rawls, *A Theory of Justice.* Cambridge: The Belknap, 1971.

[196] N. Daniels, *Just health care*. Cambridge: Cambridge University Press, 1985.

[197] A. Williams, "Cost-effectiveness analysis: Is it ethical?", *Journal of Medical Ethics*, vol. 18, no. 1, pp. 7–11, 1992.

[198] NIHR Statistics Group. (2018). Linking independent statisticians and investigators of clinical trials, [Online]. Available: https://statistics-group.nihr.ac.uk/stat-link/ (visited on 10/15/2018).

[199] J. Bindels, B. Ramaekers, I. C. Ramos, L. Mohseninejad, S. Knies, *et al.*, "Use of value of information in healthcare decision making: Exploring multiple perspectives", *PharmacoEconomics*, vol. 34, no. 3, pp. 315–322, 2016.

[200] DAMOCLES Study Group, "A proposed charter for clinical trial data monitoring committees: Helping them to do their job well", *The Lancet*, vol. 365, no. 9460, pp. 711–722, 2005.

[201] P. Glasziou and I. Chalmers. (2015). How systematic reviews can reduce waste in research, [Online]. Available: https://blogs.bmj.com/bmj/2015/10/29/how-systematic-reviews-can-reduce-waste-in-research/ (visited on 12/23/2019).

[202] B. G. O. Sully, S. A. Julious, and J Nicholl, "An Investigation of the Impact of Futility Analysis in Publicly Funded Trials", *Trials*, vol. 15, p. 61, 2014.

[203] National Health Service. (2018). Aphasia, [Online]. Available: https://www.nhs.uk/conditions/aphasia/ (visited on 12/10/2018).

[204] Royal College of Speech and Language Therapists, *RCSLT resource manual for commissioning and planning services for SLCN: Dysphagia*, 2009.

[205] R. Palmer, P. Enderby, C. Cooper, N. Latimer, S. Julious, *et al.*, "Computer therapy compared with usual care for people with long-standing aphasia poststroke: A pilot randomized controlled trial", *Stroke*, vol. 43, no. 7, pp. 1904–1911, 2012.

[206] Steps Consultancy Ltd, *Stepbystep software*, 2018.

[207] N. R. Latimer, S. Dixon, and R. Palmer, "Cost-Utility of Self-Managed Computer Therapy for People With Aphasia", *International Journal of Technology Assessment in Health Care*, vol. 29, no. 04, pp. 402–409, 2013.

[208] R. Palmer, C. Cooper, P. Enderby, M. Brady, S. Julious, *et al.*, "Clinical and cost effectiveness of computer treatment for aphasia post stroke (Big CACTUS): study protocol for a randomised controlled trial.", *Trials*, vol. 16, no. 1, p. 18, 2015.

[209] R. Palmer, M. Dimairo, C. Cooper, P. Enderby, M. Brady, *et al.*, "Self-managed, computerised speech and language therapy for patients with chronic aphasia post-stroke compared with usual care or attention control (big cactus): A multicentre, single-blinded, randomised controlled trial.", *The Lancet. Neurology*, vol. 18, no. 9, pp. 821–833, 2019.

[210] H. Brønnum-Hansen, M. Davidsen, P. Thorvaldsen, *et al.*, "Long-term survival and causes of death after stroke", *Stroke*, vol. 32, no. 9, pp. 2131–2136, 2001.

[211] Office for National Statistics. (2018). Interim life table, united kingdom, based on data for years 2007-2009., [Online]. Available: http://webarchive.nationalarchives.gov.uk/20160107151332/http://www.ons.gov.uk/ons/taxonomy/search/index.html?newquery=*&newoffset=100&pageSize=50&nscl=Life+Tables&nscl-orig=Life+Tables&content-type=Reference+table&content-type=Dataset&sortBy=pubdate&sortDirection=DESCENDING&applyFilters=true (visited on 05/08/2018).

[212] R. Ara and J. E. Brazier, "Populating an economic model with health state utility values: Moving toward better practice", *Value in Health*, vol. 13, no. 5, pp. 509–518, 2010.

[213] L. Curtis and A. Burns. (2011). Unit Costs of Health & Social Care 2015, [Online]. Available: https://www.pssru.ac.uk/project-pages/unit-costs/unit-costs-2011/ (visited on 03/26/2020).

[214] A. Filipović-Pierucci, K. Zarca, and I. Durand-Zaleski, "Markov models for health economic evaluation: The r package heemod", *arXiv preprint arXiv:1702.03252*, 2017.

[215] R. A. Alexander, "A note on averaging correlations", *Bulletin of the Psychonomic Society*, vol. 28, no. 4, pp. 335–336, 1990.

