Report for the EcoNomics of Adaptive Clinical Trials (ENACT) project: Application of a Bayesian Value-Based Sequential Model of a Clinical Trial to the CACTUS and HERO Case Studies (with Guidance Material for Clinical Trials Units)

Martin Forster[1], Laura Flight[2], Belen Corbacho[3], Ada Keding[3], Sarah Ronaldson[3], Puvan Tharmanathan[3], Charlie Welch[3], Alan Brennan[2], and Stephen Chick[4]

[1]*Department of Statistical Sciences 'Paolo Fortunati', University of Bologna, Italy; Department of Economics and Related Studies, University of York, United Kingdom*
[2]*School of Health and Related Research, University of Sheffield, United Kingdom*
[3]*York Trials Unit, Department of Health Sciences, University of York, United Kingdom*
[4]*Technology and Operations Management Area, INSEAD, FRANCE.*

September 8, 2021

Acknowledgements

The ENACT project is funded by the National Institute for Health Research (NIHR) CTU Support Funding scheme (2019 call) to support efficient/innovative delivery of NIHR research. The views expressed are those of the authors and not necessarily those of the NIHR or the Department of Health and Social Care.

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 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 expressed are those of the author(s) and not necessarily those of the National Health Service, NIHR or the Department of Health and Social Care Health Technology Assessment Programme, the Tavistock Trust for Aphasia, or the Stroke Association.

The HERO trial was funded by an Arthritis Research U.K (now Versus UK) clinical studies grant (reference 19545). Kingsbury et al. (2018) present the results of the clinical evaluation. Ronaldson et al. (2021) present the health economics findings.

Contents

List of Tables

List of Figures

Abstract

Background. There is increasing interest in the use of adaptive designs to improve the efficiency of clinical trials. Little work has been done to assess how much value these designs might deliver for health services and their patients. A Bayesian decision-theoretic model of a sequential experiment was proposed by Chick et al. (2017a) and Alban et al. (2020) and applied retrospectively to a National Institute for Health Research (NIHR)-funded clinical trial by Forster et al. (2021). The model calculates the expected health and economic value of gathering further information about the cost-effectiveness of two health technologies as the data accrue during the trial. It enables data collection to operate dynamically, governed by a series of stop-go decisions which balance the expected benefits of continuing the trial with the expected costs. The EcoNomics of Adaptive Clinical Trials (ENACT project) was part of the NIHR's Efficient Studies funding call (2019) for Clinical Trials Units (CTUs). The project set out to deliver publications, together with guidance material and training for CTUs, in the area of value-based sequential clinical trial designs.

Objectives of this report. To cover Deliverable 3 of the ENACT project: 1. to apply the Bayesian decision-theoretic model of a sequential experiment proposed by Chick et al. (2017a) and Alban et al. (2020) (the 'value-based sequential design') to two new retrospective case studies from the United Kingdom – the 'Clinical and cost-effectiveness of aphasia computer treatment versus usual stimulation or attention control long term post-stroke' (Big CACTUS) trial (Palmer et al., 2019, 2020; Latimer et al., 2020) and the 'Hydroxychloroquine Effectiveness in Reducing symptoms of hand Osteoarthritis' (HERO) trial (Kingsbury et al., 2018). 2. To compare the performance of the value-based sequential design with two comparator designs: (a) the design of the trial itself (a traditional, fixed sample size design) and (b) a fixed sample size design which maximises the expected net benefit of sampling for a Bayesian decision-theoretic model using a normal prior and a normal likelihood (the 'value-based one stage design', see Claxton and Posnett (1996) and Claxton (1999)). 3. To add to existing knowledge by exploiting special features of the Big CACTUS and HERO trials, including the use of a pilot study, carrying out a health economic modelling approach to cost-effectiveness analysis and applying methods to deal with missing data. *In addition:* 4. To contribute to a summary of the procedures that a clinical trials unit could follow if it is considering using the

value-based sequential design in a future clinical trial.

Methods. Data from each trial and other sources were used to inform the choice of parameter values for the value-based sequential design and obtain its stopping boundary, which provides a rule for halting recruitment to the trial based on the accumulating cost-effectiveness evidence. An estimate of the value delivered by the value-based sequential design was compared with the value delivered by the two comparator designs described above. A bootstrap analysis, which simulated a large number of trials, was used to estimate the sample size that could be expected from each design, as well as the expected cost and the probability of concluding that each technology is cost-effective.

Results. Under reasonable assumptions, this study shows that it is possible to apply a value-based sequential design of a clinical trial using real-world data. For the CACTUS case study, there is evidence that the value-based sequential design could have delivered additional value over the two comparator designs. For example, we estimate that such a model, operating with a maximum sample size of 528 participants, could have delivered an expected value of approximately £4.13m, 17% higher than the value that was delivered by the original trial design (£3.54m), and higher than that of the value-based one stage design (£3.85m). For the HERO case study, although the value-maximising sample size of the value-based one stage design is 43% higher than the sample size of the trial (354 participants instead of 248), it only delivers a small amount of additional value (an increase of 0.1% on £52m). The value-based sequential design delivers a slightly lower expected sample size (348 participants) and slightly higher additional value (an increase of about 0.3%).

Conclusions. 1. Considered alongside the results from the only other retrospective application of this value-based sequential design (Forster et al., 2021), we have demonstrated the potential for applying the value-based sequential design to trials run in a publicly-funded health care system such as the one in the United Kingdom. 2a. A sequential approach can provide additional value for the health care system. However, the additional value varies according to the application: the highest value is seen in the CACTUS study and negligible added value is demonstrated in the HERO (and ProFHER) case studies. 2b. Research cost savings may exist when one health technology is superior, from the effectiveness or cost-saving perspective (or both): in such scenarios, a value-based sequential design can stop a trial earlier than planned, whereas a one-stage design cannot. 3. Approaches exploring special features of trials, including the use of a pilot study, carrying out a health economic modelling approach and dealing with missing data, have been explored. 4. In guidance material, we present a summary of the procedures that a clinical trials unit could follow if it is considering using the value-based sequential design in a future clinical trial.

Chapter 1

Introduction

Martin Forster, Laura Flight, Puvan Tharmanathan, Alan Brennan and Stephen Chick

This report presents results from Deliverable 3 of the EcoNomics of Adaptive Clinical Trials (ENACT) project.[1] The ENACT project was part of the NIHR's Efficient Studies funding call (2019) for Clinical Trials Units (CTUs). The project set out to deliver publications, together with guidance material and training for CTUs, in the area of value-based sequential clinical trial designs.

Deliverable 3 had four principal objectives:

1. to apply the Bayesian decision-theoretic model of a sequential experiment proposed by Chick et al. (2017a) and Alban et al. (2020) (the 'value-based sequential design') to two new retrospective case studies from the United Kingdom – the 'Clinical and cost-effectiveness of aphasia computer treatment versus usual stimulation or attention control long term post-stroke' (Big CACTUS) trial (Palmer et al., 2019, 2020; Latimer et al., 2020) and the 'Hydroxychloroquine Effectiveness in Reducing symptoms of hand Osteoarthritis' (HERO) trial (Kingsbury et al., 2018).

2. To compare the performance of the value-based sequential design with two comparator designs: (a) the design of the trial itself (a traditional, fixed sample size design) and (b) a fixed sample size design which maximises the expected net benefit of sampling for a Bayesian decision-theoretic model using a normal prior and a normal likelihood (the 'value-based one stage design', see Claxton and Posnett (1996) and Claxton (1999)).

3. To add to existing knowledge by exploiting special features of the Big CACTUS and HERO trials, including the use of a pilot study, carrying out a health economic modelling

[1] https://www.sheffield.ac.uk/scharr/research/centres/ctru/enact. The Glossary (page 99) defines key abbreviations and provides references to relevant websites.

approach to cost-effectiveness analysis and applying methods to deal with missing data.

4. To contribute to a summary of the procedures that a clinical trials unit could follow if it is considering using the value-based sequential design in a future clinical trial.

By 'value-based', we mean a sequential trial whose stopping rule is informed by balancing the costs of continuing the clinical trial with the expected health benefits, where both costs and benefits are measured in a common unit of currency (for example, pound sterling, euro or dollars). This idea is discussed in more detail in section 1.3.

The two case studies are presented in Chapters 2 and 3. Discussion of the procedures that a clinical trials unit (CTU) should follow if it is considering using the value-based sequential design in a future clinical trial is presented in Chapter 4. Main results and ideas for future research are presented in Chapter 5.

The case studies are:

1. The CACTUS case study, which is based on the 'Clinical and cost-effectiveness of aphasia computer treatment versus usual stimulation or attention control long term post-stroke' (Big CACTUS) trial (Palmer et al., 2019, 2020; Latimer et al., 2020). Funded by the National Institute for Health Research (NIHR), the Big CACTUS trial assessed whether a computerised speech and language therapy intervention improved word naming ability compared to usual care in participants with aphasia post-stroke. The trial's budget was £1,480,713. The trial concluded that the computerised intervention was effective at improving word naming ability, but not functional conversation. Additionally, the cost-effectiveness results were very uncertain, suggesting that computerised speech and language therapy was unlikely to be cost-effective for the full population, but may have been cost-effective for some subgroups.

2. The HERO case study, which is based on the Hydroxychloroquine Effectiveness in Reducing symptoms of hand Osteoarthritis (HERO) trial (Kingsbury et al., 2018; Ronaldson et al., 2021). Funded by Arthritis Research UK (now Versus UK), the HERO trial evaluated whether hydroxychloroquine is superior to placebo for the treatment of hand osteoarthritis. The trial's budget was £900,000. The trial concluded that hydroxychloroquine was neither clinically effective, nor cost-effective.

Together with the ProFHER case study (Forster et al., 2021), the HERO and CACTUS case studies constitute emerging analyses of one approach to running a value-based, sequential, clinical trial using a Bayesian decision-theoretic method.

1.1 Overview of the funder perspective of the need for more efficient clinical trials

The NIHR, as a leading public funder of health and social care in the UK, takes a keen interest in developing methodology for the design, implementation, analysis and reporting of more efficient research studies. This commitment is demonstrated by the active invitation for industry to collaborate with the NIHR to deliver efficient studies (National Institute for Health Research, 2020) and the provision of methodological research development grants within its own research infrastructure (National Institute for Health Research, 2014). The ENACT project, which aims to explore further a value-based sequential approach to clinical trial design, complements other methodological work funded by the NIHR. This includes the development of best practice for costing in adaptive clinical trials (CAT, 2020), extending CONSORT guidance for adaptive designs (Dimairo et al., 2020) and developing a Practical Adaptive & Novel Designs and Analysis (PANDA) toolkit (PANDA, 2020).

Innovative methods, such as value-based sequential approaches, could be adopted by other research funders, both in the United Kingdom (UK) and internationally. Hence, for the purposes of the ENACT project, both NIHR-funded (CACTUS) and non-NIHR-funded (HERO) projects were included as case studies. Each case study deals with different populations and interventions and has unique features of interest (discussed below). The case studies lend insight into how clinical trials could be stopped early, or may run longer than planned, according to judgements about the cost-effectiveness of continuing the research process itself, as evidence about the cost-effectiveness of the health technologies that are being compared accumulates.

1.2 The case studies: clinical trials in the United Kingdom

The value-based sequential design which is the focus of the ENACT project has only been applied in one retrospective case study previously, in the proof-of-concept work of Forster et al. (2021). Forster et al. (2021) used data from the NIHR-funded ProFHER pragmatic trial (Handoll et al., 2015; Rangan et al., 2015), which investigated the use of sling immobilisation and surgery for the treatment of fracture of the proximal humerus in adults. The authors present the value-based sequential model as one which could complement, rather than replace, existing fixed sample size, or group sequential, designs (Forster et al., 2021, Section 1). They estimate that, had the value-based sequential design been used during the trial, it would have stopped the trial early, reducing the sample size by about 14%, saving about 5% of the research budget, and resulting in the same conclusion regarding the cost-effectiveness of the interventions as the original trial analysis (which is that surgery was not cost-effective). A bootstrap analysis of the trial's cost-effectiveness data suggests that the expected sample size would have been about

38% lower, saving around 13% of the research budget, with an estimated probability of 0.92 of showing that surgery was not cost-effective.

The two case studies presented in this report contain features which were not considered in the ProFHER case study:

- The Big CACTUS trial was preceded by an external pilot trial. This provides an opportunity to use information from the pilot to set the parameters of the prior distribution for the measure of cost-effectiveness that is used throughout this report – the expected value of incremental net monetary benefit – before the main trial commences. The Big CACTUS trial also used a model-based approach to cost-effectiveness analysis, rather than a within-trial approach. This presents an opportunity to investigate how a modelling approach might be combined with the value-based sequential design.

- The HERO trial was characterised by a considerable amount of missing data, which can lead to reduced power and efficiency, as well as bias (Bell et al., 2014). Our analysis of the HERO trial illustrates how sequential multiple imputation methods can be incorporated into the value-based sequential design. The HERO trial also recruited its participants relatively quickly, and so provides a good opportunity to investigate how the number of participants 'in the pipeline' at the time the sequential analysis starts – pipeline participants are those who have been treated, but whose health outcomes and treatment costs have yet to be observed (Hampson and Jennison, 2013) – affects the expected value delivered by the trial.

Further, the ProFHER and HERO case studies use post-trial estimates of population variances and budget spend to inform the pre-trial choice of parameter values required to run the value-based sequential design. In contrast, the CACTUS case study takes the perspective of a researcher who has reached the end of an external pilot trial and is considering using the value-based sequential design for the full scale randomised controlled trial. Only the information available before the trial begins is used to inform the parameter values of the model. In this way, the ProFHER and HERO applications can be considered as descriptive, and fully retrospective; the CACTUS case study can be considered as a first attempt to explore how the methodology could be applied prospectively.

1.3 Why the interest in a value-based, sequential approach to clinical trial design?

Interest in a value-based, sequential approach to clinical trial design is motivated by two ideas which have, traditionally, been considered independently of one another. The first is

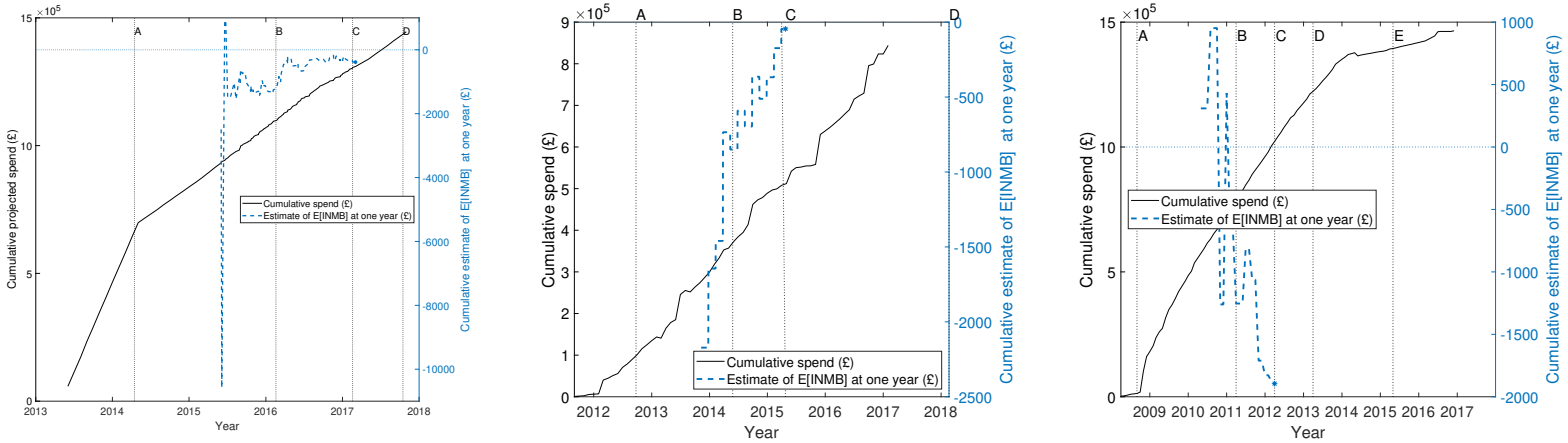
the economist's idea of 'opportunity cost': spending research monies on funding one clinical trial means that those monies are not available to fund, or part-fund, another clinical trial (or trials, or patient care, etc.). For a health care system which seeks to maximise the health that is generated from the resources that are available to it, consideration of opportunity cost captures the essence of the resource allocation problem and the idea of making decisions that are 'value-based'.[2]

The second idea is that monitoring data on the cost-effectiveness of health technologies as it accumulates over the course of a clinical trial provides the opportunity to learn, both about the technologies themselves and the way that research monies are being spent. This presents opportunities for decision-makers to stop a trial when the costs of continuing it are believed to outweigh the benefits. A trial could stop earlier than planned, offering the potential to save research costs, bring the better technology to patients sooner, and reduce the number of patients randomised into the trial. A trial could also run longer than planned, so as to improve the precision of the estimate of cost-effectiveness in the event of an equivocal signal emerging from the trial.

There are vast literatures on approaches to running value-based fixed sample size (that is, not adaptive/sequential) clinical trials (Claxton, 1999; Fenwick et al., 2020; Rothery et al., 2020) and running adaptive/sequential (that is, not value-based) clinical trials (Sutton et al., 2012; Cui et al., 2017; Hampson and Jennison, 2013; Pallmann et al., 2018). Yet, to date, these ideas tend to have been considered separately, rather than together (Flight et al., 2019).

To fix ideas, consider the two case studies that are the focus of the ENACT project, plus the application to the ProFHER pragmatic trial. Figure 1.1 presents time profiles showing our estimates of how the research budgets of these three trials were spent. The financial data comes from the accounts held by the institutions holding the research grants and, absent guidance on how they should be allocated as the trial progresses, we have made 'best guesses' about the time profile. Also plotted are estimates of how evidence of cost-effectiveness, measured by the average value of incremental net monetary benefit of the new intervention against standard or placebo at the level of an individual, accumulated over the course of the trial (the next section discusses this measure of cost-effectiveness in more detail). Average values of incremental net monetary benefit lying above zero suggest that: for the Big CACTUS trial, computerised intervention speech and language therapy is cost-effective when compared to usual care; for the HERO trial, hydroxychloroquine is cost-effective when compared to placebo; for the ProFHER trial, surgery is cost-effective when compared with sling immobilisation. Values below zero suggest the opposite. To keep things simple, we do not present estimates of the degree of uncertainty surrounding these estimates.

[2]We note that equity considerations – not just how much health is generated, but also how it is distributed – could be an important consideration for the health care system (Cookson et al., 2017). We do not consider matters concerning equity in this report.



(a) The Big CACTUS trial

(b) The HERO trial

(c) The ProFHER trial

Figure 1.1: Estimates of cumulative budget spend and expected incremental net monetary benefit for the Big CACTUS, HERO and ProFHER trials.

Notes:
Estimates of expected incremental net monetary benefit that are greater than zero suggest that: for the Big CACTUS trial, computerised intervention speech and language therapy is cost-effective; for the HERO trial, hydroxychloroquine is cost-effective; for the ProFHER trial, surgery is cost-effective. For the Big CACTUS and HERO trials, early evidence suggesting one of the two technologies was cost-effective became weaker as follow-up continued. By the end of these trials, the estimates of the 'value-added' of one technology over the other were in the region of £400 (Big CACTUS) and zero (HERO). For the ProFHER trial, evidence suggested that sling was cost-effective from an early stage and the final point estimate of incremental net monetary benefit was around £-2,000. For the meaning of letters 'A' to 'D' in (a), see Figure 2.3; for letters 'A' to 'D' in (b), see Figure 3.1; for (c): 'A' – recruitment starts; 'B' – recruitment finishes; 'C' – one year follow-up finishes; 'D' – two year follow-up finishes; 'E' – publication of principal articles (Handoll et al., 2015; Rangan et al., 2015).

The plots tell different stories about how the accumulating evidence of cost-effectiveness evolved during the trials. For the Big CACTUS and HERO trials, early evidence favouring one of the two technologies on cost-effectiveness grounds (shown by estimates of cost-effectiveness which lie far from zero in either direction) became weaker as follow-up continued: by the end of these trials, the estimates of the 'value-added' of one technology over the other are in the region of –£400 (Big CACTUS trial) and near to zero (HERO trial). For the ProFHER trial, evidence suggested that sling was cost-effective from an early stage, with the final point estimate of the expected value of incremental net monetary benefit being in the region of –£2,000 (considered with the 95% confidence interval, the result shows reasonably strong evidence that surgery is not cost-effective).