[216] T. P. Morris, I. R. White, and M. J. Crowther, "Using simulation studies to evaluate statistical methods", *Statistics in Medicine*, vol. 38, no. 11, pp. 2074–2102, 2019.

[217] S Todd, A Whitehead, N Stallard, and J Whitehead, "Interim analyses and sequential designs in phase III studies", *British Journal of Clinical Pharmacology*, vol. 51, no. 5, pp. 394–399, 2001.

[218] A. Burton, D. G. Altman, P. Royston, and R. L. Holder, "The design of simulation studies in medical statistics", *Statistics in Medicine*, vol. 25, no. august, pp. 4279–4292, 2006.

[219] S. J. Walters, R. M. Jacques, I. B. dos Anjos Henriques-Cadby, J. Candlish, N. Totton, *et al.*, "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.

[220] R. B. Nelsen, *An introduction to copulas*. Springer Science & Business Media, 2007.

[221] M. Hofert and M. Mächler, "Nested archimedean copulas meet R: The nacopula package", *Journal of Statistical Software*, vol. 39, no. 9, pp. 1–20, 2011.

[222] S. Todd, J. Whitehead, and K Facey, "Point and interval estimation following a sequential clinical trial", *Biometrika*, vol. 83, no. 2, pp. 453–461, 1996.

[223] Q. Liu and K. Anderson, "On Adaptive Extensions of Group Sequential Trials for Clinical Investigations", *Journal of the American Statistical Association*, vol. 103, no. 484, pp. 1621–1630, 2008.

[224] Z. Li and D. L. DeMets, "On the bias of estimation of a brownian motion drift following group sequential tests", *Statistica Sinica*, pp. 923–937, 1999.

[225] J. Whitehead, S. Todd, and W Hall, "Confidence intervals for secondary parameters following a sequential test", *Journal of the royal statistical society B*, vol. 62, pp. 731–745, 2000.

[226] R. Faria, M. Gomes, D. Epstein, and I. R. White, "A guide to handling missing data in cost-effectiveness analysis conducted within randomised controlled trials", *Pharmacoeconomics*, vol. 32, no. 12, pp. 1157–1170, 2014.

[227] H. W. Tuffaha, L. G. Gordon, and P. A. Scuffham, "Value of information analysis informing adoption and research decisions in a portfolio of health care interventions", *MDM Policy & Practice*, vol. 1, no. 1, pp. 1–11, 2016.

[228] T Hastie and R Tibshirani, "Generalized Additive Models", *Statistical Science*, vol. 1, no. 3, pp. 297–310, 1986.

[229] S. D. Eckermann and A. R. Willan, "Globally optimal trial design for local decision making", *Health economics*, vol. 18, pp. 203–216, 2009.

[230] G. Zhao and W. Chen, "Enhancing R&D in science-based industry: An optimal stopping model for drug discovery", *International Journal of Project Management*, vol. 27, no. 8, pp. 754–764, 2009.

[231] L. Flight and S. Julious, "Practical guide to sample size calculations: Non-inferiority and equivalence trials", *Pharmaceutical Statistics*, vol. 15, no. 1, 2016.

[232] Department of Health, *NHS reference costs 2018 to 2019*, 2019.

[233] T. van Asselt, B. Ramaekers, I. C. Ramos, M. Joore, M. Al, *et al.*, "Research costs investigated: A study into the budgets of dutch publicly funded drug-related research", *PharmacoEconomics*, vol. 36, no. 1, pp. 105–113, 2018.

[234] NHS England, *Guidance on excess treatment costs 2015*, 2017.

[235] H. Koffijberg, S. Knies, and M. P. Janssen, "The impact of decision makers' constraints on the outcome of value of information analysis", *Value in Health*, vol. 21, no. 2, pp. 203–209, 2018.

[236] A. R. Willan, R. Goeree, and K. Boutis, "Value of information methods for planning and analyzing clinical studies optimize decision making and research planning", *Journal of clinical Epidemiology*, vol. 65, no. 8, pp. 870–876, 2012.

[237] L. Andronis and P. M. Barton, "Adjusting estimates of the expected value of information for implementation: Theoretical framework and practical application", *Medical Decision Making*, vol. 36, no. 3, pp. 296–307, 2016.

[238] National Institute for Health Research. (2020). National institute for health research, [Online]. Available: https://www.nihr.ac.uk/ (visited on 09/06/2020).

[239] M. B. Gabbay, A. Ring, R. Byng, P. Anderson, R. S. Taylor, *et al.*, "Debt counselling for depression in primary care: An adaptive randomised controlled pilot trial (decoder study).", *Health Technology Assessment (Winchester, England)*, vol. 21, no. 35, p. 1, 2017.

[240] INVOLVE. (2016). Briefing note two: What is public involvement in research?, [Online]. Available: http://www.invo.org.uk/posttyperesource/what-is-public-involvement-in-research (visited on 09/20/2016).

[241] J. Boote, W. Baird, and A. Sutton, "Public involvement in the systematic review process in health and social care: a narrative review of case examples", *Health Policy*, vol. 102, no. 2, pp. 105–116, 2011.

[242] L. Flight, S. Baxter, and S. Khan, "Public involvement beyond clinical research studies", in *Trials*, BioMed Central, vol. 20, 2019, pp. 25–25.

[243] S. S. Emerson and J. M. Kittelson, "A computationally simpler algorithm for the umvue of a normal mean following a group sequential trial", *Biometrics*, pp. 365–369, 1997.

[244] M Dimairo, P Pallmann, T Jaki, *et al.*, *A practical adaptive & novel designs and analysis (panda) toolkit*, 2019.

[245] A. R. Willan, A. H. Briggs, and J. S. Hoch, "Regression methods for covariate adjustment and subgroup analysis for non-censored cost-effectiveness data", *Health Economics*, vol. 13, no. 5, pp. 461–475, 2004.

[246] A. Manca, N. Hawkins, and M. J. Sculpher, "Estimating mean qalys in trial-based cost-effectiveness analysis: The importance of controlling for baseline utility", *Health Economics*, vol. 14, no. 5, pp. 487–496, 2005.

Appendices

Appendix A

Review of Current Practice Search Strategy (Chapter 3)

Search term
Adaptive
Interim
Dose selection
Bayesian
Futility
Enrichment
Stopping rule
Seamless
Group sequential
Go/no go
Preplanned
MAMS/multi-stage/multiple stage/multiple arm
Active learning
Accumulating data
Continuous reassessment
Reanalysis
Pick the winner
Internal pilot
Drop the loser
Dose escalation
Sample size adjustment/sample size re-estimation/sample size modification

TABLE A.1: Search strategy for review of current practice in Chapter 3 adapted from Hatfield *et al.*, 2016

Appendix B

Qualitative Study Documentation (Chapter 4)

Laura Flight
Registration number: 150253644
School of Health and Related Research
Programme: PhD/Health & Related Res

Dear Laura

PROJECT TITLE: Adaptive Design Clinical Trials and their Impact on the Economic Evaluation of Healthcare Technologies - Qualitative Study
APPLICATION: Reference Number 009699

On behalf of the University ethics reviewers who reviewed your project, I am pleased to inform you that on 07/11/2016 the above-named project was **approved** on ethics grounds, on the basis that you will adhere to the following documentation that you submitted for ethics review:

- University research ethics application form 009699 (dated 18/10/2016).
- Participant information sheet 1023219 version 2 (17/10/2016).
- Participant information sheet 1023118 version 2 (17/10/2016).
- Participant consent form 1023121 version 2 (17/10/2016).
- Participant consent form 1023120 version 2 (17/10/2016).

If during the course of the project you need to deviate significantly from the above-approved documentation please inform me since written approval will be required.

Yours sincerely

Jennifer Burr
Ethics Administrator
School of Health and Related Research

Participant ID Number:

Public Interview Sheet

Question	Prompt	Notes
Have you ever participated in a clinical trial or been part of a study as a lay representative?	Familiar with research? What was your role? What input did you have?	
Are you familiar with how we decide which treatments are funded on the NHS in the UK?	Would you know where to find this information? Are you interested in this process? Reference to a news story for context –one that was funded and one that wasn't	
Are you concerned about the methods we use to decide which treatments are funded on the NHS? Why?	What would help you to understand more? What would you want to know about?	
What do you know about adaptive design clinical trials? *An adaptive design clinical trial allows us to examine data from a trial before it has ended. Based on what we see we can make changes to the rest of the trial*	Check understanding of what adaptive design is after watching video Re-iterate definition if required Flexible approach Examine data early and make changes to the trial Participated in an adaptive trial?	
Do you think adaptive designs are a good or bad thing?	Benefits of adaptive designs – flexible, save money, fewer patients Limitation of adaptive designs – less information, bias. difficult to implement, not always practical	
What do you know about cost-effectiveness and value for money?	Check understanding of what cost-effectiveness is after watching video Re-iterate definition if required Value for money of a treatment Is this something that is of interest to you?	
In any clinical trial what do you think the main aim is? *Eg. Designing a new trial to compare two drugs that are supposed to help reduce blood pressure*	Is this what you think it is in reality?	
Are there other factors that you think are important in clinical trials?	What do you want to learn from a trial? Quality of life Side effects of treatment	
Do you think it is important to learn about the cost-effectiveness of the treatment?	Is value for money important to you? Can you see how it is important for decision making?	