The question that is addressed by the value-based sequential design is the following one: imagine that the estimate of cost-effectiveness, and the uncertainty that surrounds it, are monitored as the trial progresses and there exists the option to randomise further participants into the trial or stop the trial. When is the best time to stop recruitment, from the perspective of operating a cost-effective trial? The value-based sequential design answers this question by using the point estimate of cost-effectiveness, together with its level of uncertainty and the estimated cost of continuing the trial, to identify a point when the trial should stop. This is the point at which the benefit of randomising a further pair of patients to each arm of the trial is deemed to be not worth the cost.

1.4 Methodology

The case studies use the Bayesian decision-theoretic model of a sequential experiment proposed by Chick et al. (2017a) and extended by Alban et al. (2020). This section describes the model, summarises the two comparator designs that we consider, and discusses how we assess the performance of the three models.[3]

1.4.1 Overview of the Bayesian decision-theoretic model of a sequential experiment

Chick et al. (2017a) model a two-armed sequential clinical trial in which participants are randomised, in a pairwise and sequential manner, to a new health technology, N, and a control, or standard, technology, S. Outcome and treatment cost data are available after a defined period of follow-up. The model is Bayesian and a prior distribution is used to model beliefs about the expected value of the cost-effectiveness of the two technologies before the trial starts. Cost-effectiveness is measured by the incremental net monetary benefit of the new technology versus

[3]The material in this section is based on material from Forster et al. (2021) and Chick et al. (2017a).

the standard technology. Define n as the maximum number of pairwise allocations that can be made in the trial and τ as the number of pairwise allocations that must be recruited before the first outcome data are available. Pairwise allocations $i = 1, \ldots, n$ provide the following information on cost-effectiveness:

$$\text{Net benefit of new technology} = \lambda E_{iN} - C_{iN}, \tag{1.1}$$

$$\text{Net benefit of standard technology} = \lambda E_{iS} - C_{iS}, \tag{1.2}$$

$$\begin{aligned} \text{Incremental net monetary benefit} &= \text{Net benefit of new} - \text{Net benefit of standard} \\ &= (\lambda E_{iN} - C_{iN}) - (\lambda E_{iS} - C_{iS}) \\ &= \lambda(E_{iN} - E_{iS}) - (C_{iN} - C_{iS}), \end{aligned} \tag{1.3}$$

where E is effectiveness, C is treatment cost and λ is the maximum willingness to pay of a reimbursement agency for one unit of effectiveness.[4] The prior beliefs about the unknown expected value of incremental net monetary benefit are assumed to have a normal distribution with a specified expected value and variance. Sampling is assumed to take place from a normal distribution with a known variance so that, as observations on incremental net monetary benefit arrive sequentially from participants who have been followed-up, application of Bayes' rule leads to a posterior distribution for the expected value of incremental net monetary benefit which is also normal, with expected value and variance given by standard expressions (see, e.g. Spiegelhalter et al. (1994a)).

The solution of the value-based sequential design is a policy, or rule, which determines whether the clinical trial should stop or randomise another pair of patients to each arm. The policy is informed by the posterior mean and variance for the expected value of incremental net monetary benefit and maximises the expected benefit of the trial. This is defined as the difference between the expected benefit accruing to the patients whose treatment will be determined by the adoption of the preferred technology (minus any costs of switching technologies), minus the cost of carrying out the trial. For the two case studies presented in this report, it is assumed that the cost of switching technologies is small enough to be treated as zero.

There are three distinct stages to the trial, defined according to whether participants are being randomised to the two arms and/or whether there is the option to randomise further participants:

1. during *Stage I*, participants are randomised to the two treatments, but no outcome and cost data have been observed;

2. during *Stage II*, the prior distribution for the expected value of incremental net monetary

[4] For example, it is common for the UK National Health Service to use a value between £20,000 and £30,000 per Quality Adjusted Life Year (QALY) (NICE, 2013).

benefit is updated using Bayes' rule as cost-effectiveness data is observed and there is the option to randomise further participants or stop recruitment to the trial. If, during Stage II, the expected benefit of randomising a further pair of participants is less than the expected cost, Stage II finishes and the trial moves to Stage III.

3. during *Stage III*, recruitment has finished, follow-up data for participants who remain in the pipeline are observed and Bayesian updating continues.

Chick et al. (2017a) note two situations in which it is not optimal to run the trial as a sequential trial. In the first, the expected value of running the trial using a fixed sample size design (see, e.g., Claxton and Posnett (1996) and Claxton (1999)) is higher than the expected value of running the value-based sequential design. In the second, the value of the prior mean is so high (or so low) that it is not optimal to run either a fixed sample size or a sequential trial. Instead, the technology adoption decision selects the new technology if the prior mean is greater than zero or the existing technology if it is less than zero. A summary of the rule which is used to stop the sequential trial, and to determine which design is optimal, is shown in Figure 1.2. This presents a stopping boundary for the trial in (pairwise allocations $\times$ prior/posterior mean) space. The three stages of the sequential trial, as just described, are marked as 'I', 'II' and 'III'.

How is the optimal choice of trial design determined? The value of the prior mean is compared with the regions indicated by the letters 'A' to 'D' in Figure 1.2. It is optimal to run the value-based one stage design if the prior mean lies between 'A' and 'C' or 'D' and 'B'; it is optimal to run no trial if the prior mean is above A (in which case the new technology should be adopted immediately) or below B (in which case the standard technology should be adopted immediately). It is optimal to run the value-based sequential design if the prior mean lies between 'C' and 'D'.

If the value-based sequential design is the preferred design, how is the stopping rule deployed in practice? During Stage II, as data on incremental net monetary benefit arrives from participants who have been followed-up, Bayesian updating takes place at a series of so-called 'interim analyses'. If, at an interim analysis, the posterior mean for the expected value of incremental net monetary benefit lies in the continuation region, it is optimal to continue to randomise participants to the two arms of the trial. Once the path for the posterior mean crosses the stopping boundary, it is optimal to halt recruitment and move to Stage III.

The model proposed by Chick et al. (2017a) did not allow for the possibility that patients were being treated with the new technology when the trial started. An extension proposed by Alban et al. (2020) accounts for mixed clinical practice at the start of the trial, which was a feature of the ProFHER trial (Forster et al., 2021), but not the case studies presented in this report.

In the two case studies that are presented in this report, we use information from the tri-

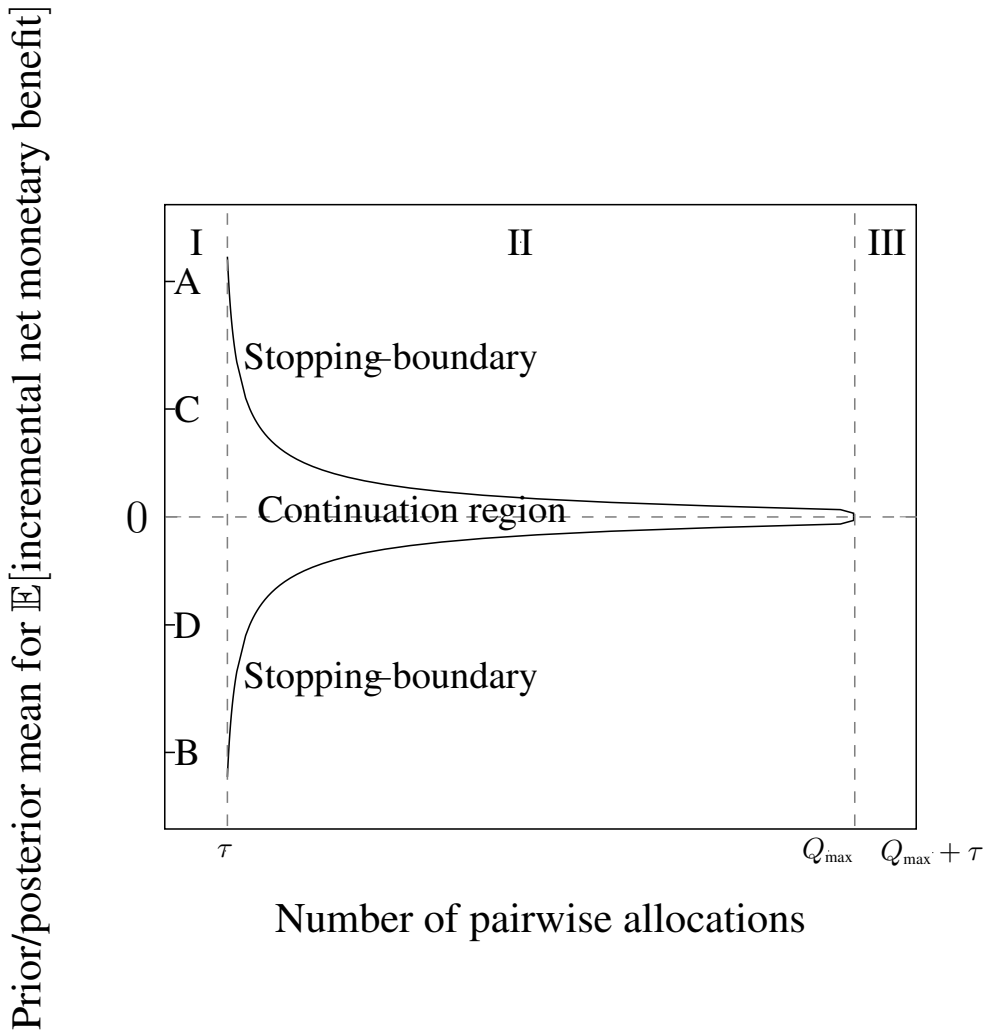


Figure 1.2: Stopping boundary for the value-based sequential design, showing the three stages of the trial (marked 'I', 'II' and 'III'), the stopping boundary and the continuation region. [Source: adapted from Chick et al. (2017a) and Forster et al. (2021)].

Notes:
During Stage I, participants are randomised to the two arms of the trial but no outcome data for cost-effectiveness are observed owing to the delay in observing these outcomes. The first cost-effectiveness data are observed at the start of Stage II and are used to obtain the posterior mean for the expected value of incremental net monetary benefit. During Stage II, as cost-effectiveness data arrive in sequence, the prior/posterior mean is updated and there is the option to randomise further participants into the trial or stop the trial and move to Stage III. It is optimal to continue the trial as long as the posterior mean lies in the 'continuation region'; Stage II stops once the posterior mean crosses the stopping boundary. During Stage III, outcome data for the pipeline participants – those who have been treated but not yet followed-up – are observed and Bayesian updating continues. Once all outcomes have been observed, the new technology is deemed cost-effective if the posterior mean lies above zero, otherwise it is deemed not to be cost-effective.

als themselves, plus other sources, to estimate the parameters that are required to solve the value-based sequential model and we take the perspective of a research team which has been commissioned to run the trial on behalf of the NIHR. This means that we assume that the decision to spend the fixed costs of the trial – these include the set-up costs of the trial, plus the writing-up costs, plus any other costs which are incurred independently of those which result from patient recruitment and monitoring – has already been made by the NIHR.[5] Savings resulting from stopping the trial early are therefore restricted to savings in variable costs alone. This constitutes a conservative approach to the analysis and it could underestimate the cost savings that are offered by the value-based sequential design. However, in the absence of guidance about how to cost sequential clinical trials, it was believed to be the best approach that was available.

1.4.2 The three designs that are compared in this report

As already noted in the introduction to this section, the CACTUS and HERO case studies compare the performance of the value-based sequential design with two alternative, fixed sample size, designs:

1. the original trial design;

2. the fixed sample size Bayesian decision-theoretic design referred to above, using a normal prior–normal likelihood model, whose sample size maximises the expected net benefit of sampling (the 'value-based one stage design', see Claxton and Posnett (1996) and Claxton (1999)).

The value-based sequential design that we solve requires a maximum sample size to be set, prior to the trial commencing. For comparability with the value-based one stage design, our base case value-based sequential design sets its maximum sample size equal to the optimal sample size of the value-based one stage design. To test the sensitivity of our results to the choice of maximum sample size, we also consider variants of this model in which the maximum sample size is varied.

1.4.3 Assessing the performance of the three trial designs

We assess the performance of the three designs by using what we term 'operating characteristics', the main ones being:

1. *The sample size/expected sample size*. For the original trial design, this is just the trial's sample size (sometimes expressed as the sample size of the trial divided by two, so that

[5] Section 4.3.2 discusses fixed and variable costs in more detail.

the unit of measurement is the number of pairwise allocations. So, for example, the ProFHER pragmatic trial planned to recruit 250 patients, which equates to 125 pairwise allocations (125 randomised to surgery and 125 to sling)). For the value-based one stage design, the sample size that is used is equal to twice the number of pairwise allocations which maximise the expected net benefit of sampling. The net benefit of sampling is defined as the expected value of sample information (EVSI) minus the total variable cost of sampling (Claxton and Posnett, 1996; Claxton, 1999). For the value-based sequential design, the number of pairwise allocations is an expected value based on the results of a bootstrap analysis. The bootstrap analysis uses random draws from the trial's data to generate many artificial 'trials'. For each such trial, a path for the posterior mean for the expected value of incremental net monetary benefit is compared with the Stage II stopping boundary at a series of pre-defined interim analyses. Recruitment to the trial is deemed to have stopped, and pipeline patients are followed up, the first time the posterior mean lies outside this stopping boundary. Pipeline patients are then followed up and the final value for the posterior mean is calculated. The expected sample size is the average value of the sample size across all of the bootstrapped trials.

2. *Expected cost*. The expected cost of the original trial design is the final estimated cost of the trial, based on information from the trial's accounts. The expected cost of the value-based one stage design sums the trial's fixed costs and the total variable costs, defined as the optimal number of pairwise allocations (calculated in 1 above) multiplied by the estimated marginal cost per pairwise allocation. The expected cost of the value-based sequential design sums the trial's fixed costs and the total expected variable costs, defined as the expected number of pairwise allocations (calculated in 1 above) multiplied by the estimated marginal cost per pairwise allocation. We assume that the fixed costs and marginal cost per pairwise allocation are the same, regardless of the design of the trial under consideration.

3. *Expected net benefit*. The expected net benefit of a trial measures the additional benefit which is expected to accrue to patients as a result of choosing the preferred (cost-effective) health technology at the end of the trial, minus the cost of carrying out the trial. The net benefits of each technology are defined in Eqs. (1.1) and (1.2), so that the additional net benefit is given by the difference between the two (the 'incremental net monetary benefit', see Eq. (1.3)). The value of the trial in reducing uncertainty uses the Bayesian decision-theoretic perspective referred to at the start of this section and is measured by the expected value of sample information (EVSI). For the purposes of this report, the expected net benefit of the Big CACTUS trial and the HERO trial is equal to the expected value of sample information minus the total variable cost of the trial, eval-

uated at the actual sample size of the trial. The expected net benefit of the value-based one stage design is equal to the expected value of sample information minus the total variable cost of the trial, evaluated at the optimal number of pairwise allocations (using the methods of Claxton and Posnett (1996); Claxton (1999)). The expected net benefit of the value-based sequential design is equal to the expected value of entering Stage II of the sequential trial, evaluated at the posterior mean for the expected value of incremental net monetary benefit, minus the cost of reaching Stage II (the variable cost per patient pair multiplied by the number of patients recruited to the trial when Stage II commences). Further details of these calculations are provided in the case study chapters.

4. *The estimated probability that the new technology is cost-effective*. The estimated probability of concluding that the new technology is cost-effective is obtained from the bootstrap analysis by dividing the number of 'trials' for which the final value of the posterior mean is greater than zero by the total number of 'trials' used for the bootstrap.

1.4.4 Computer code

Matlab code to obtain the stopping boundaries for the value-based sequential design, as well as the equivalent of Figure 1.2 for a particular application, is available from Chick et al. (2017b).

1.5 Value-based adaptive designs which are not the focus of this report

We conclude this section by noting that this report considers a specific kind of value-based sequential design, namely, the one proposed by Chick et al. (2017a) and Alban et al. (2020). This is appropriate for a two-armed clinical trial which measures the cost-effectiveness of the health technologies under consideration and which assumes that Bayesian updating takes place with a normal prior and a normal likelihood.

We note that there is a growing literature on other value-adaptive models which can fit alternative types of clinical trial. These are not considered in this report. However, we briefly review them in Chapter 4.

Chapter 2

The CACTUS case study

Laura Flight, Alan Brennan and Stephen Chick

With thanks to Cindy Cooper, Munya Dimairo, Steven Julious, Nick Latimer and Rebecca Palmer for providing comments on early drafts of the chapter and sharing their expertise from conducting the CACTUS pilot and Big CACTUS trials.

2.1 Introduction

2.1.1 Background and aims

The Big CACTUS trial evaluated the clinical and cost-effectiveness of a computerised speech and language therapy (CSLT) in patients with aphasia following stroke. The original target sample size was 95 participants per arm (a total of 285 for the three arms). The study recruited 278 participants in total between October 2014 and August 2016 to three treatment arms

1. computerised speech and language therapy (CSLT),
2. usual care (UC),
3. attention control (AC).

Follow-up was at 6-months for the clinical outcomes and 12-months for the economic outcomes. The trial was funded by the National Institute for Health Research (NIHR) Health Technology Assessment (HTA) programme (HTA - 12/21/01) with a budget of £1,422,284 to cover research costs[1]. The long-term cost-effectiveness of the CSLT was assessed using a model based cost-utility analysis. A full description of the model is given in Section 2.1.4.

[1]Taken from https://www.journalslibrary.nihr.ac.uk/programmes/hta/122101/. Last accessed: 14Aug20

The Big CACTUS trial showed that CSLT led to significant improvements in word-finding ability that were maintained irrespective of time post stroke. However, these improvements did not generalise to conversation or participant perceptions of communication participation and quality of life. The cost-effectiveness analysis used a final data set with n = 270 participants and involved multiple imputation of some missing data. The main findings suggested that CSLT is unlikely to be considered cost-effective for the whole population investigated but may be cost-effective for people with mild and moderate word-finding difficulties (Palmer et al., 2020; Latimer et al., 2020).

As mentioned in Chapter 1, the Big CACTUS trial has two features of interest: it was preceded by a pilot trial and it used a model-based, rather than a within-trial, cost-effectiveness analysis. The base case analysis for this case study uses pilot trial data to define the prior distribution for expected INMB before the main trial, measures research costs as forecast in the research proposal, according to those that are covered by the funder (see chapter 4 for more details) and uses the multiple imputation data set with $n = 270$ participants as described by Palmer et al. (2020). For simplicity our modelling and re-analysis of the Big CACTUS trial is based on the comparison of the CSLT and UC arms only (ignoring the AC arm). This comparison reflects the potential new intervention compared with standard practice in the NHS setting.

The aims of the CACTUS case study presented here are to:

1. generate a value-based sequential design, that could have been used for the Big CACTUS trial in place of the fixed sample design using data only the information available immediately prior to the start of the main study, including from the CACTUS pilot trial.

2. simulate the possible results of the value-based sequential design and quantify the Expected Value of Sample Information (EVSI) for that design and compare that to the EVSI of different possible fixed sample size designs

3. incorporate research costs to the analysis and quantify the Expected Net benefit of Sampling (ENBS) (EVSI minus expected research costs) for both the value-based sequential design and the fixed sample designs

4. layer in the actual dataset from the Big CACTUS trial and undertake a what-if process imagining analysing data as it actually accrued over time using the value-based sequential design methods to see at what point the Big CACTUS would have been stopped.

2.1.2 The CACTUS research programme

The CACTUS research programme involved

- a pilot trial to assess the feasibility of conducting a large scale trial into the effectiveness of self-managed computer treatment for people with long-standing aphasia post stroke
- an early economic evaluation based on the results of the pilot trial
- a full scale randomised controlled trial Big CACTUS with three arms and clinical evaluation
- an economic evaluation using a health economic model to evaluation the cost-effectiveness of CSLT based on the full evidence from the Big CACTUS trial.

The following sections discuss each of these elements in turn.