Participant ID Number:

I am interested in whether we need to change how we calculate value for money when we use an adaptive design clinical trial and also if we can use information about value for money more in the decisions we make during the trial.		
These methods could potentially make value for money more important/influential. Does this concern you?	Scarce resources so unethical to fund treatments that aren't cost-effective Continuing with a trial where there is evidence of clinical effectiveness but not cost-effectiveness Continuing with a trial where there is evidence of cost effectiveness but not clinical effectiveness	
Would you remain in a trial where the treatment was known to work but we needed to carry on to learn about value for money?	Why would you stay? Why would you leave? Greater good What would motivate a patient to stay in a trial when efficacy known (philanthropy/self-interest)	
If you were participating in a trial that used these methods (an adaptive design and cost-effectiveness played an important role in the decisions made) what information would you like to know about before you agreed to take part?	Informed consent -how to make sure a patient is fully aware of the design and the impact this has on them? Would this impact on recruitment – Do you think people would be put off joining the trial? Would you want to re-consent after each adaptation if there were changes made to the trial? How much information would you want about the analysis at the interim and the changes made?	
Do you think it's important to have a patient representative involved in the decision making at the interim?	What training would they need? Who would be suitable for this role? Would they sit on the DMEC/TMG/TSC?	
Anything else you would like to add or reiterate?		

Participant ID Number:

Researcher Interview Sheet

Question	Prompt	Notes
What is your current role in the Health Technology Assessment/decision making process?	Reviewing grants/publications	
At the moment what is your relationship with other stakeholders in a trial/decision making process?	When are you involved? How much are you involved and at what points? Who do you work with in the study team?	
What is your understanding and experience of adaptive designs?	Designed/analysed/reviewed an adaptive trial? Used in health economic analysis/evidence synthesis? How was adaptive nature of study handled?	
What is your opinion/perception of adaptive designs?	Benefits /Limitations Barriers to their use (training/resources/bridge funding/lack of awareness)	
What is your understanding and experience of health economics?	How often do you see health economics used in design/planning of clinical trial? Barriers to use in the design (large sample size/priorities) Are you aware of how the adaptive design might influence your economic evaluation?	
What is your understanding and experience of value of information analysis?	Have you used this in planning/design of research? How often have you seen this used? Barriers to its use (complexity) Do you think this is a useful tool? How do you think evidence from an adaptive design might influence the analysis?	
In any clinical trial what do you think the main aim is?	Clinical effectiveness? Is this what you think it should be?	
Are there other factors that you think are important in clinical trials?	What do you need to know? Clinical, resource and economic endpoints (Briggs) Safety/acceptability/quality of life	
What role do you think cost-effectiveness plays in clinical trials and decision making?	More or less priority Differ depending on stakeholder? Currently clinical effectiveness dominates the design – are you happy making decisions about cost-effectiveness on reduced power? Working within the clinical effectiveness constraints? Secondary outcomes Piggy backed/add on	

Participant ID Number:

I am interested in applying health economics and adaptive designs together, so this might include using health economics to design an adaptive trial but also using information about health economics to inform the decision and adaptations that we make at the interim analysis of an adaptive design.		
Have you seen this approach used?	Examples Successfully funded? People researching this?	
What do you think the advantages of this approach would be?	Financial incentive/Market driven Better evidence on cost-effectiveness Stopping trials that aren't likely to be funded on cost-effectiveness grounds Increased efficiency Value for money important to decision makers	
What do you think the disadvantages of this approach would be?	Increased complexity Funder/regulator attitude	
The methods proposed potentially change the role of cost-effectiveness in the trial making it more important/influential. Do you have any concerns about the ethical implications of this?	Informed consent how to make sure a patient is fully aware of the design and the impact this has on them? Would this impact on recruitment Biased estimates from adaptive designs used in cost-effectiveness decisions Scarce resources mean it is unethical to fund treatments that aren't cost-effective Continuing with a trial when there is evidence of clinical effectiveness but not cost-effectiveness Continuing with a trial where there is evidence of cost effectiveness but not clinical effectiveness What would motivate patient to stay in a trial when efficacy known (philanthropy/self-interest) Blinding Collective ethics vs individual ethics	
What concerns would you have if we used cost-effectiveness and value of information at interim analyses?	Changes to your role Accuracy of information Training and knowledge	
What challenges do you think you would face that might limit the use of the methods in practice?	Extra analysis at the interim may delay or lengthen study duration when could have been collecting data Time and resources Acceptability to funders/regulators Acceptance by grant reviewers – space available to report design Collection of long term outcomes (QALY)	