2.1.3 The CACTUS pilot trial and early economic evaluation

The CACTUS (Cost-effectiveness of Aphasia Computer Treatment Compared to Usual Stimulation) pilot trial assessed the feasibility of conducting a large scale trial into the effectiveness of self-managed computer treatment for people with long-standing aphasia post stroke (Palmer et al., 2012). The pilot was a single blind parallel group, randomised controlled trial. Participants received either CSLT, designed to improve word finding ability through language exercises for people with aphasia, or UC control. The pilot recruited 34 participants to each arm, with 28 participants followed-up at 5-months. The difference between CSLT and UC 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 CSLT.

Latimer et al. (2013) reported an early assessment of the likely long-term cost-effectiveness of CSLT based on the pilot trial data. The pilot health economic analysis concluded that CSLT had an estimated incremental cost-effectiveness ratio (ICER) of £3,058 per QALY compared with UC (INMB=£2,636 for a WTP threshold of £20,000 per QALY), based on a deterministic model analysis. The probability that CSLT was cost-effective at the £20,000 per QALY threshold was approximately 75.8%. However, the results were based on a small number of participants and a larger sample was required to reduce uncertainty in estimates of cost-effectiveness. It was concluded that a full scale randomised controlled trial was feasible and so the Big CACTUS trial was planned.

2.1.4 The Big CACTUS clinical evaluation

The Big CACTUS trial had co-primary outcomes of improvement in the number of words named correctly and functional communication measured from baseline to 6-month follow-up. A multiple linear regression model adjusting for baseline score (either word finding score or functional communication score depending on the model considered), aphasia severity and site

was used as the primary analysis. The primary analysis set was based on a modified intention to treat principle. In this report we consider the multiple imputation (MI) analysis set with 270 participants[2], that was considered as a sensitivity analysis by Palmer et al. (2020). This choice is justified by the relatively low levels of missing data for the primary outcomes and EQ-5D data (approximately 20%) and that the Big CACTUS results were found to be consistent between the different analysis sets considered.

2.1.5 The Big CACTUS health economic model and economic evaluation

The QALYs of participants at one year were estimated using an accessible EuroQol-5 Dimensions, five-level version. This was combined with the measurement of treatments costs. A five state Markov model was used to conduct a cost-utility analysis, from the UK NHS and personal social services perspective following National Institute for Health and Care Excellence (NICE) guidance (NICE, 2013). Data from Big CACTUS were used to inform key model parameters, including transition probabilities and health state utilities. The base case health economic analysis used a probabilistic model analysis using MI trial data (as reported by Palmer et al. (2020); Latimer et al. (2020)). This analysis estimated an ICER of £42,686 per QALY gained for CSLT compared to UC (INMB=-£388.73 for a £20,000 per QALY WTP threshold). For mild and moderate word-finding difficulty subgroups the ICER was £22,371 (INMB= -£69.49) and £21,262 (INMB= -£48.44) per QALY, respectively. The study concluded that CSLT is unlikely to be cost-effective in the whole population but could be cost-effective in patients with mild and moderate word finding difficulties especially when considering a £30,000 per QALY threshold. The contrasting conclusions around the potential cost-effectiveness of CSLT between CACTUS pilot trial and the final Big CACTUS highlight the high levels of uncertainty in the exploratory pilot analysis that was based on a small sample size.

2.2 Methods

2.2.1 Overview of methods for the re-analysis of clinical and health economics data using the Big CACTUS trial

For simplicity the re-analysis of the Big CACTUS trial is based on a comparison of the CSLT and UC arms only. This comparison reflects the potential new intervention compared with standard practice in the NHS. The health economics analysis presented in this chapter is based

[2]This simplifying assumption is made to allow us to demonstrate the first application of the value-based sequential trial using a model-based health economic analysis. In reality, a researcher will need to decide whether to conduct the multiple imputation at each interim analysis of the trial. This is discussed in Chapter 3 in relation to the HERO case study.

on a replication in R of the original model developed specifically for the purposes of the ENACT project. The Latimer et al. (2020) Big CACTUS health economic model was recoded from Excel or R in order to undertake efficiently our analysis. Simplifications have been made during this process compared to the original health economic analysis because this enables a more tractable probabilistic sensitivity analysis and enabled us to illustrate the value-based sequential design approach.

Table 2.A.1 shows the results are almost identical; the R recoded model with the original Excel model produce very similar results for costs and QALYs in each arm of the comparison. The R recoded estimated an ICER of £42,582 per QALY gained for CSLT compared to UC (INMB=-£423 for a £20,000 per QALY WTP threshold).

2.2.2 Context for selecting and comparing designs for the CACTUS two-arm trial

To illustrate how the value-based sequential design can be used to guide the design and conduct of a trial we imagine that we are at the end of the CACTUS pilot, planning the full scale trial with a cost-effectiveness analysis. We refer to this as the CACTUS trial to distinguish it from the CACTUS pilot and Big CACTUS trials. Our main focus when designing the CACTUS trial will be to assess the clinical and cost-effectiveness of the CSLT compared to UC since this is the primary comparison of interest as per the Big CACTUS trial design.

We consider the perspective of the research team prior to commencing the trial, but post trial commissioning. Hence, we assume that a decision to commission the research and commit the fixed costs of the trial has already been taken. This includes all costs that are incurred in the set up, conduct and analysis of the trial that are incurred regardless of the trial design. Analysis therefore assumes zero fixed costs and obtains the optimal design accounting for the variable costs (per pairwise allocation) of the trial only.

2.2.3 Specifying the prior distributions required for the designs based on the pilot trial

We inform parameter value choices using only the data that were available to the investigator at the end of the CACTUS pilot trial. To implement the value-based approach we define a number of key parameters, summarised in Table 2.1 and discussed in detail in Appendix 2.B. These parameters include data from the CACTUS pilot and assume a reference prior before the pilot began to inform the parameters of the prior distribution of INMB.

We assume a WTP threshold of £20,000 per QALY gained to illustrate the methods, but acknowledge that NICE typically consider a range of between £20,000 and £30,000 per QALY

and that other researchers suggest other values for this threshold (Claxton et al., 2015).

2.2.4 Estimating the costs of sampling per study participant

To estimate the cost of conducting the research, referred to as the cost of sampling (Willan and Kowgier, 2008), for the CACTUS case study, the costs provided in the Big CACTUS grant application for the three arm trial are used. This gives an approximate fixed cost for the trial and a variable cost per patient randomised. For simplicity, we make the same assumptions as the Big CACTUS trial in terms of number of centres, required sample size per arm and length of recruitment even though we are focussed on the two arm comparison between CSLT and UC. As summarised in Table 2.1, and based on the three arm trial approximately £581,765 of the £1,445,565 budget was spent before the first participants were recruited in trial set up costs and £193,044 was spent after follow-up of the final patient. These before and after costs are assumed to be zero in the following calculations as we assume the research has already been commissioned and these costs have been committed (see Section 2.2.2). The resulting estimates were that £1,647 was spent on each patient during recruitment and £706 during follow-up giving a cost per patient of £2,353.

A two arm CACTUS trial that plans to recruit 95 participants per arm (CSLT and UC) has a planned budget of £1,221,879 (including fixed costs) and a variable cost component of £447,070 (£2,353 $\times$ 95 participants $\times$ 2 arms). [3]

Unlike the approach taken in the HERO and ProFHER case studies, this allocates all costs incurred during this time to a per patient cost such as staff costs, indirect costs as well as randomisation and data collection costs.

2.2.5 Expected value of sample information (EVSI) for fixed sample size designs (a value-based one stage approach)

A 'one-stage' value of information analysis can be used to establish whether it is worthwhile conducting further research, and if it is, to compare different designs for fixed sample size trials (Claxton and Posnett, 1996; Claxton, 1999). As described in Section 2.2.2, we want to plan a clinical trial to investigate the clinical and cost-effectiveness of the CSLT compared to UC. The expected value of perfect information (EVPI) is calculated using the standard Normal loss function (Raiffa and Schlaifer, 1961) and the parameter values in Table 2.1 to give an individual-level EVPI of £22.73. The total EVPI (for a population to benefit of 215,378) is £4.9 million. This suggests that it could be worthwhile conducting further research into

[3] This does not include the AC arm that was considered in the Big CACTUS grant application that estimated the costs of the trial to be £1,445,567 described in Section 2.3.3.

Parameter	Definition	Value	Source
P	Population expected to benefit from adoption decision	215,378 patients	Palmer et al. (2020)
σ_X	Standard deviation for INMB in population	£4600.51	Appendix 2.B
n_0	Effective sample size of the prior distribution	7	Pairwise allocations, CACTUS pilot Palmer et al. (2012)
μ_0	Prior mean for expected INMB	3190.42	Re-analysis of CACTUS pilot model-based analysis (based on Latimer et al. (2013)
Δ	Delay for observing INMB endpoint (in years)	1	Palmer et al. (2020)
	Estimated annual rate of recruitment to trial (allocations per arm)	55 allocations	Big CACTUS protocol
τ	Delay for observing INMB endpoint (allocations per arm)	55 allocations	Annual rate of recruitment
	Time horizon of trial	639 days	Big CACTUS protocol
I_{N}	Fixed cost of adopting CSLT	£0	Assumption
	Estimated spend on fixed costs prior to starting trial	£581,765	Big CACTUS grant application
	Estimated spend during recruitment (per patient)	£1,647	Big CACTUS grant application
	Estimated spend during follow-up (per patient)	£706	Big CACTUS grant application
	Estimated spend on fixed costs post follow-up	£193,044	Big CACTUS grant application
c_{fixed}	Total spend on fixed costs	£774,809	Big CACTUS grant application
	Total spend	£1,445,565	Big CACTUS grant application
c	Estimated cost per allocation to three (two) arms	£7,059 (4,706)	Big CACTUS grant application
λ	Maximum willingness to pay for one QALY	£20,000	NICE (2013)

Table 2.1: Parameter values for the CACTUS case study. CSLT: Computerised Speech and Language Therapy, INMB: Incremental Net Monetary Benefit, QALY: Quality Adjusted Life Year, UC: Usual Care

the comparison of CSLT and UC if the costs associated with the research do not exceed the population EVPI.[4]

The expected value of sample information (EVSI) is calculated for a range of fixed sample size designs with increasing sample sizes to determine the most cost-effective fixed sample size design for the comparison of CSLT to UC. These trial designs do not allow for the early termination of the trial, hence have just one stage. The ENBS is calculated for each of these sample sizes by subtracting an estimate of the cost of sampling. The design with the highest ENBS is considered to be optimal (a cost-effective use of resources), from a non-sequential perspective to running the trial.

2.2.6 Computing the stopping boundaries of the value-based sequential approach

We use the methodology outlined in section 1.4.1 to obtain the stopping boundary for the value-based sequential model. This uses the `Matlab` code of Chick et al. (2017a).

2.3 Results

In this section we first summarise the patterns of the data from the Big CACTUS trial as it accrued during the trial for the clinical, QALY and cost data. We then consider how the proposed research budget for the Big CACTUS trial would be spent during the study period and how the model-based estimate of cost-effectiveness of CSLT compared to UC changed over the duration of the study. This is followed by the results of the one-stage EVSI analysis giving the optimal sample size for a fixed design based on the costs and benefits of conducting the research and an illustration of the stopping boundary for value-based sequential designs with different maximum sample sizes. The final sections then compare the operating characteristics of the original fixed sample size design, the value-based one stage design and the value-based sequential designs.

2.3.1 Analysis of the accruing clinical outcome data

Figure 2.1a plots the estimate from accruing trial data, as observations on each pairwise allocation are observed, of the difference in the improvement in words named correctly at the 6-month

[4]These results differ from the original CACTUS pilot EVPI analysis reported by Latimer et al. (2013) because of the different analytical methods. Latimer used Monte Carlo simulation with 500 inner loops and 100 outer loops. Our aim is not to directly replicate this analysis, but use the case study to illustrate how the value-based sequential approach can be used in similar settings.

follow-up between CSLT and UC using MI data, together with 95% confidence intervals. Values above zero show CSLT to be superior to UC. The figure shows that word naming ability was higher for CSLT than UC on average throughout follow-up. Figure 2.1b plots the cumulative estimate of the difference in the improvement in functional conversation at the 6-month follow-up between CSLT and UC. Values above zero show CSLT to be superior to UC. The figure shows that functional conversation was lower for CSLT than UC on average throughout follow-up. The trial's final result showed that CSLT resulted in a clinically significant improvement in word naming ability (MI dataset: 16.44 words, 95% CI: 13.40 to 19.49). However, the intervention did not improve functional conversation when compared to UC (MI dataset: -0.05 95% CI: -0.20 to 0.10) (Palmer et al., 2019, 2020).[5]

2.3.2 Analysis of accruing cost-effectiveness data

Figure 2.2a shows the cumulative estimate of the difference in the model-based QALY between CSLT and UC using a deterministic analysis and MI data, at 12-months follow-up. It shows a small positive increase in QALY for participants in the CSLT arm compared to UC.

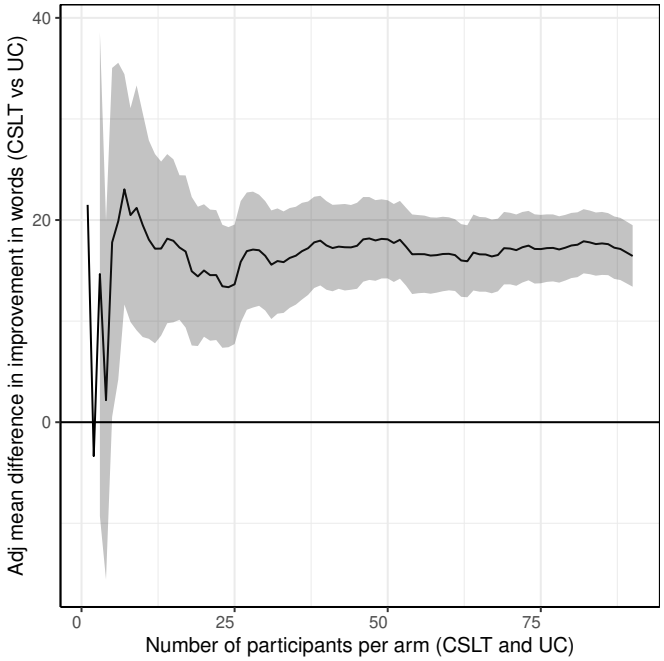
Figure 2.2b shows the cumulative estimate of the difference in model-based costs between CSLT and UC using a deterministic analysis and MI data. It shows that CSLT is more expensive than UC throughout the follow-up period. As shown in Table 2.A.1, once data from all 270 participants were collected, the `R` recoded model showed the difference between CSLT and UC costs was estimated to be £726.58 and the difference in QALYs was 0.0171.

2.3.3 Profile of research budget spend and cost-effectiveness data

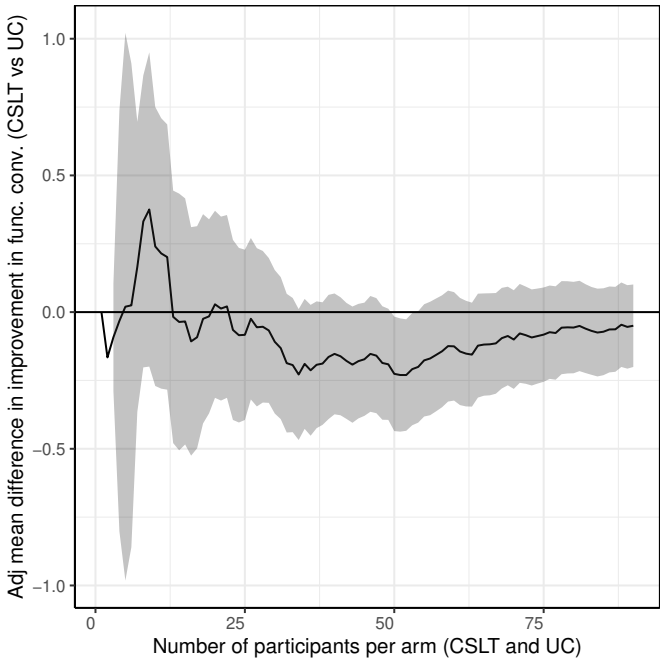
Figure 2.3 plots the projected cumulative spend of the Big CACTUS research budget over the course of the project (scale on left axis), using information from the Big CACTUS grant application based on a three arm trial. Although the actual spend could be estimated from the Big CACTUS expenditure (as in the HERO case study), this information would not be available to researchers when planning a trial. We have estimated this total projected expenditure for the Big CACTUS trial to be £1,445,567.[6] This includes costs incurred before, during and after recruitment. We have assumed there would be a constant recruitment rate of one patient per centre per month over a 21-month period until a target sample size of 95 allocations to CSLT, UC and AC are reached, based on the original frequentist design for the trial. The costs considered include research costs covered by the NIHR. Other costs, such as those incurred by the NHS to provide research support and treatment costs, are not considered. Chapter 4

[5]The analysis conducted for the CACTUS case study analysis has small differences compared with the results reported by Palmer et al. (2020) as we have used a simple replication of the approach taken rather than the comprehensive original analysis.

[6]This is the value taken from the grant application that is slightly different to the final amount reported

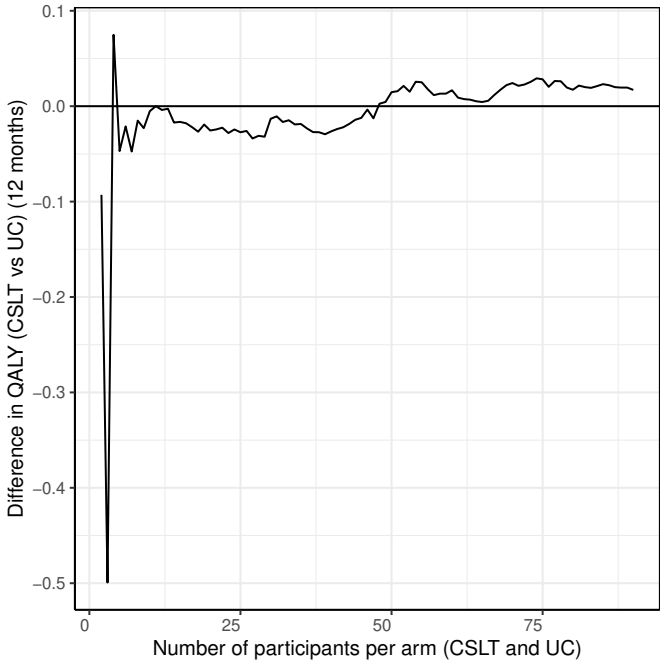


(a) Difference in improvement in words named correctly between CSLT and UC

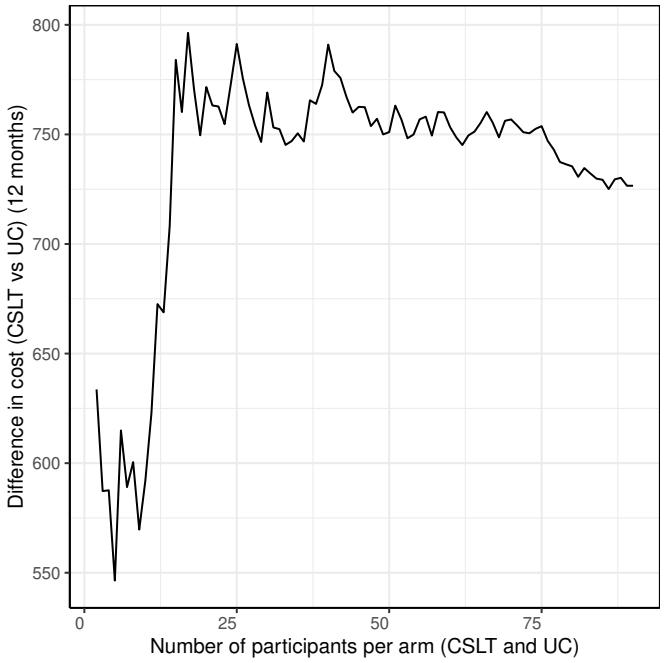


(b) Difference in improvement in functional conversation between CSLT and UC

Figure 2.1: Clinical outcomes - adjusted estimates from accruing trial data of the differences in clinical outcomes at the 6-month follow-up between CSLT and UC with 95% confidence intervals. CSLT: Computerised Speech and Language Therapy, UC: Usual Care



(a) Incremental QALYs



(b) Incremental costs

Figure 2.2: Estimate of the difference between health economic outcomes from accruing Big CACTUS trial data for CSLT and UC. CSLT: Computerised Speech and Language Therapy, INMB: Incremental Net Monetary Benefit, QALY: Quality Adjusted Life Year, UC: Usual Care

discusses in more detail the different perspectives that might be taken when calculating these costs.

Figure 2.3 (scale on right axis) also plots the actual estimate of the model-based incremental net monetary benefit (INMB) at one year based on a deterministic model analysis and using accumulating, imputed data from the Big CACTUS trial after each patient pair reached the 12-month timepoint. Positive values suggest that CSLT is cost-effective compared to UC for a willingness to pay threshold of £20,000 per QALY. Key milestones in the project are denoted by the letters 'A' (recruitment starts), 'B' (recruitment finishes), 'C' (one year follow-up finishes) and 'D' (publication of results of the clinical evaluation). Apart from a spike in INMB based on data from the first 12 participants to be randomised, the cumulative estimate of INMB is consistently below zero suggesting that CSLT is not cost-effective compared to UC for a willingness to pay threshold (WTP) of £20,000 per QALY. The figure shows a steep increase in research costs before recruitment begins, followed by an increase in costs for the remainder of the trial at a smaller gradient.

2.3.4 The optimal sample size of the value-based one stage design

The results for EVSI and ENBS are shown in Figure 2.4. The graph plots the EVSI, ENBS and total variable costs as a function of the trial's sample size for the value-based one stage design. The optimal one stage design has a sample size of 132 pairwise allocations to the CSLT and UC arms (264 participants). The population EVSI for this $n = 132$ per arm fixed designs is £4,217,830 and ENBS value is £3,596,638. In comparison, the actual proposed design of $n = 95$ per arm has an estimated EVSI of £3,980,727 and an ENBS of £3,533,657.

2.3.5 Stopping boundaries for two possible value-based sequential designs ($n = 95$ and $n = 132$ per arm)

Figure 2.5 shows the stopping boundaries for the value-based sequential designs with $n = 95$ and $n = 132$ pairwise allocations. With a maximum sample size of $n = 95$ pair allocations (Figure 2.5a), there is a short window of opportunity, during Stage II of the value-based sequential design, to stop the trial early due to the long follow-up time for the health economic outcomes relative to the length of recruitment. Stage II is extended when we assume 132 allocations (Figure 2.5b) can be made and the length of recruitment is increased.

For large values of the prior mean for the expected INMB it is optimal to not run any trial and to make the treatment adoption decision based on current information. Figure 2.5 also shows that it is never optimal to run a value-based one-stage trial if the prior mean for the expected INMB exceeds A or B.

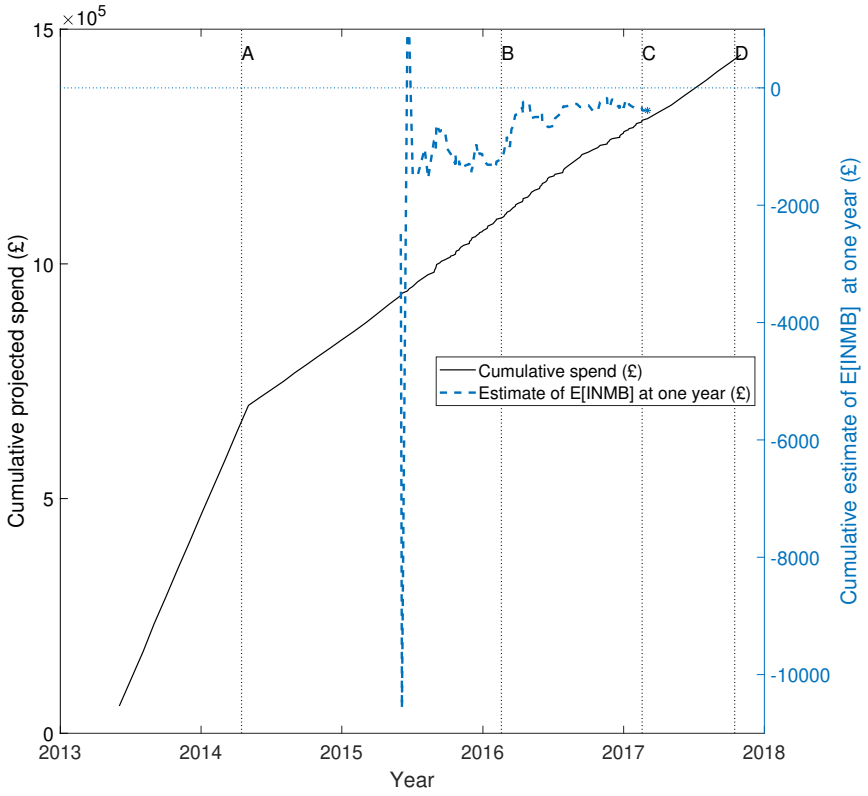


Figure 2.3: Cumulative budget spend over the course of the three arm Big CACTUS trial based on projected costs incurred by the funder (the NIHR) and the cumulative estimate of the expected INMB between CSLT and UC. Key milestones: 'A' - recruitment starts; 'B' - recruitment finishes; 'C' - one year follow-up finishes; 'D' - principal publication (Palmer et al., 2020)

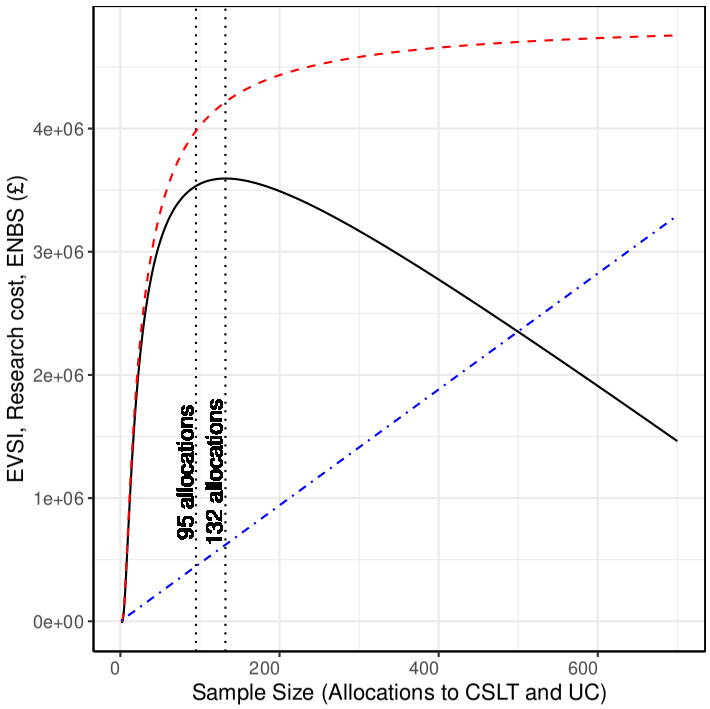


Figure 2.4: ENBS (black solid line), population EVSI (red dashed line) and total variable cost of sampling (blue dot-dashed line) for range of sample sizes for the CACTUS trial (pairs of patients randomised to CSLT and UC). ENBS: Expected Net Benefit of Sampling, EVSI: Expected Value of Sample Information, CSLT: Computerised Speech and Language Therapy, UC: Usual Care

2.3.6 The expected value of different research designs

Table 2.2 summarises the results in this section that compare the expected net benefit of the

1. fixed sample size design with 95 pairwise allocations (based on the frequentist sample size calculation of the Big CACTUS trial),

2. value-based one stage design (with 132 pairwise allocations, see Figure 2.4),

3. value-based sequential designs with a maximum of 95 to 792 pairwise allocations.

In each scenario, the same assumptions are made about the recruitment rate into the trial taken from the Big CACTUS trial (described in Appendix 2.B). Section 2.3.8 explores the characteristics of a selection of these designs in more detail using bootstrapping.

Table 2.2 shows that the value-based one stage design with 132 pairwise allocations provides a small (+1.7%) increase in expected net benefit of sampling (ENBS=£3.60m) compared to the original proposed 95 pairwise allocations sample size (ENBS=£3.54m) based on the frequentist sample size calculation of the Big CACTUS trial. This was illustrated by the dashed vertical lines earlier in Figure 2.4 and shows that the original sample size is close to the peak

Trial design	Number of pairwise allocations	EVSI/EVVBS (£) Individual	EVSI/EVVBS (£) Population	Variable cost (£) Total	Variable cost (£) To Stage II	Expected net benefit (£)	Percentage increase in expected net benefit
1. CACTUS case study	95	18.49	3.98×10^6	4.48×10^5	-	3.54×10^6	-
2. Value-based one stage	132	19.58	4.22×10^6	6.21×10^5	-	3.60×10^6	1.7
3. Value-based sequential	132	19.10	4.11×10^6	-	2.59×10^5	3.85×10^6	9.0
4. Value-based sequential	264	20.38	4.39×10^6	-	2.59×10^5	4.13×10^6	16.8
5. Value-based sequential	396	20.78	4.48×10^6	-	2.59×10^5	4.22×10^6	19.3
6. Value-based sequential	528	20.96	4.51×10^6	-	2.59×10^5	4.26×10^6	20.4
7. Value-based sequential	660	21.05	4.53×10^6	-	2.59×10^5	4.28×10^6	20.9
8. Value-based sequential	792	21.11	4.55×10^6	-	2.59×10^5	4.29×10^6	21.3
9. EVPI	∞	22.73	4.90×10^6	-	-	-	-

Table 2.2: Performance of different trial designs from a value-based perspective. EVPI: Expected Value of Perfect Information, EVSI: Expected Value of Sample Information, EVVBS: Expected Value of Value-Based Sequential design.

Notes: The column headed 'Number of pairwise allocations' records the actual number of pairwise allocations made in the Big CACTUS trial (row 1), the optimal sample size for the one stage value-based design (row 2) and the maximum number of pairwise allocations that could be made in the value-based sequential trials (rows 3 to 8). The columns headed 'EVSI/EVVBS' record the expected value of sample information (EVSI) for the designs in rows 1 and 2 and the expected value of the value-based sequential (EVVBS) design for the designs in rows 3 to 8. The column headed 'Variable cost' records the unavoidable variable cost of sampling. This is the total variable cost for the fixed sample size designs and the variable cost of reaching Stage II (the point where data start to be observed for the first participants in the trial) for the value-based sequential designs. The column headed 'Expected net benefit' records EVSI minus the total variable costs for the fixed sample size designs and EVVBS minus the variable costs to Stage II for the value-based sequential designs. The final column records the percentage increase in expected net benefit of the relevant value-based design over the expected net benefit delivered by the Big CACTUS trial listed in row 1.

of the ENBS curve but that $n = 132$ pairwise allocations could provide a marginally greater expected net benefit.

Table 2.2 also shows the results of the value-based sequential designs, with rows three to eight showing results for the value-based sequential design with a capped maximum sample size ranging from 132 pairwise allocations up to seven times higher (792 pairwise allocations). The value-based sequential designs show a marked increased in expected value compared with the fixed designs. The value-based sequential design with a maximum of 132 pairwise allocations has an expected value of £3.85m which is $+9.0\%$ higher than the $n = 95$ pair fixed design and $+6.9\%$ higher than the $n = 132$ pair fixed design.

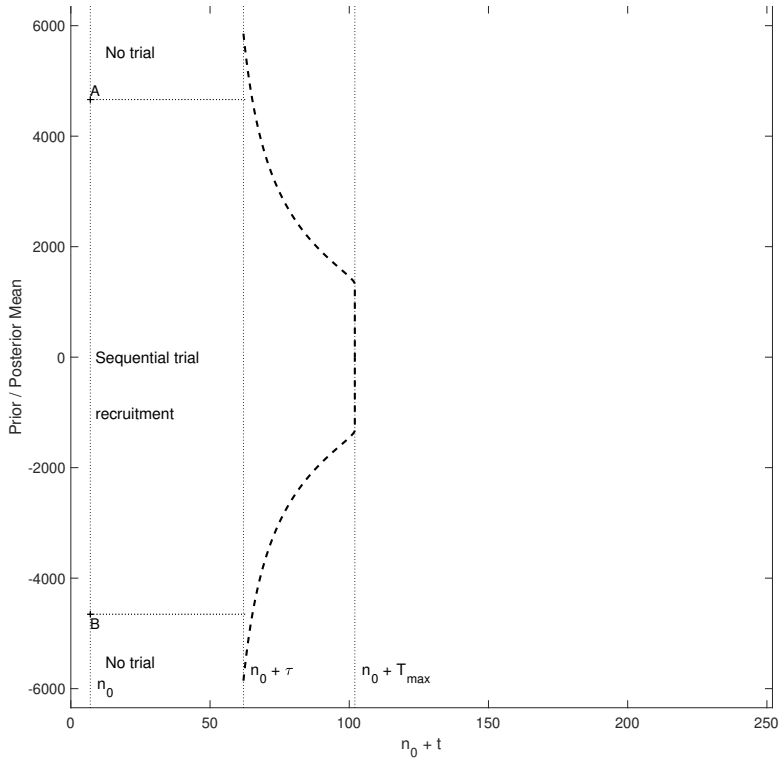
Table 2.2 also shows increasing expected net benefit of sampling for value-based sequential designs as the maximum number of pairwise allocations is increased. By 396 pairs, the ENBS is £4.22m ($+19.3\%$ higher than the original $n = 95$ pair fixed design). There are diminishing returns however, and further increases in the maximum sample size do not greatly increase the expected net benefit. This is explained by a diminishing marginal additional value of further data to decision making as measured by the expected value of the value-based sequential design as we have higher sample sizes and approach the EVPI but continuously increase research costs (similar to what we see in Figure 2.4 for fixed sample size EVSI and ENBS in the value-based one stage designs).

2.3.7 When would the CACTUS case study value-based sequential design trial have stopped?

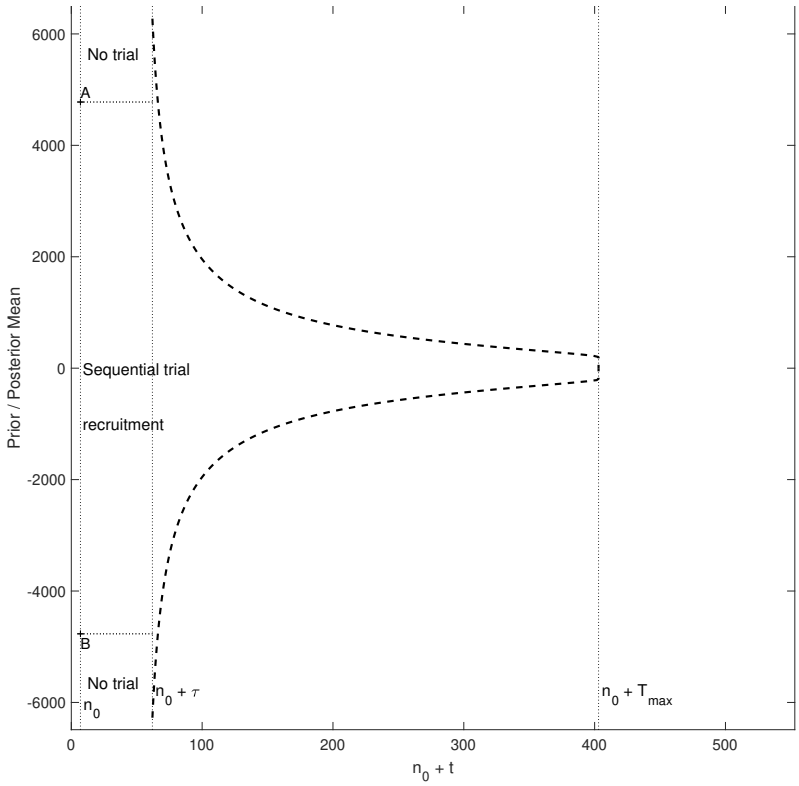
In this section we examine what would have happened if the Big CACTUS trial had been designed as a value-based sequential design with two arms. Three scenarios are examined for the maximum sample size in terms of allocated pairs ($n = 95, n = 132, or n = 396$). In each case we consider what would have happened as the actual accrued data from the real CACTUS trial came in.

There were some methodological issues to address in implementing this. When applying the value-based sequential approach the data can be analysed as frequently as after every patient pair (a purely sequential design) or after groups of patients have reached the outcome of interest (similar to a group sequential design). While a purely sequential design offers a greater opportunity to stop the trial as soon the stopping boundary has been crossed and the maximum value, it may be unrealistic in practice. In the CACTUS case study the health economic model is evaluated at each analysis, while the computation burden is small when considering the CACTUS trial data only, when evaluating different designs using bootstrapping of the individual trial data (see Section 2.3.8) this becomes infeasible. Here we consider a block size of five allocations to the CSLT and UC arms.

The results are shown in Figure 2.6. This plots the cumulative path of the expected INMB based on the Big CACTUS trial data for CSLT and UC with the stopping boundaries for the value-based sequential designs with a maximum of 95 (red), 132 (blue) and 396 (purple) allocations. The CACTUS sample path (black line) does not cross the stopping boundaries during the trial indicating that if the value-based sequential design had been used, there would never have been a point where it was optimal to stop the CACTUS case study trial early. A possible explanation for this is due to the small differences in QALY observed during the trial (see Figure 2.2a) and the relatively small cost difference of the CSLT intervention. This is in contrast to the ProFHER trial, where there was strong evidence that the expected INMB was positive. For CACTUS the trial may have stopped early had we observed a large difference in QALY or costs giving a strong positive or negative INMB and consequently a large probability of cost-effectiveness for one of the treatments. Instead the INMB estimate remained close to zero and within the stopping boundary, signalling that as the data emerged the value-based sequential analysis would have shown CSLT to be around the borderline of £20,000 per QALY gained, and that therefore collecting more data to reduce uncertainty and clarify which side of that borderline the cost-effectiveness of CSLT versus UC is would be useful.



(a) 95 pairwise allocations



(b) 132 pairwise allocations

Figure 2.5: Stopping boundaries for the value-based sequential model using the Big CACTUS parameter values based on a maximum sample size of 95 allocations (from the original Big CACTUS fixed design) and 132 allocations to illustrate the performance of the value-based sequential design not constrained by the fixed sample size calculation

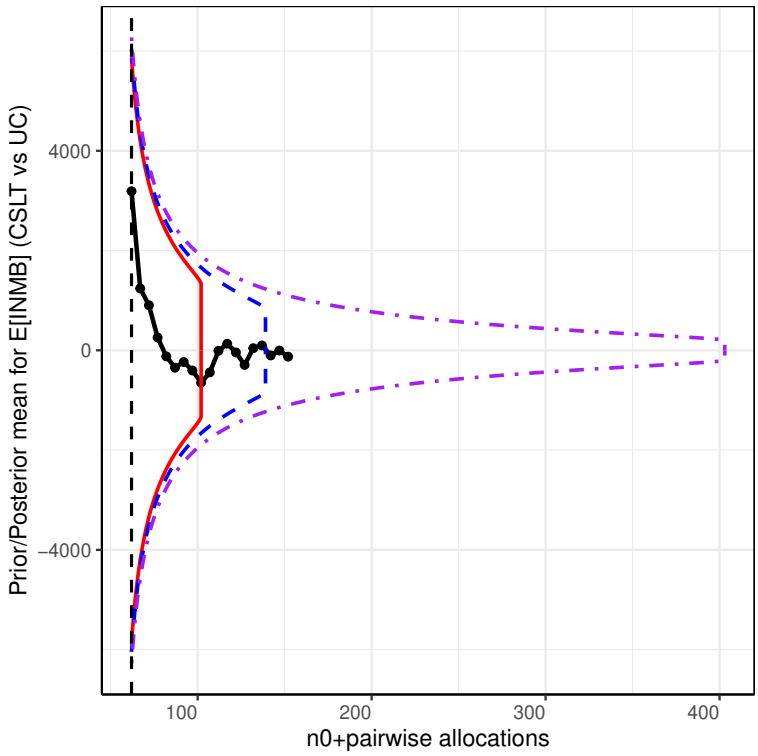


Figure 2.6: Stopping boundaries for the value-based sequential model for a maximum sample size of 95 (red- solid), 132 (blue- dashed) and 396 (purple - dotdashed) pairwise allocations where n_0 is the effect sample size of the prior information. CSLT: Computerised Speech and Language Therapy, INMB: Incremental Net Monetary Benefit, UC: Usual Care

2.3.8 What else could have happened? Results from simulating 5000 possible trials

In the previous section we saw how a value-based sequential design would have turned out given the actual data accrual pathway in the Big CACTUS trial. In this section, we explore what else could have happened. The actual pathway was just one realisation of how the trial data emerged. Here we explore how the data might have come in a different order over time. For example, the early data could have come in showing some people doing very well in the CSLT arm. Or, it could have happened that the first cases in the UC arm did unusually well. We explore how the random arrival of data showing good or not so good improvement in their outcomes could have made a difference to the value-based sequential results. Again, we examine the three scenarios for the maximum sample size in terms of allocated pairs ($n = 95, n = 132, n = 396$). In particular we look at the average sample size, average research budget and the average posterior mean expected INMB.