	Additional complexity Data management Blinding Interim Reporting – how should this be handled? Solutions to issues (observational data/only in research recommendations)	
Are there any contexts or scenarios where you think this approach would or would not be feasible? Why?	Disease area (rare diseases/end of life – cost-effectiveness in these areas less important) Study population (orphan diseases) Long term outcomes Intervention and available current standard Acceptability to patients	
How would using these methods affect your role?	Relationship with other stakeholders Timing of role within the trial Inclusion of patients? Validation of planning/design/analysis with the whole study team Resources required to make this possible?	
In your clinical role…	How would you use information about a treatment from an adaptive clinical trial? Do you view this information differently? Is evidence based on how cost-effective a treatment is important to you? Does it influence the treatments you provide to patients?	
What training would you require to be able to implement/review/utilise methods?	Education about methods Resources that would be useful Guidance documents	
What would you need to be able to implement these methods?	Guidelines/NICE DSU document Software Audit trail and pre-specification – would this be achievable? Case studies Relationships with other study team members	
What role do you think the DMEC would play in adaptive trials that examine cost-effectiveness?	Decision making left to them? Who would need to be on the committee (health economist as well as statistician)? What training would they need? How would their role be defined?	
Anything else you would like to add or reiterate?		

Appendix C

CACTUS Case Study Ethics Approval (Chapter 5)

Laura Flight
Registration number: 150253644
School of Health and Related Research
Programme: Health and Related Research (PhD/Health & Related Res FT)

Dear Laura

PROJECT TITLE: Secondary Data Analysis - Adaptive Design Clinical Trials and their Impact on the Health Economic Analysis of Healthcare Technolgoies
APPLICATION: Reference Number 014510

On behalf of the University ethics reviewers who reviewed your project, I am pleased to inform you that on 13/07/2017 the above-named project was **approved** on ethics grounds, on the basis that you will adhere to the following documentation that you submitted for ethics review:

- University research ethics application form 014510 (dated 12/07/2017).

If during the course of the project you need to deviate significantly from the above-approved documentation please inform me since written approval will be required.

Yours sincerely

Jennifer Burr
Ethics Administrator
School of Health and Related Research

Appendix D

Explaining the Correlation Values in Simulation Study Two (Chapter 7)

In Simulation Study Two of Chapter 7, despite there being high correlation values between outcomes measured at the same time point (such as baseline utility and baseline percentage words named correctly) the estimated correlations between the primary and health economic outcomes is small (Table 7.6). This can partly be explained using covariance theory.

The results are shown for generic variables a, b, c, d and then illustrated for total costs and QALY to show why the correlations may be small in the simulation study. Let,

$$var(a) = \mathbb{E}(a^2) - \mathbb{E}(a)^2 = \sigma_a^2, \tag{D.1}$$

$$cov(a, b) = \mathbb{E}\left((a - \mathbb{E}(a))(b - \mathbb{E}(b))\right) = \mathbb{E}(ab) - \mathbb{E}(a)\mathbb{E}(b), \tag{D.2}$$

$$corr(a, b) = \frac{cov(a, b)}{\sigma_a \sigma_b}. \tag{D.3}$$

The correlation between two variables (a, b) can be written as,

$$corr(a-b, c-d) = \frac{cov(a-b, c-d)}{\sigma_{(a-b)}\sigma_{(c-d)}}, \tag{D.4}$$

$$= \frac{\mathbb{E}\left((a-b)(c-d)\right) - \mathbb{E}\left(a-b\right)\mathbb{E}\left(c-d\right)}{\sqrt{\mathbb{E}\left((a-b)^2\right) - \mathbb{E}\left((a-b)\right)^2}\sqrt{\mathbb{E}\left((c-d)^2\right) - \mathbb{E}\left((c-d)\right)^2}}, \tag{D.5}$$

$$= \frac{\mathbb{E}\left(ac - ad - bc + bd\right) - \left[\mathbb{E}\left(a\right)\mathbb{E}\left(c\right) - \mathbb{E}\left(a\right)\mathbb{E}\left(d\right) - \mathbb{E}\left(b\right)\mathbb{E}\left(c\right) + \mathbb{E}\left(b\right)\mathbb{E}\left(d\right)\right]}{\sqrt{\mathbb{E}\left(a^2 - 2ab + b^2\right) - \mathbb{E}\left(a\right)^2 + 2\mathbb{E}\left(a\right)\mathbb{E}\left(b\right) - \mathbb{E}\left(b\right)^2}\sqrt{\mathbb{E}\left(c^2 - 2cd + d^2\right) - \mathbb{E}\left(c\right)^2 + 2\mathbb{E}\left(c\right)\mathbb{E}\left(d\right) - \mathbb{E}\left(d\right)^2}}, \tag{D.6}$$