We use a non-parametric bootstrapping analysis to simulate possible trial results. The Big CACTUS trial data for CSLT and UC are sampled at random with replacement 5000 times independently for each treatment arm. The data are ordered so that one participant is randomised to each arm at each allocation. The cumulative health economic model parameters are then estimated from the accumulating data once there are five allocations to each arm. This gives the cumulative estimate of the model-based expected INMB during the trial. These data are combined with the prior estimate of the expected INMB to give the posterior mean expected INMB. Each bootstrapped sample path for the posterior mean is then compared with the value-based stopping boundary to determine when the trial would have stopped, the adoption decision that would have been made, the cost of the trial and the posterior mean expected INMB allowing for pipeline patients.

Table 2.3 summarises the operating characteristics of the value-based sequential and fixed sample size designs with the maximum sample sizes of 95, 132 and 396 allocations. For sample size the potential efficiency of the value-based designs are striking. For the value-based sequential design with a maximum sample size of 95 pairwise allocations, the average sample size is 81 allocations compared to 95 for the fixed design. The minimum sample size before deciding to stop is $n = 60$ pairs. For a maximum sample size of $n = 132$, the average sample is $n = 100$, and for a maximum of $n = 396$, the average sample size before stopping is $n = 179$. In each case the minimum sample size was $n = 60$.

The research budget savings (or extra budget required) are also of interest. For the value-based sequential design with a maximum sample size of $n = 95$ the average budget saving compared to a fixed design is estimated at £67,000 (approximately 15% saving). The peak saving which could occur is £165,000 when the trial stops at $n = 60$ pairs. For the larger

maximum sample size of the value-based sequential design on average there is a slightly higher expected research cost than the original $n = 95$ fixed design ($+£22,000$).

Table 2.3 also shows the effects on the estimated cost-effectiveness and its precision; how certain are we about the cost-effectiveness result differs for the different designs. All of the designs show that on average across the 5,000 simulation results there is a negative expected INMB for CSLT versus UC. For each design we would expect to find CSLT is not cost-effective compared to UC. The precision of the estimate depends on the design. The largest fixed design has a standard deviation in the posterior expected INMB of £282 across the 5000 simulations, whereas the value-based designs have two to three times greater standard deviation reflecting their lower average sample size. In a sense this reflects that a fixed $n = 396$ trial is 'overkill' in terms of sample size, because it would cost a much larger research budget (an extra £1,417,000 compared to $n = 95$ fixed design) and it be possible to decide on cost-effectiveness with a much lower sample size.

Table 2.4 shows the proportion of the 5000 simulated trials that cross the upper and lower stopping boundary as well as the proportion of times CSLT and UC are considered to be most cost-effective. Due to the delay in observing all the trial data after the trial crosses the boundary there is a small chance the simulated trial results stop because the lower boundary has been crossed, but once all the data are observed CSLT is declared cost-effective. The results suggest that increasing the maximum sample size increases the probability that UC is chosen as the most cost-effective of the two interventions. This is true for the value-based sequential and fixed sample size designs. The probability for the fixed sample size design with 95 pairwise allocations is 68%, similar to the probability of cost-effectiveness reported by the original Big CACTUS trial. This probability increases to 92% for a sample size of 396 pairs for the fixed design, again suggesting that there is a reduction in uncertainty from this larger sample size.

In the final three rows of Table 2.4 the probability that the equivalent fixed and value-based sequential designs result in the same technology adoption decision is above 75% for all designs. However, this decreases as the maximum sample size increases, suggesting that the designs diverge slightly in their decisions as the sample size of the fixed sample size design is much larger than the expected sample size of the value-based sequential design. It is important to keep in mind that while these larger fixed sample size designs may increase accuracy, they incur much greater costs.

	Average	SD	Min	Max
Max. sample size =95				
Value-based sequential				
Posterior mean for $\mathbb{E}(\text{INMB})$	-132	834	-5816	7107
Sample size (pairwise allocations)	81	15	60	95
Change in budget (£000) (% of total spend)	-67 (-15)	-70	-165	0
Fixed sample size				
Posterior mean for $\mathbb{E}(\text{INMB})$	-200	661	-3175	3919
Sample size (pairwise allocations)	95	0	95	95
Change in budget (£000) (% of total spend)	0 (0)	0	0	0
Max. sample size =132				
Value-based sequential				
Posterior mean for $\mathbb{E}(\text{INMB})$	-156	818	-4237	10048
Sample size (pairwise allocations)	100	32	60	132
Change in budget (£000) (% of total spend)	22 (5)	149	-165	174
Fixed sample size				
Posterior mean for $\mathbb{E}(\text{INMB})$	-261	544	-2455	4174
Sample size (pairwise allocations)	132	0	132	132
Change in budget (£000) (% of total spend)	174 (39)	0	174	174
Max. sample size =396				
Value-based sequential				
Posterior mean for $\mathbb{E}(\text{INMB})$	-213	772	-5039	9151
Sample size (pairwise allocations)	179	119	60	396
Change in budget (£000) (% of total spend)	394 (88)	562	-165	1417
Fixed sample size				
Posterior mean for $\mathbb{E}(\text{INMB})$	-399	282	-1436	824
Sample size (pairwise allocations)	396	0	396	396
Change in budget (£000) (% of total spend)	1417 (317)	0	1417	1417

Table 2.3: Operating characteristics for the CACTUS case study based on 5000 re-sampled paths for the comparison of the value-based sequential and fixed sample size designs with maximum sample sizes of 95, 132 and 396 pairwise allocations. INMB: Incremental Net Monetary Benefit, SD: Standard Deviation

	Final decision		
	CSLT	UC	Total
Max. sample size =95			
Value-based sequential			
First crossing lower part of boundary	0.08	0.39	0.47
First crossing upper part of boundary	0.27	0.25	0.52
Total	0.35	0.64	1.00
Fixed sample size			
Total	0.32	0.68	1.00
Max. sample size =132			
Value-based sequential			
First crossing lower part of boundary	0.07	0.45	0.52
First crossing upper part of boundary	0.27	0.21	0.48
Total	0.34	0.66	1.00
Fixed sample size			
Total	0.27	0.73	1.00
Max. sample size =396			
Value-based sequential			
First crossing lower part of boundary	0.05	0.58	0.62
First crossing upper part of boundary	0.23	0.14	0.37
Total	0.28	0.72	1.00
Fixed sample size			
Total	0.08	0.92	1.00
Fixed and value-based sequential designs with the same maximum sample size give the same decision			
Max. sample size =95			0.91
Max. sample size =132			0.87
Max. sample size =396			0.78

Table 2.4: Further operating characteristics for the CACTUS case study based on 5000 re-sampled paths for the comparison of the value-based sequential and fixed sample size designs with maximum sample sizes of 95, 132 and 396 pairwise allocations. CSLT: Computerised Speech and Language Therapy, UC: Usual Care.

2.3.9 Summary of key findings

The value-based sequential approach has been applied to the CACTUS case study to explore how the methods can be applied where data are available from a pilot trial and a model-based health economic analysis is required. We found that applying the value-based sequential design was feasible in this context. It was possible to use the pilot trial data to generate prior distributions required to implement the value-based sequential design. An existing health economic model was recoded into `R` software to enable the use of a model-based estimate of cost-effectiveness to be incorporated into the approach.

Under the assumption of a variable research cost per recruited sampled in the trial of £2,353 per participant, we were then able to generate traditional EVSI and ENBS calculations to show that a value-based one stage design of $n = 132$ pairs could have provided greater value to decision makers given its research costs than the originally chosen $n = 95$ pairs fixed design based on frequentist sample size calculations. More importantly, we were able to generate stopping boundaries for proposed value-based sequential designs in which accumulating data on cost-effectiveness would be monitored during the trial and used to determine whether the trial should stop or continue recruiting participants based on a trade off between the estimated cost-effectiveness of the intervention, uncertainty in that cost-effectiveness and the ongoing research costs.

The results showed that for each of the value-based sequential designs examined, the ENBS was greater than that of a fixed design of $n = 95$ or $n = 132$ pairs. That is, from a health economic perspective, the value-based sequential design with a maximum sample of $n = 95$, $n = 132$ or $n = 396$ would be expected to offer greater value for money to the NIHR than a fixed design. This is partly because in one circumstance, where it becomes apparent early on in the trial that the accrued evidence is showing one of the treatments is substantially more cost-effective than the other, then we could terminate the trial earlier and make a clear decision on future technology adoption without spending more research monies. And it is also partly because there can be a second circumstance, where the data accrued so far shows evidence that we are still uncertain and the cost-effectiveness of the intervention compared to usual care is borderline, and in this circumstance then we can flexibly collect that additional evidence and inform the decision on effectiveness and cost-effectiveness with greater precision.

In the context of the CACTUS case study, using the observed data on the CSLT and UC treatments from the Big CACTUS trial, had the trial been run as a value-based sequential design and the data accrued as it actually did, then the trial would not have stopped early. So the second of the circumstances discussed in the paragraph above would have been operating.

We were further able to show by simulating 5000 alternative possible trial results for the two-arm trial, that the value-based sequential design with maximum sample size of $n = 95$ pairs would have finished earlier and been cheaper on average than a fixed design with $n = 95$

pairs. But with a larger maximum sample size the value-based sequential design would have gone on slightly longer and cost marginally more than the fixed $n = 95$ pair design. But that this would have been 'worth it' to the NIHR and the NHS England, because it would have generated greater certainty about cost-effectiveness with only marginal additional research cost.

Whilst all the above is positive, it should be stressed that the absolute scale of differences between the value-based sequential and fixed designs are not large. There was not a huge potential for the CACTUS trial to stop early during the value-based sequential design. This is reflected in the probability of choosing each intervention (70%:30% UC:CSLT). There were fairly small differences between the interventions in terms of costs and benefits and so extreme results were not observed. This means savings were small (order of £10,000 to £100,000) when comparing the value-based sequential design and original fixed sample size design. However, the benefits of the value-based sequential approach are not solely focussed on stopping a trial early. Other benefits include a framework for comparing the value of fixed and value-based sequential designs and potentially running a trial for longer with the view to providing stronger evidence for a health technology adoption decision.

In summary, we have shown that the potentially larger sample size offered by the value-based sequential design can reduce the uncertainty of the INMB estimate whilst not incurring the same large costs as the equivalent sample size for a one stage design. This will help decision makers when informing a treatment adoption decision.

2.3.10 Limitations

Limitations to this analysis include some simplifications to the original cost-effectiveness modelling used in the Big CACTUS trial when it was recoded into `R` software. However, the results are very similar and we do not judge this to be of concern. Additionally, the real Big CACTUS trial had three arms and we have only operationalised the value-based sequential design methods on a two arm case study. Further work will explore how the value-based approach can be extended to the three arm comparison (discussed in Chapter 5).

Our analysis assumes that the research costs of interest in trading off the costs and benefits of different research designs are the variable costs of recruitment per participant. It is a strength that we have used actual research costs proposed from the original study to estimate this, since many EVSI studies make much simpler assumptions. Nevertheless, this was a simplification of this case study. In essence we have assumed that the fixed costs for a traditional fixed sample design for set up before the recruitment, and analysis afterwards would be identical if we instead ran a value-based sequential design. There are reasons to think this might not be true – it might require more analytical resource to undertake the data accrual analysis, there could be other fixed costs of finishing trials early not accounted for in the study. These are relatively

easy to incorporate into the analysis and do not represent methodological difficulties but they do mean our example and its results are somewhat stylised and further work could strengthen that side of the research costs analysis.

2.3.11 Implications for NIHR and stakeholders

The case study here, and the others in this report, signal that the methods for value-based sequential design are now able to be operationalised on real world case studies. NIHR, CTUs and all the stakeholders involved in the design, commissioning and delivery of trials will need to work together in concert for the potential benefits of value-based sequential designs to be realised in a practical way within existing structures.

The ENACT report has written a separate discussion on the various opportunities, challenges and implied recommended actions for the key players in this field including how commissioners, clinical trial designers, NHS, members of the public and HTA agencies can contribute or have awareness to develop a pathway to further testing and implementation of these methods in practice.

2.3.12 Generalisability of findings

This is a new case study. The specific patters of relative costs and benefits of fixed designs versus value-based sequential design would depend upon the interventions, the trial designs, the health economic model and the prior uncertainty in another condition/intervention of interest. The scale or direction of findings here cannot therefore be generalised.

The principles can however. It is clear that value-based sequential designs can offer the potential to end trials early or continue them longer depending on how the evidence accrues during the trial and therefore tailor the investment of research resources to areas where the reduction of uncertainty add value to decision making bodies such as NICE and other organisations deciding on health technology adoption on the basis of effectiveness and cost-effectiveness.

The methods demonstrated here apply equally well to any government funded healthcare system with government funded health research investment - they translate across international boundaries.

2.3.13 Summary

The CACTUS case study demonstrates further that the theories and methods of value-based sequential designs can be operationalised in a real world settings to design and analyse trials in which decisions are made as the data accrues.

Appendix

2.A Big CACTUS health economic model

The original pilot health economic model and the Big CACTUS health economic model were created in `Excel`. However, given the requirement to repeatedly run this model during the sequential trial the models were replicated in `R`. Table 2.A.1 provides a comparison of the cost-effectiveness results based on the deterministic analysis of the original model and the model recreated in `R` for this analysis using multiply imputed data.

	Outcome	**Intervention**	**Comparator**	**Difference mean**
Original deterministic analysis using MI data in `Excel`				
CSLT vs UC	Cost	728.50	0.00	728.50
	QALY	4.2179	4.2026	0.0153
			INMB	-423
			ICER	47614
CSLT vs AC	Cost	728.50	38.18	690.32
	QALY	4.2179	4.2022	0.0157
			INMB	-376
			ICER	43891
Deterministic analysis using MI data in `R`				
CSLT vs UC	Cost	726.58	0.00	726.58
	QALY	4.1984	4.1813	0.0171
			INMB	-385
			ICER	42582
CSLT vs AC	Cost	726.58	38.49	688.09
	QALY	4.1984	4.1808	0.0176
			INMB	-336
			ICER	39057

Table 2.A.1: Summary of the deterministic model results based on multiple imputed data for the original Excel model and the R model built to replicate this analysis. CSLT: Computerised Speech and Language Therapy, ICER: Incremental Cost Effectiveness Ratio, INMB: Incremental Net Monetary Benefit, MI: Multiple Imputation, QALY: Quality Adjusted Life Year, UC: Usual Care

2.B Big CACTUS parameter values

2.B.1 Calculating model parameters during the trial

In the CACTUS case study we assume that patients arrive to the trial in pairs and are randomly allocated one to each treatment (CSLT or UC). In Big CACTUS the random allocation was based on stratified block randomisation, a similar process could be adopted here. At early interim analyses, when data are only provided from one or two patient pairs there is insufficient information to calculate all of the model parameters. For example, if none of the participants have reached the 6, 9 or 12-month follow-ups there will be no data for those transition probabilities. If at a given analysis, there is not yet sufficient information to calculate a model parameter it is assumed to be zero. Palmer et al. (2020) describe the health economic model parameter values in detail.

2.B.2 Incidence rate (cases per annum)

According to the value of information analysis conducted at the end of the Big CACTUS trial, an average of 21,538 people could be treated with CSLT per year giving an estimate of 215,378 people over a 10-year period (Palmer et al., 2020).

2.B.3 Recruitment rate and delay

The proposed recruitment rate for the Big CACTUS trial was one patient per centre per month and that each centre would recruit participants for approximately 15 months (Palmer et al., 2015). A 21-month recruitment period is proposed for the CACTUS trial. It is assumed that the recruitment rate is constant across the duration of the trial and that all sites recruit in every month. Therefore 285 participants need to be recruited over the 21-month recruitment window (95 treatment allocations are made). Over a 12-month period (the length of delay in observing the health economic outcomes) 55 allocations are made. A constant recruitment rate is assumed for each of the designs considered. This means that doubling the number of participants would double the length of the recruitment period, doubling the costs incurred during this period. This implies that reducing the number of participants needed in the trial would proportionately reduce the cost of this stage of the trial (and therefore the overall costs of the trial).

2.B.4 Proportion treated with the 'new' technology at the start of the trial

It is assumed that the patients with aphasia were not being treatment with CSLT at the start of the trial based on correspondence with the chief investigator of Big CACTUS.

2.B.5 Specification of the prior distribution for the expected incremental net monetary benefit

Let X be the random variable denoting population-level INMB, which we assume has a Normal distribution with expected value μ_X and variance σ_X^2. Let x denote an observation of INMB such that

$$\bar{x}_m = \sum_{i=1}^{m} \frac{x_i}{m}, \tag{2.1a}$$

$$s_m^2 = \frac{1}{m-1} \sum_{i=1}^{m} (x_i - \bar{x})^2 \tag{2.1b}$$

where, for a sample comprising m pairwise allocations,

$$\bar{x}_m \text{ is the sample mean INMB,}$$
$$s_m^2 \text{ is the sample variance.}$$

We define the expected value and variance of the prior distribution as μ_0 and σ_0^2, respectively. Applying Bayes' rule for a Normal prior–Normal likelihood model, appropriate in this setting, the following expressions define the mean and variance of the posterior distribution for the expected value of INMB, given a sample size of m pairwise allocations and an effective sample size of the prior distribution equal to $n_0 = \sigma_X^2/\sigma_0^2$

$$\mu_m = \frac{n_0\mu_0 + m\bar{x}_m}{n_0 + m}, \tag{2.2a}$$

$$\sigma_m^2 = \frac{\sigma_X^2}{n_0 + m}. \tag{2.2b}$$

For the base case analysis, before the pilot study commenced, one possibility is to assume a reference prior distribution, meaning that $n_0 \to 0$ in Eq. (2.2a) and Eq. (2.2b) (see, e.g. Spiegelhalter et al. (1994b, Section 4.1.1)). Under this assumption, we obtain the following expressions for the posterior mean and variance after the pilot study

$$\mu_m = \bar{x}_m, \tag{2.3a}$$

$$\sigma_m^2 = \frac{\sigma_X^2}{m}, \tag{2.3b}$$

$$= \frac{s_m^2}{m}. \tag{2.3c}$$

where we assume that we can replace σ_X^2 with its unbiased estimator from the pilot, s_m^2. The CACTUS pilot trial had a sample size of $m = 7$ giving a sample mean $\bar{x}_m = \bar{x}_7$ and a sample

variance $s_m^2 = s_7^2$. These expressions become the prior mean and variance for the main trial

$$\mu_7 = \bar{x}_7, \tag{2.4a}$$

$$\sigma_7^2 = \frac{s_7^2}{7}. \tag{2.4b}$$

The effective sample size of the prior distribution for the main trial is therefore

$$n_7 = \frac{\sigma_X^2}{\sigma_7^2}, \tag{2.5a}$$

$$= \frac{s_7^2}{\frac{s_7^2}{7}}, \tag{2.5b}$$

$$= 7. \tag{2.5c}$$

Using the R based model of the CACTUS pilot analysis the prior mean $\mu_7 = \bar{x}_7 = 3190.42$. The posterior variance ($\sigma_7^2 = s_7^2$) can be calculated using the probability of cost-effectiveness. The probability of cost-effectiveness for a willingness to pay threshold of £20,000 per QALY is calculated as the probability that the INMB is greater than zero at this threshold value. This is calculated using

$$P(\mathrm{X} > 0) = 1 - \Phi\left(\frac{0 - \bar{x}_7}{\sqrt{s_7^2}}\right), \tag{2.6}$$

where

Φ is the cumulative density function of the standard Normal distribution,

$\bar{x}_7$ is the expected value of the sample mean INMB from the pilot trial,

s_7^2 is the sample variance of the INMB from the pilot trial.