$$= \frac{\left[ac - \mathbb{E}\left(a\right)\mathbb{E}\left(c\right)\right]\left[ad - \mathbb{E}\left(a\right)\mathbb{E}\left(d\right)\right] - \left[bc - \mathbb{E}\left(b\right)\mathbb{E}\left(c\right)\right] + \left[bd - \mathbb{E}\left(b\right)\mathbb{E}\left(d\right)\right]}{\sqrt{\mathbb{E}\left(a^2\right) - \mathbb{E}\left(a\right)^2 + \mathbb{E}\left(b^2\right) - \mathbb{E}\left(b\right)^2 - 2\mathbb{E}\left(ab\right) + 2\mathbb{E}\left(a\right)\mathbb{E}\left(b\right)}\sqrt{\mathbb{E}\left(c^2\right) - \mathbb{E}\left(c\right)^2 + \mathbb{E}\left(d^2\right) - \mathbb{E}\left(d\right)^2 - 2\mathbb{E}\left(cd\right) + 2\mathbb{E}\left(c\right)\mathbb{E}\left(d\right)}}, \tag{D.7}$$

$$= \frac{cov(a,c) - cov(a,d) - cov(b,c) + cov(b,d)}{\sqrt{var(a) + var(b) - 2Cov(a,b)}\sqrt{var(c) + var(d) - 2cov(c,d)}}. \tag{D.8}$$

The covariance between two variables (a, b) can be re-written as,

$$cov(a, b) = \rho_{a,b} \times \sigma_a \times \sigma_b, \tag{D.9}$$

where $\rho_{a,b}$ is the correlation between a and b.

When calculating the correlation between the primary outcome (improvement in percentage words from baseline to 6-months) and total costs the correlation formula becomes,

$$corr(a-b, c) = \frac{Cov(a-b, c)}{\sigma_{a-c}\sigma_c} \tag{D.10}$$

$$= \frac{\mathbb{E}\left[(a-b)(c)\right] - \mathbb{E}\left(a-b\right)\mathbb{E}\left(c\right)}{\sqrt{var(a) + var(b) - 2Cov(a,b)}\sqrt{var(c)}} \tag{D.11}$$

$$= \frac{\mathbb{E}\left[ac\right] - \mathbb{E}\left(a\right)\mathbb{E}\left(c\right) - \mathbb{E}\left[bc\right] + \mathbb{E}\left(b\right)\mathbb{E}\left(c\right)}{\sqrt{var(a) + var(b) - 2cov(a,b)}\sqrt{var(c)}} \tag{D.12}$$

$$= \frac{cov(a,c) - cov(b,c)}{\sqrt{var(a) + var(b) - 2cov(a,b)}\sqrt{var(c)}} \tag{D.13}$$

$$= \frac{\rho_{a,c}\sigma_a\sigma_c - \rho_{b,c}\sigma_b\sigma_c}{\sqrt{var(a) + var(b) - 2cov(a,b)}\sqrt{var(c)}} \tag{D.14}$$

In Simulation Study Two it is assumed that $\sigma_a = \sigma_b$, $\rho_{a,c} = \rho$ as the total costs and percentage words named correctly at 6-months are both measured at 6-months and $\rho_{b,c} = \rho^3$ as total costs and baseline percentage words named correctly are collected one time point apart. In this case, the numerator for the correlation between total costs and improvement in percentage words

named correctly will be small when ρ and ρ^3 are close in value, hence giving a small correlation between these two outcomes.

To calculate the QALY the sum of the utility at baseline and 6-months is multiplied by 0.25 by the triangle rule. Ignoring the multiplication factor the correlation formula between the primary outcome and the sum of baseline and 6-month utility becomes,

$$cov(a-b, c+d) = \frac{cov(a-b, c+d)}{\sigma_{a-b}\sigma_{c+d}} \tag{D.15}$$

$$= \frac{cov(a,c) + cov(a,d) - Cov(b,c) - cov(b,d)}{\sqrt{var(a) + var(b) - 2cov(a,b)}\sqrt{var(c) + var(d) + 2cov(c,d)}} \tag{D.16}$$

$$= \frac{\rho_{a,c}\sigma_a\sigma_c + \rho_{a,d}\sigma_a\sigma_d - \rho_{b,c}\sigma_b\sigma_c - \rho_{b,d}\sigma_b\sigma_d}{\sqrt{var(a) + var(b) - 2cov(a,b)}\sqrt{var(c) + var(d) + 2Cov(c,d)}} \tag{D.17}$$

In Simulation Study Two it is assumed that $\sigma_a = \sigma_b$ and $\sigma_c = \sigma_d$. The correlation $\rho_{a,c} = \rho_{b,d} = \rho^3$ as the baseline utility and percentage words named correctly at 6-months are one time point apart, as are baseline percentage words at baseline and utility at 6-months. $\rho_{a,d} = \rho_{b,c} = \rho$ as baseline percentage words and baseline utility and percentage words at 6-months and utility at 6-months are collected at the same time point. In this case, the numerator is close to zero as the correlations cancel out. This explains why the correlation between the primary outcome and QALY in Simulation Study Two are small even for high values of ρ for the data generation.