If the probability of cost-effectiveness and the sample mean INMB are available then it is possible to re-arrange Equation 2.6 to estimate the sample variance of the INMB. This approach assumes that the sample mean INMB is Normally distributed, a reasonable assumption for large samples by central limit theorem also used by authors such as Willan and Pinto (2005). The probability of cost-effectiveness is estimated to be 0.756 using the R model-based health economic analysis of the CACTUS pilot data, giving a sample variance of

$$s_7^2 = 4600.51^2, \tag{2.7}$$

an unbiased estimator of σ_7^2.

Chapter 3

The HERO case study

Charlie Welch, Martin Forster, Belen Corbacho, Ada Keding, Sarah Ronaldson and Puvan Tharmanathan

3.1 Introduction

The HERO trial was a double-blind, randomised, clinical trial carried out in 13 primary and secondary care centres across England. It evaluated whether hydroxychloroquine is superior to placebo for the treatment of hand osteoarthritis (OA). Follow-up took place at six months post-randomisation for the clinical evaluation and at 12 months post-randomisation for the economic evaluation. The study was funded by Arthritis Research UK (now Versus UK) and had a budget of £900,000. The trial protocol is published in Kingsbury et al. (2013) and results of the clinical evaluation are published in Kingsbury et al. (2018). Results of the within-trial economic evaluation are reported in Ronaldson et al. (2021). The trial's budget was held by the University of Leeds and the University of York.[1]

Recruitment to the trial took place between 24 September 2012 and 27 May 2014, with follow-up completed on 25 April 2015. The trial recruited 248 patients, with an average age of 63 years, presenting with symptomatic pain and radiographic hand OA. The recruitment profile is shown in Figure 3.A.1 of the appendix to this chapter. Patients were randomised to receive either: 1. hydroxychloroquine in 200mg, 300mg or 400mg doses or 2. placebo. The base case analysis used the intention to treat principle. The primary endpoint was average hand pain severity during the previous two weeks, measured on an eleven-point (0 to 10) numerical

[1]The trial had research ethics committee approval from the Leeds East Research Ethics Committee and the UK Medicine and Healthcare Products Regulatory Agency. The trial was registered as ISRCTN91859104. All participants gave written informed consent before screening. Costs in the study were measured in UK £ sterling, at 2015 prices. Analysis of the original trial data used Stata 13 (StataCorp, 2013) and the multiple imputation used the `ice` command of Royston (2004, 2005). Analysis in this chapter uses Matlab R2019a (The Math Works, Inc., 2019).

rating scale (NRS), at six months, where a higher score represents a worse level of pain. Secondary endpoints, including quality of life, were also recorded. The trial used the EQ-5D-5L instrument to measure quality of life at baseline, 6 months and one year.

The economic evaluation consisted of a cost-utility analysis (estimating the cost per Quality Adjusted Life Year (QALY) at one year follow-up) and a cost-effectiveness analysis (estimating the cost per unit reduction in pain score). The trial was characterised by a considerable amount of missing data. This is a frequent problem in RCT data, irrespective of how good the data collection is (Carpenter and Kenward, 2012; Faria et al., 2014).[2] The base case economic analysis reported in Ronaldson et al. (2021) uses multiple imputation by chained equations to fill in missing values (Royston, 2004, 2005; StataCorp, 2013) and takes the perspective of the UK National Health Service and Personal Social Services. A secondary analysis takes the societal perspective and includes resource costs such as the child care and travel costs that were incurred by participants for health care attendances (Ronaldson et al., 2021). Sensitivity analysis uses complete cases only, that is, those cases with no missing data on outcome and cost variables over the twelve months of follow-up (approximately 31% of the total number enrolled in the trial).

The trial found that hydroxychloroquine was not superior to placebo in terms of its effect on expected severity of pain at six months (Kingsbury et al., 2018) and expected QALYs at one year (Ronaldson et al., 2021). The expected levels of hand pain were estimated to be 5.49 (placebo) and 5.66 (hydroxychloroquine), a difference of -0.16 (95% confidence interval of (-0.73, 0.40), $p = 0.57$). This result was robust to various adjustments for adherence, missing data and rescue medication. The base case economic analysis found that hydroxychloroquine was not superior to placebo on cost-effectiveness grounds: using a maximum willingness to pay of £30,000 per QALY, the estimate of expected incremental net monetary benefit (INMB) of hydroxychloroquine compared to placebo was –£144.34 (95% confidence interval of (–£158.67, –£130.02)) and the probability that hydroxychloroquine is cost-effective was estimated to be 0.39 (Ronaldson et al., 2021).[3]

The incremental analysis that is presented in this chapter subtracts the estimate for placebo from the estimate for hydroxychloroquine and assumes a maximum willingness to pay of £30,000 per QALY. A basic descriptive analysis of the sequence of outcome and cost data is presented in section 3.2. Section 3.3 describes our methodology. Section 3.4 assesses the value

[2]The problem is amplified when there are frequent measurements and a high number of variables, such as in the HERO trial. Since treatment costs and QALYs are cumulative measures, missing data for a patient at one follow-up point (e.g. a patient's QALYs at 6 months) implies measures at subsequent points (the patient's total QALYs at 12 months) are also missing.

[3]Results for the complete case analysis change little: when the willingness to pay threshold is £30,000 per QALY, the estimate of expected INMB is -£144.30 (95% confidence interval of (-£169.12,£119.45)). When the willingness to pay threshold is £20,000 per QALY, the base case results are -£92.30 (-£102.11,-£82.49) and the complete case analysis results are -£81.67 (-£98.21,-£65.14).

delivered by the various trial designs that we consider, together with a range of other operating characteristics. Section 3.5 concludes. Additional information is presented in Appendices 3.A to 3.D.

3.2 Descriptive analysis of the sequence of outcome and cost data for the HERO trial

In this section we present a sequential descriptive analysis of how some of the key data from the HERO trial evolved as the trial progressed. We present an overview of how the budget was spent and compare it with a description of how the estimate of expected INMB at one year and its constituent parts (incremental QALYs and incremental costs at one year) evolved. Analysis is based on imputed data, using a model that does not condition on any explanatory variables other than randomised group. To complement this material, we also show how the primary outcome measure – hand pain severity in the previous two weeks, measured at 6 months – evolved, using the complete cases. Analysis of the hand pain data is accompanied by a description of how incremental costs at six months accumulated, also using the complete cases.

In recognition of the fact that continuous monitoring of data is unlikely to be feasible during a sequential trial, the analysis assumes that outcomes were observed in blocks of ten pairwise allocations at a time, that is, ten patients allocated to hydroxychloroquineand ten patients allocated to placebo. Although this is not what happened in practice (randomisation did not balance the numbers in each arm of the trial as it progressed), we used our best judgement to approximate the process. We assume a constant rate of recruitment into the trial so that, once follow-up of outcomes has commenced, information on successive blocks is available approximately every 49 days. In light of the fact that the cost-effectiveness analysis takes place at one year post-randomisation, this implies that approximately 74 pairwise allocations (that is, 60% of the total sample size planned for the trial) had been made by the time the first data for cost-effectiveness were observed.

The overall picture to emerge from this analysis is that, had health outcome and treatment cost data been monitored sequentially over the course of the HERO trial, there would have been little evidence that hydroxychloroquine was preferred to placebo on either effectiveness grounds (as measured by incremental hand pain severity) or cost-effectiveness grounds (by calculating the estimate of expected INMB).

3.2.1 Profile of research budget spend and evidence of cost-effectiveness

Figure 3.1, which replicates Figure 1.1c of Chapter 1, plots the cumulative spend of the HERO research budget over the course of the project, using data from the financial accounts (black

continuous line, scale on the left axis).[4] Also plotted is the estimate of the cost-effectiveness of hydroxychloroquine compared with placebo as evidence accumulated, as measured by the average value of incremental net monetary benefit at one year (blue dashed line, scale on the right axis; the series is based on the imputed data as described in the introduction to this section). Positive values of the estimate suggest that hydroxychloroquine is cost-effective; negative values suggest that it is not. Key milestones in the project are denoted by the letters 'A' (recruitment starts), 'B' (recruitment finishes), 'C' (one year follow-up finishes) and 'D' (publication of Kingsbury et al. (2018), presenting the results of the clinical evaluation).

Figure 3.1 shows that, during follow-up, the estimate of expected INMB was never greater than zero, meaning that there was never evidence that hydroxychloroquine was cost-effective. The first estimate, based on the first twenty pairs of patients allocated to the trial, is equal to –£2172.[5] By the end of follow-up, the estimate had risen to –£45. Comparison of these two profiles is important because it illustrates how much (or how little) of the research budget might have been saved if early stopping of the trial had occured. The figure shows that just under one third of the budget had been spent when one year follow-up commenced and just over half had been spent by the time it had finished.

3.2.2 Sequential descriptive analysis of QALYs and treatment costs

Figure 3.2 breaks down the estimate of expected INMB at one year that is plotted in Figure 3.1 into a plot of the estimate of expected incremental QALYs (Figure 3.2a) and incremental treatment costs (Figure 3.2b) at one year. Limits showing plus and minus two standard errors are also indicated. Values above zero show hydroxychloroquine to be more effective (Figure 3.2a)/more costly (Figure 3.2b). The plots show that hydroxychloroquine was estimated to be less effective than placebo throughout the follow-up period. Figure 3.2b shows that treatment with hydroxychloroquine was estimated to be more expensive than placebo throughout the follow-up period, except at the very end, when it was estimated to be £39 cheaper.

3.2.3 Sequential descriptive analysis of the clinical outcomes and treatment costs

Figure 3.3 plots sequential estimates of incremental effectiveness and cost at 6 months, together with limits at plus and minus two standard errors, using the complete case data. Figure 3.3a

[4] Spend includes all costs recorded in the financial accounts, for whatever reason. These include the costs of hydroxychloroquine and placebo, together with the costs of encapsulating the drugs prior to supply to the recruiting sites, so that the drugs would appear identical.

[5] It was not possible to obtain an estimate for the first ten patient pairs owing to data sparsity, which caused difficulties for the chained equations algorithm used for the multiple imputation.

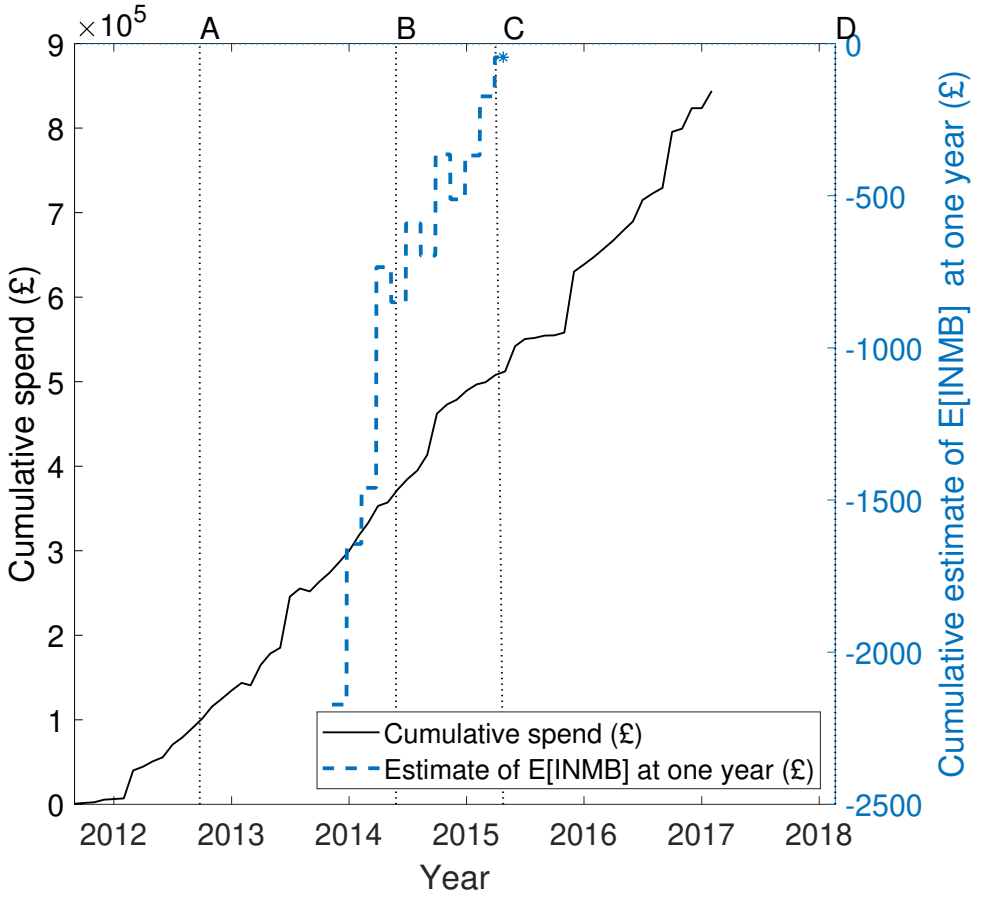


Figure 3.1: Estimate of cumulative budget spend and point estimate of expected incremental net monetary benefit at one year for the HERO trial.

Notes:
The estimate of cumulative budget spend is measured on the left axis and is plotted as a black continuous line. The estimate of the expected value of incremental net monetary benefit at one year is measured on the right axis and is plotted as a blue dashed line. The point estimate of cost-effectiveness at the end of follow-up is marked with a *. Key milestones are: 'A' – recruitment starts; 'B' – recruitment finishes; 'C' – one year follow-up finishes; 'D' – principal publication of the clinical analysis (Kingsbury et al., 2018). The cost-effectiveness data are presented in column 4 of Table 3.B.1. The calculation of the estimate of the expected value of incremental net monetary benefit assumes a maximum willingness to pay for one QALY that is equal to £30,000.

plots the estimate of expected incremental hand pain severity, as measured on the numerical rating scale. Values above zero show hydroxychloroquine to be inferior to placebo. Figure 3.3b plots the estimate of expected incremental treatment costs. The two plots tell a similar story to the plots in Figure 3.2, albeit using a different health outcome measure. Evidence suggested that hand pain severity for patients treated with hydroxychloroquine was no better than, or very similar to, hand pain severity for patients treated with placebo throughout follow-up. Hydroxychloroquine was estimated to be more expensive than placebo throughout follow-up, with the final point estimate suggesting that there was little difference in expected cost between the two treatments.

3.3 Methods

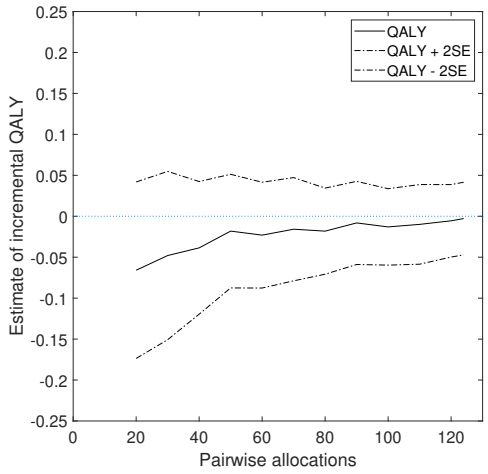
To apply the value-based sequential design to the HERO case study, we used the model and methods outlined in section 1.4. We estimated the parameter values required for the HERO case study to obtain the value-based sequential design's stopping boundary and, in turn, we used this to assess the performance of the value-based sequential design.

The parameter values are presented in Table 3.D.1 and details about how we chose them are presented in Appendix 3.D. We chose the maximum willingness to pay for one Quality Adjusted Life Year to be £30,000, this being one of the two values used in Ronaldson et al. (2021) (the other being £20,000 per QALY). After reviewing literature on the prevalence and incidence of hand OA within the United Kingdom, we set the size of the population to benefit from the adoption decision to 24,500. We set the cost of randomising a pair of patients to the two arms of the trial to £1,650, under the assumption that the approximate split between fixed and variable costs in the trial was 50:50. We set the prior mean for the expected value of incremental net monetary benefit to be zero and the prior variance to give a low weight on prior information relative to the data from the trial (equivalent to an effective sample size of 2 pairwise allocations). This is intended to reflect the lack of information available to investigators prior to the trial taking place. Finally, we set the follow-up period for the cost-effectiveness data to be one year, the same as the follow-up period used in the trial. As already noted, assuming a constant rate of recruitment to the trial, this means that 74 pairwise allocations have been made by the time Stage II of the value-based sequential design starts.

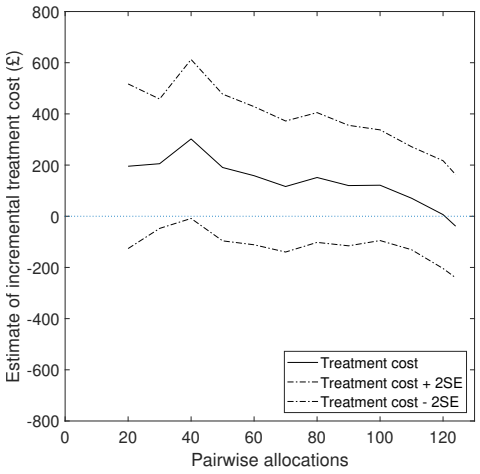
We compared the performance of the original fixed sample size design of the HERO trial with the following value-based models:

1. the value-based one stage design for a normal prior–normal likelihood (Claxton and Posnett (1996) and Claxton (1999));

2. three versions of the value-based sequential design of Chick et al. (2017a) and Alban



(a) Estimate of incremental QALYs.

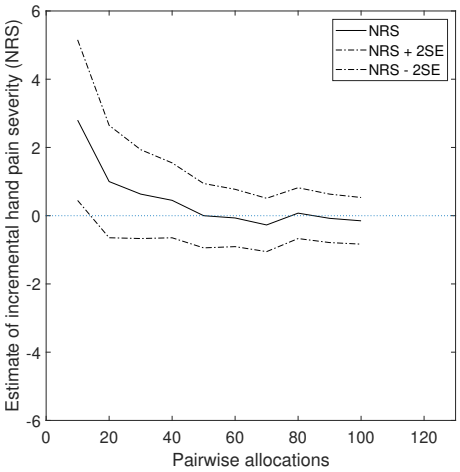


(b) Estimate of expected incremental treatment cost.

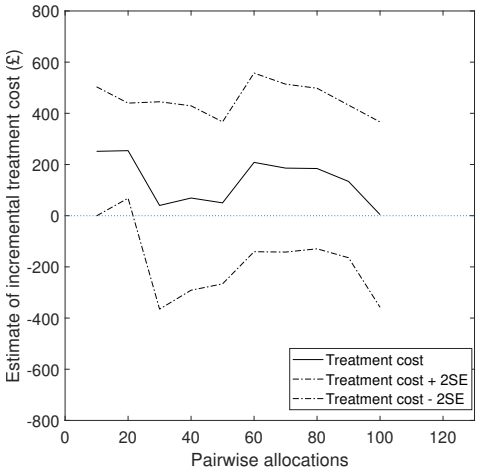
Figure 3.2: Estimate of expected incremental QALYs and treatment costs at one year as evidence accumulated, together with limits at ± two standard errors, using imputed data.

Notes:

Expected incremental QALYs: values above zero suggest hydroxychloroquine to be superior to placebo. Expected incremental treatment cost: values above zero suggest hydroxychloroquine to be more expensive than placebo. The plots show that hydroxychloroquine was estimated to be less effective and more expensive than placebo for the majority of the follow-up period. The data series for Figure 3.2a is presented in column 2 of Table 3.B.1. The data series for Figure 3.2b is presented in column 3.



(a) Estimate of incremental hand pain severity.



(b) Estimate of incremental treatment cost.

Figure 3.3: Estimate of expected incremental hand pain severity and incremental treatment costs at six months as evidence accumulated, together with limits at ± two standard errors (SE), using the complete cases.