In Chapter 7 it shown that as the correlation increased above 0.6 there was a decrease in the correlation between the primary outcome and health economic outcomes. This is a consequence of the correlation structure chosen for the data generation described in Section 7.6.1 and the results shown above where the correlation is driven by additions and subtractions of ρ, ρ^3 and ρ^4. As shown in Figure D.1, as x increases $x - x^3$ increases, until x reaches 0.6, at which point $x - x^3$ begins to decrease.

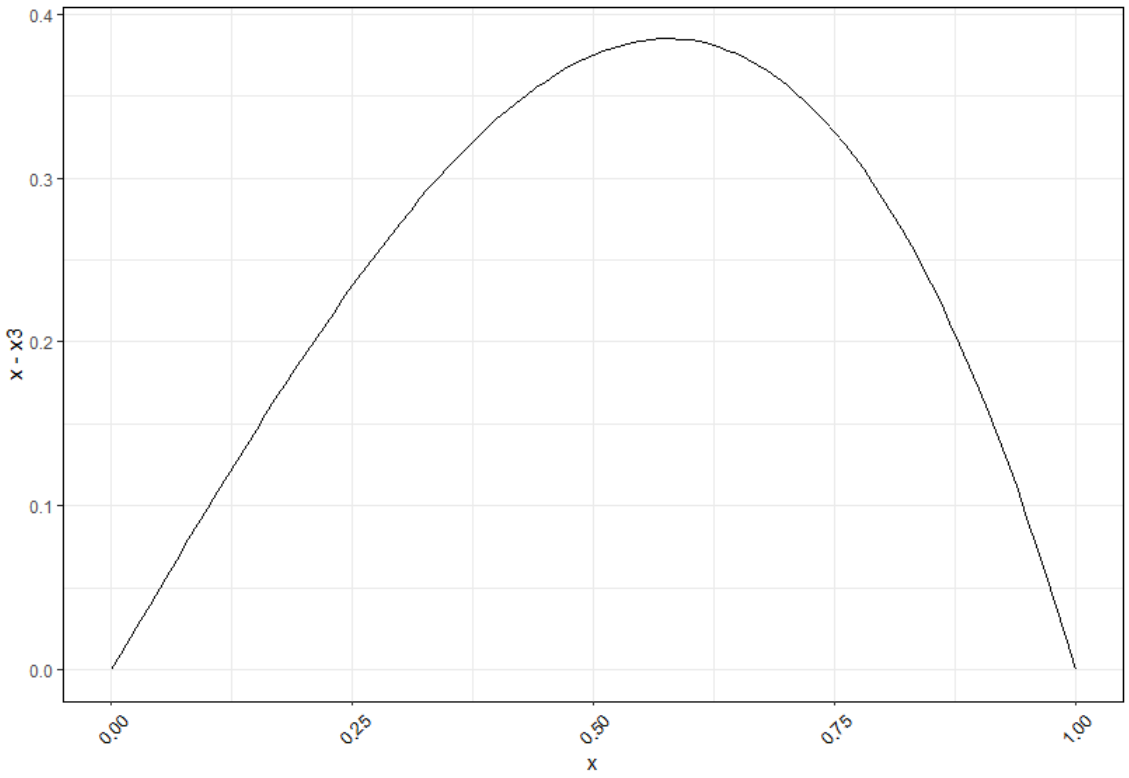


FIGURE D.1: Illustration of how the correlation reduces despite increasing values of the baseline correlation in Simulation Study Two

Appendix E

Comparison of Adjusted and Unadjusted Expected Value of Sample Information and Model Parameters (Chapter 8)