Notes:

Expected incremental hand pain severity: values above zero suggest hydroxychloroquine to be inferior to placebo. Expected incremental treatment cost: values above zero suggest hydroxychloroquine to be more expensive than placebo. The plots are based on the complete case data, rather than the imputed data that is used in Figure 3.2. They show that hydroxychloroquine was estimated to be less effective, and more costly, than placebo for the majority of the follow-up period.

et al. (2018): (a) the first with a maximum sample size, $Q_{\max}$, equal to the sample size of the HERO trial (248 patients or 124 pairwise allocations); (b) the second with $Q_{\max}$ set to the optimal sample size of the value-based one stage design described in point 1 above; (c) the third with $Q_{\max}$ equal to double the optimal sample size from (b). We used the third version to assess how much additional value is generated when the maximum sample size of the trial is increased.

Some operating characteristics, such as the value-based one stage design's optimal sample size and the value delivered by the one-stage and sequential designs, are obtained from the numerical solutions to the models. We obtained other operating characteristics using a bootstrap analysis. Each iteration of the bootstrap resampled from the un-imputed quality of life and cost data from the trial, arranged in blocks of ten pairwise allocations. Missing data were cumulatively imputed by block and used to calculate blockwise point estimates of incremental net monetary benefit as the data accumulated. For example, the estimate for the 6th block was based on the point estimate of incremental net monetary benefit calculated using the 60 pairwise allocations in blocks 1 to 6. These point estimates were used to derive a path for the posterior mean of expected value of incremental net monetary benefit, assuming a normal prior distribution for expected value of incremental net monetary benefit and a normal likelihood. Each bootstrapped path was compared with the stopping boundary of the model, in the manner described in section 1.4, and a stopping time calculated. Operating characteristics were then calculated. Further details of the bootstrap analysis are presented in Appendix 3.C.

3.4 Results

3.4.1 A comparison of expected net benefits

Table 3.1 compares the expected net benefits of the designs that we consider, assuming a maximum willingness to pay for one Quality Adjusted Life Year of £30,000 (results for the one stage designs are shown in rows 1 and 2 and results for the sequential designs in rows 3 to 5). The table also lists the total variable costs of the two one stage designs (column 5) and the variable cost of reaching Stage II for the sequential designs (column 6). We note that, regardless of the design under consideration, the fixed costs are assumed to be equal to zero from the perspective of the trials unit team running the trial (see section 1.4.1). Also shown (row 6) is the individual-level expected value of perfect information (EVPI), which is equal to £2148. For a population of 24,500 patients benefitting from the adoption decision over a ten year horizon, the population EVPI is £52.6m. This value may be interpreted as the value to the health care system of eliminating all of the uncertainty that is captured by the prior distribution for the expected value of incremental net monetary benefit of hydroxychloroquine compared with

(1) Trial design	(2) (Maximum) number of pairwise allocations	(3) EVSI/EVVBS (£) Individual	(4) EVSI/EVVBS (£) Population	(5) Total variable cost (£)	(6) Variable cost to Stage II (£)	(7) Expected net benefit (£)	(8) Percentage increase in expected net benefit
1. HERO trial	124	2131	5.22×10^7	2.06×10^5	-	5.20×10^7	-
2. Value-based one stage design	177	2136	5.23×10^7	2.92×10^5	-	5.20×10^7	0.067
3. Value-based sequential design	124	2129	5.22×10^7	-	1.22×10^5	5.20×10^7	0.066
4. Value-based sequential design	177	2133	5.23×10^7	-	1.22×10^5	5.21×10^7	0.252
5. Value-based sequential design	354	2137	5.24×10^7	-	1.22×10^5	5.22×10^7	0.451
6. EVPI	-	2148	5.26×10^7	-		-	-

Table 3.1: Performance of different trial designs from a value-based perspective.

Notes: Column 2 records the number of pairwise allocations made in the HERO trial (row 1), the optimal sample size for the value-based one stage design (row 2) and the maximum number of pairwise allocations that could be made in the value-based sequential designs (rows 3 to 5). Columns 3 and 4 record the expected value of sample information (EVSI) for the trials listed in rows 1 and 2 and the expected value of running Stage II of the value-based sequential designs (EVVBS) for the trials listed in rows 3 to 5. Column 5 records the total variable cost of sampling for the two fixed sample size trials in rows 1 and 2. Column 6 records the unavoidable total variable costs of reaching Stage II for the value-based sequential designs in rows 3 to 5. Column 7 records the expected net benefit of the design, equal to the design's EVSI minus total variable costs (for the fixed sample size trials listed in rows 1 and 2) and the EVVBS minus the cost of reaching Stage II (for the value-based sequential designs listed in rows 3 to 5). Column 8 records the percentage increase in expected net benefit of the relevant value-based model over the expected net benefit delivered by the HERO trial (row 1), where the expected net benefits are those reported in column 7.

placebo.

The fixed sample size designs

As noted previously, the HERO trial recruited 248 participants, which translates to 124 pairwise allocations. Figure 3.4 plots the expected value of perfect information, the expected value of sample information, the variable research cost and the expected net benefit of sampling as a function of the number of pairwise allocations made in a fixed sample size (one stage) trial. Also marked are the sample size of the HERO trial ($p_{\text{trial}} = 124$ pairwise allocations) and the number of pairwise allocations which maximise the expected net benefit of sampling for the value-based one stage design, $p^* = 177$. p^* is approximately 43% larger than p_{trial}. However, as Figure 3.4 and Table 3.1 show, although there is a large difference between p^* and p_{trial}, in both absolute and percentage terms, there is a relatively small difference between the expected net benefit of sampling. That is, the maximised expected net benefit of sampling delivered by the value-based one stage design – this is the largest achievable gain for the health care system as a result of reducing uncertainty about the expected value of incremental net monetary benefit, as captured by the prior distribution, and accounting for the variable costs of running the trial – is approximately 0.067% higher (+£35,000) than the expected net benefit of sampling at p_{trial}.

These results are due to the relatively flat expected net benefit of sampling function between the points p_{trial} and p^*. The shape of this function is determined by the nature of the expected value of sample information and variable cost functions (see Figure 3.4):

- Figure 3.4 shows that the majority of the information gain that results from running a fixed sample size trial to reduce uncertainty in the prior distribution for the expected value of incremental net monetary benefit takes place before the sample size of the HERO trial, p_{trial}, is reached. This results in a relatively flat EVSI function for all sample sizes greater than p_{trial} and a range of sample sizes immediately below it. Table 3.1 shows that the expected value of sample information provided by the HERO trial is £52.2m, which is approximately 99% of the expected value of perfect information.

- The cost of randomising an additional pair of patients into the trial is relatively low, so that the slope of the variable cost function is close to the slope of the expected value of sample information function for a large range of values of the sample size of the trial, including those in the interval $[p_{\text{trial}}, p^*]$.

The optimal sample size for the value-based one stage design is defined by the sample size at which the marginal benefit of increasing the sample size (the slope of the EVSI function that is plotted in Figure 3.4) is equal to the marginal cost (the slope of the variable cost function that is plotted in Figure 3.4). Given the features of these two functions as just described, this results

in two sample sizes – p_{trial} and p^* – which are reasonably far from each other in both absolute and relative terms, but two expected net benefits which are not.

The value-based sequential design

Figure 3.5 presents the stopping boundaries for the value-based sequential design for two of the three designs that we considered. Figure 3.5a sets the maximum sample size of the sequential trial equal to the sample size of the HERO trial (124 pairwise allocations). Figure 3.5b sets the maximum sample size equal to the optimal sample size of the value-based one stage design (177 pairwise allocations). The Stage II stopping boundaries are marked (circled points linked by a continuous line), together with the ranges of the prior mean over which no trial, a value-based one stage design and the value-based sequential design are optimal. Where the value-based one-stage design is optimal, the optimal sample sizes are marked by circles.[6]

Rows 3 to 5 of Table 3.1 report the expected values delivered by the value-based sequential designs. The results show that these provide additional net benefit over that which is delivered by the HERO trial and the value-based one stage design, but it is small. For example, when the maximum sample size is 177 pairwise allocations, which is equal to the number of pairwise allocations which maximise the expected net benefit of sampling for the value-based one stage design, the gain of the sequential design over the one stage design is less than 1%. When it is doubled to 354 pairwise allocations, the gain remains less than 1%. Figure 3.6 shows the absolute value of the expected gain over the design of the HERO trial as a function of the maximum sample size chosen for the value-based sequential design. As the maximum sample size increases, the gain increases, but at a decreasing rate. For example, when the maximum sample size is set to 1000 pairwise allocations, the gain is approximately £266,000 (0.51%).

A preliminary assessment of the reasons behind this result suggest it is related to the number of patients that are recruited in the value-based sequential design before Stage II can commence (Stage II being the period of the trial during which the posterior mean is being updated and there is the option to continue to randomise or stop the trial). Firstly, given the assumption about the rate of recruitment, 74 pairwise allocations have already been randomised to the two arms of the trial by the time Stage II commences. This 'pre-commits' a large fraction of the trial's total variable research costs before there is any chance to halt recruitment to the trial. Secondly, these 74 pairwise allocations deliver a large proportion of the expected value of perfect information available in the trial: the EVPI at the start of the trial is £52.6m (recall row 6 of Table 3.1) and

[6] It should be noted that the value-based one-stage designs referred to here are not the same as the value-based one stage design whose results are reported in row 2 of Table 3.1. The value-based one stage design reported in row 2 of Table 3.1 solves for an optimal number of pairwise allocations, assuming that the prior mean for the expected value of incremental net monetary benefit is equal to zero. The one-stage designs in Figure 3.5 solve for an optimal number of pairwise allocations over a range of non-zero values of the prior mean (those plotted on the vertical axes).

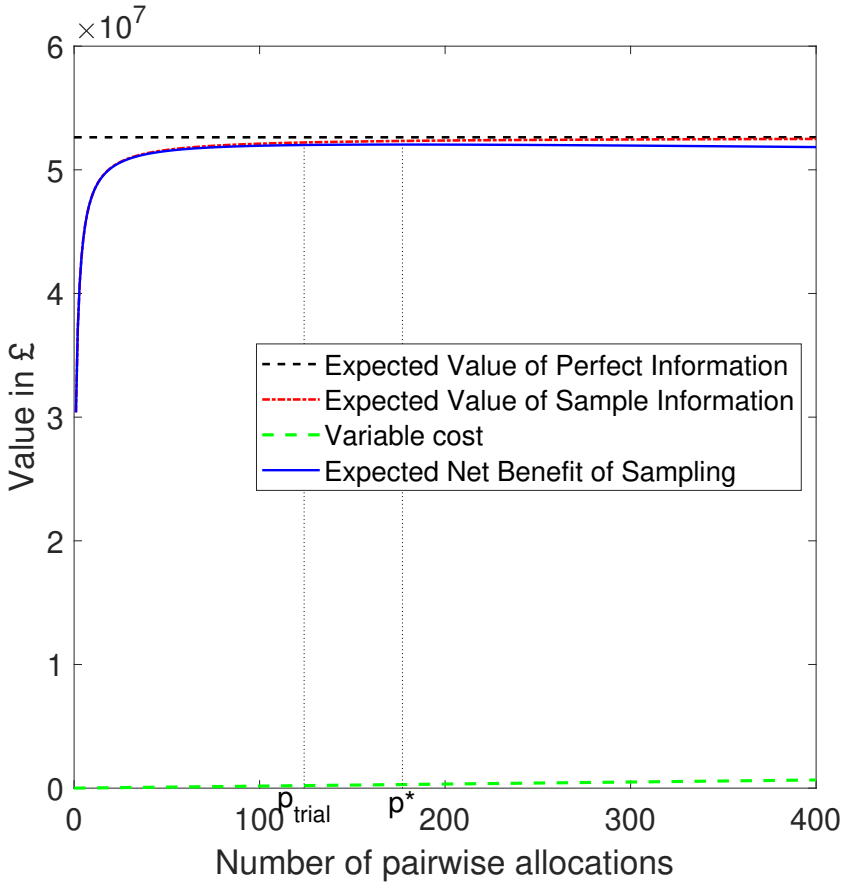
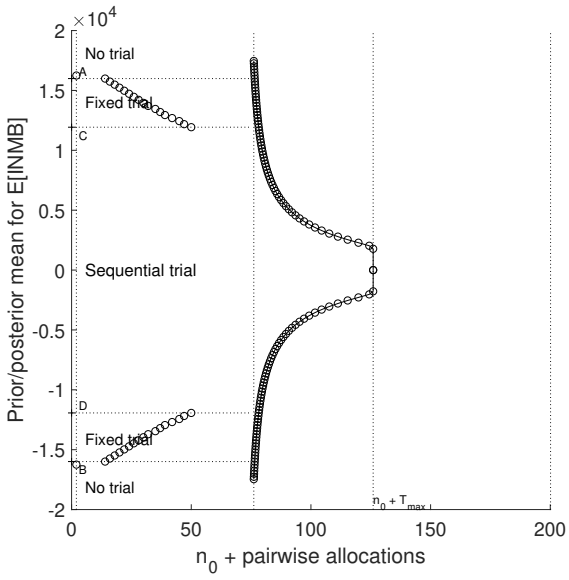
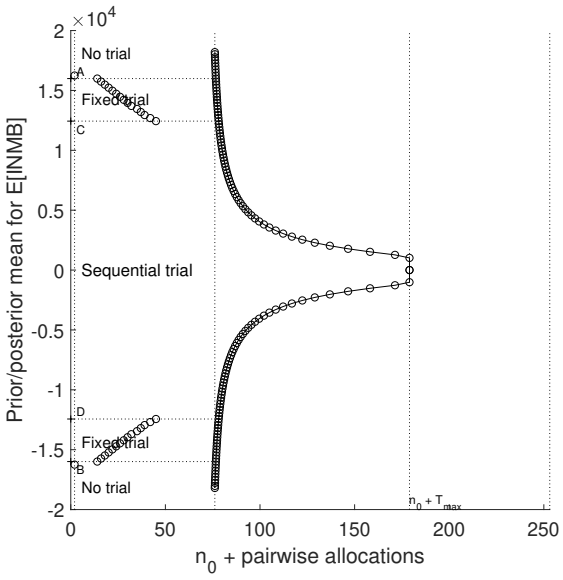


Figure 3.4: Expected value of perfect and sample information, variable cost and expected net benefit of sampling for the HERO trial.

Notes:
The figure shows the expected value of perfect information, the expected value of sample information, the expected net benefit of sampling and the variable cost of the HERO trial and as a function of the number of pairwise allocations made in a fixed sample size (one stage) trial. The expected net benefit of sampling is equal to the expected value of sample information minus the variable cost of sampling. Also marked are the number of pairwise allocations made in the HERO trial ($p_{\text{trial}} = 124$) and the number of pairwise allocations which maximise the expected net benefit of sampling for the value-based one stage design, $p^* = 177$. p^* is approximately 43% larger than p_{trial}. However, the figure shows that the majority of the information gain from sampling takes place over the interval [0,100] and the marginal cost of sampling (the slope of the variable cost function) is relatively low. As a result, the expected net benefit of sampling function is relatively flat in the interval [100, 400] and so the 'added value' of the value-based one stage design over the design of the actual HERO trial, defined as the difference between the expected net benefit of sampling at p^* and at p_{trial}, is low.



(a) Maximum number of pairwise allocations = 124 (sample size of the HERO trial)



(b) Maximum number of pairwise allocations = 177 (optimal choice for the value-based one stage design)

Figure 3.5: Stage II stopping boundaries and policies for the value-based sequential designs with a maximum sample size of 124 and 177 pairwise allocations.

Notes:
The figures show the stopping boundaries for Stage II of two Bayesian value-based trial designs. Figure 3.5a sets the maximum sample size equal to 124 pairwise allocations, which is the maximum sample size of the HERO trial itself. Figure 3.5b sets the maximum sample size equal to 177 pairwise allocations, equal to the optimal sample size of the value-based one stage design. Also shown are regions for the prior mean in which no trial, a fixed sample size trial, and a sequential trial are optimal. In the regions where a fixed sample size trial is optimal, some optimal sample sizes are marked with circles.

the EVPI remaining after the outcomes of 74 pairwise allocations have been made is £8.5m. This reduces the total amount of information which remains to be exploited by the sequential design during Stage II. The combination of these two effects – both of which are related to the number of patients who are recruited to the trial before Stage II starts – reduces the additional value that the value-based sequential designs can generate over the actual trial (and the best value-based one stage design).

3.4.2 Results from the bootstrap analysis

Figure 3.7 plots the two Stage II stopping boundaries that are plotted in Figure 3.5a (as a red, continuous line) and Figure 3.5b (as a blue, dashed line), one superimposed upon the other. Also plotted is the path of the posterior mean for the expected value of incremental net monetary benefit from the trial (solid black line, with interim analyses marked as circles, drawn using the data from Table 3.B.1 and the parameter values defining the prior distribution , see

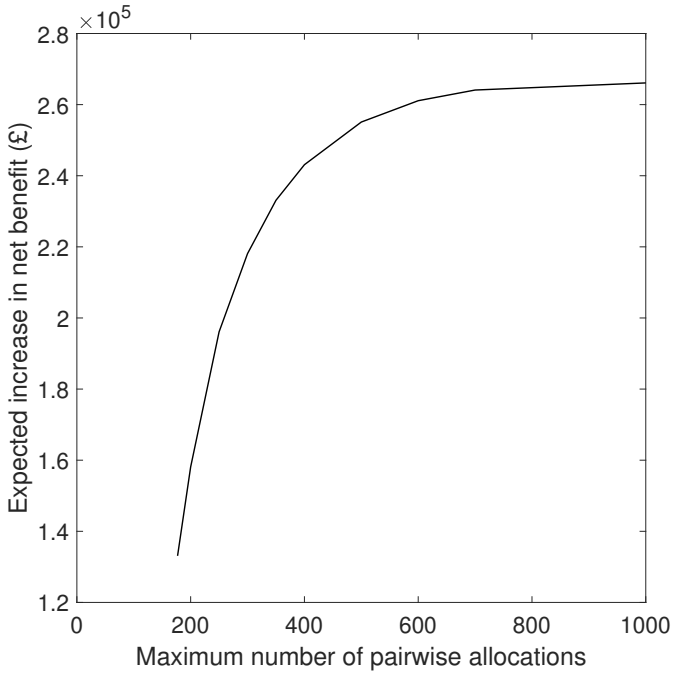


Figure 3.6: Difference in expected net benefit provided by the value-based sequential designs with the maximum sample size indicated and the design of the HERO trial.

Notes:
The figure shows that, for the range of the maximum number of pairwise allocations that is plotted, the additional expected net benefit delivered by the value-based sequential design over the original design of the HERO trial is between approximately £120,000 and £266,000 and is an increasing and strictly concave function of the maximum sample size chosen for the value-based sequential design.

Table 3.D.1). This path starts at 20 pairwise allocations owing to the sparsity of the data for the chained equations approach that was used to impute missing values (see section 3.2). Two bootstrapped paths for the posterior mean are also indicated, using pink and green dashed lines.

Figure 3.7 shows that, had a value-based sequential design with a maximum sample size of 124 pairwise allocations been used to obtain the stopping boundary for the trial, the trial would not have stopped early because the path for the posterior mean does not cross the red stopping boundary before the maximum sample size is reached. The same result holds if the maximum sample size had been set to 177 pairwise allocations. This result is due to the equivocal signal about incremental net monetary benefit which emerged from the trial as data accumulated (refer to Figure 3.2). Although it is not clear from Figure 3.7, the posterior mean was negative once all of the pipeline patients had been followed up, meaning that, upon its conclusion, the posterior mean would have indicated that hydroxychloroquine was not cost-effective.