Design		FIX	OBF 2	OBF 5	POC 2	POC 5	FIX	OBF 2	OBF 5	POC 2	POC 5	FIX	OBF 2	OBF 5	POC 2	POC 5
Correlation				0.0					0.4					0.8		
Pilot Data with Reduced Prior Uncertainty																
EVSI	**adjusted**	7.242	5.071	3.854	4.865	4.321	10.358	6.969	6.285	6.615	4.198	8.553	6.654	5.148	6.126	2.561
	unadjusted	6.886	4.638	3.521	4.288	3.461	9.069	5.286	4.361	5.282	2.768	7.902	5.325	4.776	4.387	2.179
	difference	2.520	4.462	4.509	6.312	11.052	6.636	13.736	18.075	11.206	20.522	3.961	11.095	3.743	16.537	8.077
ENBS	**adjusted**	-0.068	-0.423	-0.660	-0.429	-0.473	0.792	0.113	0.014	0.060	-0.499	0.294	0.020	-0.295	-0.076	-0.944
	unadjusted	-0.167	-0.543	-0.752	-0.589	-0.711	0.436	-0.352	-0.517	-0.308	-0.894	0.114	-0.347	-0.397	-0.556	-1.050
	difference	-41.855	-12.375	-6.507	-15.669	-20.060	28.982	-194.910	-105.764	-148.246	-28.331	44.142	-112.575	-14.833	-75.971	-5.302
Pilot Data																
EVSI	**adjusted**	35.908	26.641	21.612	25.619	17.992	34.606	31.319	26.747	30.370	20.589	37.706	27.956	24.671	29.095	22.517
	unadjusted	32.912	23.814	22.259	24.019	17.965	35.574	28.438	27.512	28.396	18.733	36.946	26.822	26.232	26.693	21.173
	difference	4.354	5.604	-1.476	3.223	0.074	-1.379	4.821	-1.409	3.360	4.720	1.019	2.070	-3.067	4.305	3.078
ENBS	**adjusted**	7.789	5.516	4.215	5.262	3.268	7.429	6.797	5.633	6.556	3.980	8.285	5.885	5.077	6.224	4.528
	unadjusted	6.961	4.736	4.394	4.820	3.260	7.696	6.001	5.844	6.011	3.467	8.075	5.571	5.508	5.561	4.156
	difference	5.610	7.617	-2.076	4.383	0.112	-1.767	6.217	-1.840	4.339	6.883	1.284	2.734	-4.073	5.628	4.276
Resource cost intervention arm	**adjusted**	201.692	201.446	201.578	201.324	202.556	202.635	202.356	202.321	202.543	202.813	203.863	204.126	204.199	203.740	203.664
	unadjusted	201.692	201.441	201.488	201.321	202.461	202.635	201.824	201.471	202.044	201.482	203.863	203.617	203.600	203.271	202.456
	difference	0.000	0.001	0.022	0.001	0.023	0.000	0.132	0.211	0.123	0.329	0.000	0.125	0.147	0.115	0.298
Resource cost control arm	**adjusted**	270.003	269.703	269.613	269.663	270.012	270.379	270.574	270.634	270.568	270.187	271.058	270.850	270.510	270.737	270.465
	unadjusted	270.003	269.708	269.535	269.666	269.921	270.379	271.106	271.315	271.067	271.579	271.058	271.359	271.372	271.205	271.816
	difference	0.000	-0.001	0.014	-0.001	0.017	0.000	-0.098	-0.126	-0.092	-0.257	0.000	-0.094	-0.159	-0.086	-0.249
Utility increment	**adjusted**	-0.004	-0.004	-0.004	-0.004	-0.003	0.260	0.255	0.254	0.256	0.250	0.217	0.213	0.211	0.213	0.209
	unadjusted	-0.004	-0.004	-0.004	-0.004	-0.003	0.260	0.258	0.259	0.258	0.257	0.217	0.216	0.215	0.216	0.215
	difference	0.000	0.037	-0.243	0.060	0.265	0.000	-0.588	-0.863	-0.541	-1.535	0.000	-0.618	-0.909	-0.558	-1.598
Probability of a good response	**adjusted**	0.504	0.505	0.505	0.505	0.508	0.504	0.506	0.504	0.506	0.508	0.503	0.505	0.505	0.505	0.507
	unadjusted	0.487	0.491	0.491	0.490	0.496	0.487	0.491	0.491	0.490	0.496	0.487	0.491	0.491	0.491	0.496
	difference	1.713	1.474	1.388	1.501	1.255	1.702	1.467	1.268	1.521	1.152	1.648	1.451	1.343	1.488	1.117
Probability of relapse	**adjusted**	0.009	0.009	0.009	0.009	0.010	0.009	0.009	0.009	0.009	0.010	0.009	0.009	0.009	0.009	0.009
	unadjusted	0.009	0.010	0.011	0.010	0.012	0.009	0.010	0.011	0.010	0.012	0.010	0.010	0.011	0.010	0.012
	difference	-4.652	-5.882	-11.612	-5.714	-10.406	-4.280	-5.285	-12.148	-5.079	-11.072	-4.718	-6.092	-12.060	-6.031	-11.282
Treatment effect	**adjusted**	0.129	0.129	0.129	0.129	0.130	0.128	0.129	0.128	0.129	0.130	0.129	0.130	0.130	0.130	0.131
	unadjusted	0.129	0.135	0.136	0.134	0.143	0.128	0.134	0.135	0.133	0.143	0.129	0.135	0.138	0.135	0.144
	difference	0.000	-1.949	-2.654	-1.803	-4.797	0.000	-1.931	-2.805	-1.792	-4.817	0.000	-1.940	-2.790	-1.767	-4.815

TABLE E.1: Adjusted and unadjusted estimates of the EVSI, ENBS, health economic model parameters and the clinical primary outcome