The two bootstrapped paths show the different results when the maximum sample size is set to 124 pairwise allocations. Resampled path 1 (pink) remains in the continuation region so that, had it been the 'true' path for the posterior mean during the trial, the trial would not have

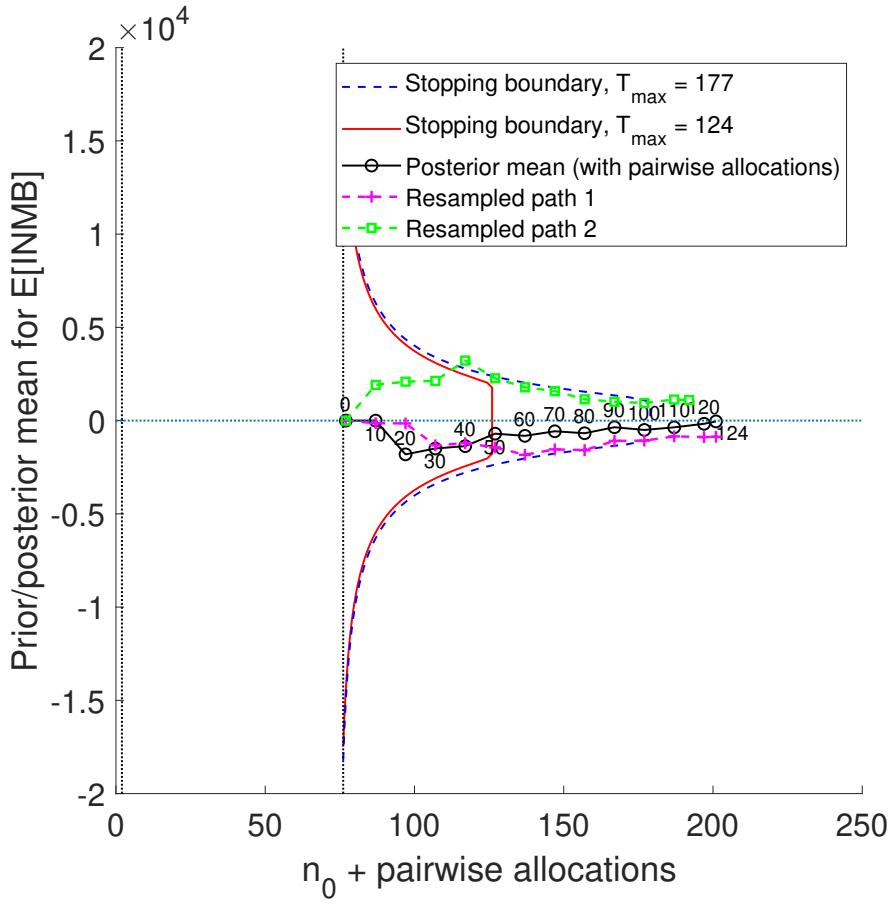


Figure 3.7: Stopping boundaries for the HERO trial, together with the path of the posterior mean for expected incremental net monetary benefit and two bootstrapped paths.

Notes:
The figure shows the stopping boundary for the value-based sequential design when the maximum sample size is equal to 124 (red, continuous line) and 177 (blue, dashed line) pairwise allocations. The path for the posterior mean for the expected value of incremental net monetary benefit of hydroxychloroquine compared with placebo using the trial data is shown in black, with interim analyses marked as '○'. The number of pairwise allocations contributing to each interim analysis is also marked. Two boostrapped paths (Resampled path 1 is pink and Resampled path 2 is green) are shown as dashed lines with markers. The black and pink paths remain in the continuation region for a value-based sequential design whose maximum sample size is 124 pairwise allocations. The green path crosses the upper stopping boundary. 'Resampled path 1' suggests that hydroxychloroquine is not cost-effective once all information on the pipeline patients has arrived (the posterior mean is less than zero at the final observation point), whereas 'Resampled path 2' suggests that it is (the posterior mean is greater than zero at the final observation point). The path of the posterior mean using the trial data is close to zero but negative once the trial concludes, so that it too suggests that hydroxychloroquine is not cost-effective at the end of the trial.

	Average	Standard deviation	Minimum	Maximum
Maximum sample size = 124 pairwise allocations				
HERO trial				
Posterior mean for $\mathbb{E}$[incremental NMB]	−65.03	653.09	−1961.64	1840.14
Value-based sequential design				
Posterior mean for $\mathbb{E}$[incremental NMB]	−65.58	659.15	−2306.28	1851.22
Sample size (pairwise allocations)	123.60	2.46	94	124
Maximum sample size = 177 pairwise allocations				
value-based one stage design				
Posterior mean for $\mathbb{E}$[incremental NMB]	−76.12	556.35	−1790.51	1775.98
Value-based sequential design				
Posterior mean for $\mathbb{E}$[incremental NMB]	−80.25	590.12	−2306.28	1851.22
Sample size (pairwise allocations)	173.05	12.39	94	177
Maximum sample size = 354 pairwise allocations				
Trial with a fixed sample size				
Posterior mean for $\mathbb{E}$[incremental NMB]	−71.09	396.99	−1165.70	1039.62
Value-based sequential design				
Posterior mean for $\mathbb{E}$[incremental NMB]	−85.05	537.64	−2219.82	1797.54
Sample size (pairwise allocations)	311.77	65.05	94	354

Table 3.2: Bootstrap analysis: comparison of the performance of the original fixed sample size design with the value-based sequential versions.

Notes:
The table reports the average value and standard deviation of the posterior mean at stopping, together with the average sample size and standard deviation of fixed designs and value-based sequential designs, assuming a maximum sample size of 124 pairwise allocations (the sample size of the HERO trial), 177 pairwise allocations (the optimal sample size of the value-based one stage design) and 354 pairwise allocations (twice the optimal sample size of the value-based one stage design). The standard deviations refer to the standard deviations of the values reported in the data that are used to calculate the average values, that is, the standard deviations of the data that result from the bootstrap analysis.

	Final decision	
	Hydroxychloroquine not cost-effective	Hydroxychloroquine cost-effective
Maximum sample size = 124 pairwise allocations		
HERO trial	0.543	0.457
Value-based sequential design	0.544	0.456
Maximum sample size = 177 pairwise allocations		
Value-based one stage design	0.562	0.438
Value-based sequential design	0.559	0.441
Maximum sample size = 354 pairwise allocations		
Trial with a fixed sample size	0.584	0.416
Value-based sequential design	0.578	0.422

Table 3.3: The proportion of bootstrap paths which suggest that hydroxychloroquine is cost-effective, for various fixed sample size and value-based sequential designs.

Notes:
The table reports the proportion of paths which show hydroxychloroquine to be cost-effective, assuming a maximum sample size of 124 pairwise allocations (the sample size of the HERO trial), 177 pairwise allocations (the value-maximising sample size for a value-based one stage design) and 354 pairwise allocations (twice the value-maximising sample size for a value-based one stage design).

stopped early. Resampled path 2 (green) crosses the stopping boundary after outcomes for 40 pairwise allocations have been observed so that, had it been the 'true' path, the trial would have stopped early. The two bootstrapped paths provide different cost-effectiveness information once all outcomes have been followed up. The posterior mean at stopping for resampled path 1 is negative, suggesting that hydroxychloroquine is not cost-effective; that for resampled path 2 is positive, suggesting that it is.

Results for the posterior mean of the expected value of incremental net monetary benefit at the time of stopping the trial, together with the expected sample sizes of the various designs and the proportion of bootstrapped paths which conclude that hydroxychloroquine is cost-effective, are presented in Tables 3.2 and 3.3. When the maximum sample size of the value-based sequential design is set to 124 pairwise allocations, very few bootstrapped paths stop before the maximum sample size. This result is due to the small number of interim analysis points that are available during Stage II, combined with the fact that the signal from the trial is equivocal. As a result, the operating characteristics of the value-based sequential design and the fixed sample size models are similar.

When the maximum sample size of the value-based sequential design increases, Table 3.2

shows that the expected sample size increases. This is an unsurprising result. These designs deliver a slightly lower posterior mean for expected incremental net monetary benefit at stopping. However, there is little difference in the proportion of bootstrapped paths which show hydroxychloroquine to be cost-effective across the designs (Table 3.3). For example, when the maximum sample size is set to 354 pairwise allocations, the proportion of paths which show hydroxychloroquine to be cost-effective is 0.416 for the value-based one stage design and 0.422 for the value-based sequential design.

3.5 Discussion and conclusions

Before discussing our results, we note that they are dependent upon the choices that we have made for the parameter values and the particular models that we study. For example, to assess the models using our Bayesian decision-theoretic set-up, we assume a normal prior–normal likelihood and we assume that we know the sampling variance for incremental net monetary benefit. For the bootstrap analysis, we assume that interim analyses take place during Stage II in blocks of ten pairwise allocations at a time, informed using the procedure outlined in section 3.C. We also note that the parameter values that we have chosen are estimates; section 3.D discusses their choice in detail. Directions for future research, discussed below, include investigating the sensitivity of the results to changes in these assumptions.

From a qualitative perspective, the results from the HERO case study are unsurprising, and are consistent with those of the CACTUS case study: the value-based one stage design creates more value – measured in terms of the extra net benefit that the trial can generate for the health care system (the expected net benefit of sampling) – than the HERO trial itself, because the HERO trial did not set its sample size equal to the sample size which maximises the expected net benefit of sampling. Further, a value-based sequential design, which uses as its maximum sample size the optimal sample size from the value-based one stage design, delivers additional value, over and above the added value provided by the value-based one stage design. This, too, is an unsurprising result, since the additional flexibility introduced by 'thinking and acting sequentially' permits the trial to stop in situations where the posterior mean for the expected value of incremental net monetary benefit is extreme enough to make the expected benefit of randomising a further pair of patients to the two arms of the trial not worth the cost. Finally, the added value delivered by the value-based sequential design is shown to be an increasing and strictly concave function of the choice of maximum sample size – putting it another way, giving the trial the opportunity to 'run long' adds value – a result which is consistent with the finding of the CACTUS case study.

However, from the quantitative perspective, the 'value added' offered by both the one-stage and sequential value-based models, when compared with the HERO trial, is estimated to be

small (less than 1%). Preliminary assessment of the reasons for this result suggest that, for the value-based one stage design, it is related to the large amount of information provided by the early participants recruited to the trial, reducing the amount of information which remains to be exploited at higher sample sizes, and the relatively low marginal cost of sampling. These features, when combined, lead to a relatively flat expected net benefit of sampling function in the region of the sample size of the trial itself and the value-based one stage design's optimal sample size (refer to Figure 3.4). In addition, for the value-based sequential design, the large number of patients recruited to the trial and 'in the pipeline' by the time Stage II starts means that a large amount of the trial's total variable cost is already spent before interim analyses can take place. This reduces the savings that are available from stopping the trial early.

The bootstrap analysis shows that no model is able to discriminate clearly between hydroxychloroquine and placebo on cost-effectiveness grounds. This is due to the equivocal 'signal' in the data from the trial, which found no evidence that hydroxychloroquine is more effective than placebo, and little evidence of a large cost difference between the two treatments. This result is different to the results of the application of the value-based sequential design to the ProFHER trial (Forster et al., 2021). The ProFHER trial found that there was little difference between the estimates of effectiveness of the two health technologies under investigation, but a large difference in the estimates of their costs. Consequently, the path of the posterior mean for the expected value of incremental net monetary benefit, plus the majority of bootstrapped paths, indicated that early stopping on cost-effectiveness grounds was the best thing to do. Early stopping reduced both the expected number of patients to be randomised to the two arms of the trial (by about 38%), as well as the trial's expected cost (by about 13%).

Future research using the HERO case study could investigate how sensitive the above results are to changes in both the way the model is set up and solved and the values of the parameters that are chosen. Some of the most promising ideas include:

- the model assumes that the sampling variance for incremental net monetary benefit is known. This could be relaxed using the methods outlined in Chick et al. (2017a);
- the bootstrap analysis used only 1000 replications and a multiple imputation by chained equations approach which did not adjust for baseline covariates (see section 3.B). The existing bootstrap analysis could be repeated with more replications and a new analysis carried out with between group differences for each block estimated using an analysis model that adjusts for informative baseline covariates;
- parameter values could be varied. For example, the HERO case study has treated the fixed and variable research costs differently to the way the CACTUS case study treated them and it has made some strong assumptions about the way the number expected to benefit from the adoption decision has been calculated, in particular that the 'patient horizon' is

fixed and the future stream of patients to benefit is not discounted (see section 3.D). In addition, we have assumed that there would be no fixed cost incurred by the health care system if it adopted hydroxychloroquine for the population of interest;

- the analysis could be extended to incorporate the fixed costs of operating the trial into the decision about which design (no trial, value-based one stage design or value-based sequential design) is preferred. This could investigate the potential for different trial designs having different fixed costs.

Directions for future research for value-based sequential models in general is discussed in chapter 5.

Appendix

3.A Recruitment profile for the HERO trial

The recruitment profile for the HERO trial is shown in Figure 3.A.1. It records the date of randomisation ('Rand'), together with the number of participants providing outcome data at six months ('M6') and twelve months ('M12'). Recruitment took place between 24 September 2012 and 27 May 2014, with follow-up completed on 25 April 2015 (Kingsbury et al., 2018).

3.B Data for the clinical and cost-effectiveness analysis in the HERO trial

3.B.1 Complete case analysis

The complete case cost-utility analysis used only 76 participants (42 from the hydroxychloroquine arm and 34 from the placebo arm) out of the 248 randomised to the two arms of the trial, owing to large amounts of missing data. Missing values were present for variables recording both health care resource use and health outcome (patients not returning their EQ-5D-5L questionnaires and/or not replying to all of the questions). Using data on the complete cases only, the estimate of expected value of incremental net monetary benefit (hydroxychloroquine compared with placebo) at 12 months was £144.30 (95% CI of (–£169.15, £119.45)) (Ronaldson et al., 2021).

3.B.2 Unadjusted imputed data for descriptive analysis

The descriptive analysis presented in this chapter is based on imputed data which does not make adjustments for covariates such as the baseline measure of utility. This is the same as the approach taken in the ProFHER pragmatic trial (Forster et al., 2021) and works under the assumption that the data are missing completely at random and that the randomisation process has successfully balanced the intervention and control groups. These imputed data, arranged in blocks of ten pairwise allocations, are presented in Table 3.B.1

	Sep-12	Oct-12	Nov-12	Dec-12	Jan-13	Feb-13	Mar-13	Apr-13	May-13	Jun-13	Jul-13	Aug-13	Sep-13	Oct-13	Nov-13	Dec-13	Jan-14	Feb-14	Mar-14	Apr-14	May-14	Jun-14	Jul-14	Aug-14	Sep-14	Oct-14	Nov-14	Dec-14	Jan-15	Feb-15	Mar-15	Apr-15	May-15	Jun-15
Rand	1	3	6	13	23	36	55	75	97	109	122	134	149	164	179	192	204	216	232	244	248													
M6							1	1	1	12	20	32	43	56	67	80	92	98	111	118	133	141	170	176	190	204	208	208	210	211				
M12														2	4	9	18	28	39	49	65	74	90	91	103	119	134	144	148	159	173	177	183	190

Figure 3.A.1: Recruitment profile for the HERO trial.

Notes:
'Rand' records the numbers randomised into the trial in the months recorded in the columns. 'M6' denotes the number of participants available for analysis at 6 months' follow-up. 'M12' denotes the number of participants available for analysis at 12 months' follow-up.

(1) Number of pairwise allocations	(2) Estimate of expected value as evidence accumulated Incremental QALY	(3) Incremental cost (£)	(4) INMB (£)
10	-	-	-
20	–0.0659	195.448	–2172.148
30	–0.0480	205.335	–1644.585
40	–0.0386	302.034	–1459.959
50	–0.0181	190.567	–734.587
60	–0.0230	158.461	–849.986
70	–0.0158	116.218	–590.775
80	–0.0182	151.452	–696.664
90	–0.0081	119.886	–364.102
100	–0.0130	121.334	–511.874
110	–0.0100	70.439	–367.916
120	–0.0056	6.455	–173.505
124	–0.0028	–38.749	–45.058

Table 3.B.1: Unadjusted point estimates of incremental QALYs, incremental costs and incremental net monetary benefit, together with the point estimates of expected value of incremental net monetary benefit, at one year as evidence accumulated.

Notes: The data come from multiple imputation data sets and are ordered assuming that an interim analysis takes place every ten pairwise allocations. The row corresponding to 124 pairwise allocations refers to point estimates upon conclusion of the trial. The data in column 2 are plotted in Figure 3.2a; the data in column 3 are plotted in Figure 3.2b and the data in column 4 are plotted in Figure 3.1. The calculations assume a maximum willingness to pay for one QALY of £30,000.

3.B.3 Adjusted imputed data for bootstrap analysis

The main cost-effectiveness analysis in the HERO trial used imputed data, and adjusted for covariates and baseline cost (Ronaldson et al., 2021). The resulting ICER is £2,267 per QALY (cost per QALY lost), the incremental net monetary benefit is –£144.34 (95% CI equal to (–£158.67, –£130.02)), using a maximum willingness to pay per QALY of £30,000. The probability that hydroxychloroquine is cost-effective is equal to 0.40, assuming a maximum willingness to pay of £20,000 per QALY and 0.39 assuming a maximum willingness to pay of £30,000 per QALY.

3.C Procedure used for the bootstrap

500 bootstrap paths of the posterior mean for expected incremental net monetary benefit were used to assess the operating characteristics of the models that are reported in Tables 3.2 and 3.3 (for the value-based sequential designs we set the maximum sample size to 124, 177 and

354 pairwise allocations). We used the 'Common Random Numbers' approach in an attempt to reduce the variance involved in comparing results. This was carried out by running the bootstrap for the value-based sequential design with a maximum sample size of 354 pairwise allocations and then using these results to obtain the bootstrap trials for the models with 124 and 177 pairwise allocations. The following steps describe the procedure used to generate one bootstrap path for the trial with 354 pairwise allocations.

1. Draw (with replacement) 354 observations from each allocation stratum (354 from hydroxychloroquine, 354 from placebo) from the original HERO dataset.

2. Randomly sort this resample and place into 36 blocks of 10 pairwise allocations based on this random order (35 blocks of 10 pairwise allocations and 1 block of 4 pairwise allocations).

3. For blocks $i = 2, \ldots, 36$, missing values were imputed as follows:

   (a) use all patients in blocks 1, ..., i (e.g. for block 8, use the 80 hydroxychloroquine patients in blocks 1–8 and the 80 placebo patients in blocks 1–8);

   (b) if all patients have a value of either 0 or missing for a given variable, then replace the missing values with 0 (this step is required primarily for some of the cost variables in the first few blocks);

   (c) identify variables with missing values that will lead to severe collinearity in the univariate models fitted as part of the multiple imputation by chained equations (MICE) procedure and exclude them from that procedure (this generally applied to cost variables in early blocks, where the majority of values are either 0 or missing, with only a small proportion of non-zero, non-missing values);

   (d) impute incomplete variables using MICE, generating five datasets where these variables are complete;

   (e) impute missing values in variables excluded from the MICE procedure (if any) using hotdeck multiple imputation (i.e. missing values are imputed with observed values in proportions that match the observed data, and imputed values can vary within individual between imputed datasets).

4. For blocks 2,...,36, calculate a point estimate of expected value of incremental net monetary benefit for each of the five imputed datasets and combine these point estimates by taking their mean (as per Rubin's rules), resulting in point estimates of expected value of incremental net monetary benefit for blocks 2, ...,36.

5. Use the point estimates of expected value of incremental net monetary benefit for blocks 2,...,36 (using the accumulated data) to calculate point estimates of expected value of incremental net monetary benefit for each block.

6. Use the point estimates of expected value of incremental net monetary benefit for each block and block sizes, together with Bayesian updating (normal prior and likelihood and known variance) to calculate a path of the prior/posterior mean of expected value of incremental net monetary benefit.

3.D Parameter values used for the HERO case study

The parameter values used for the analysis are sourced and calculated as follows:

1. Estimate of fixed costs of adopting hydroxychloroquine: these were estimated to be zero.

2. Estimate of the sampling standard deviation, σ_X. We considered two possible values. The first is based on the complete cases and is equal to £7632. The second obtains an estimate from the multiple imputation analysis by using Rubin's rules and is equal to £7615. In view of the facts that: 1. the two values are so similar and 2. there is a large amount of missing data in the study, we used the latter value.

3. Estimate of P, the number of patients affected by the adoption decision. This requires an estimate of the incidence rate of the condition and the time horizon over which the decision will apply. P may then be estimated according to:

$$P = \sum_{i=0}^{T} \frac{I_t}{(1+\delta)^t}, \tag{3.1}$$

where T is the time horizon and δ is the discount rate (Claxton and Posnett, 1996; Philips et al., 2008; Rothery et al., 2020).

We consulated a range of publications to estimate I_t (Yu et al., 2015, 2017; Morgan et al., 2019; Swain et al., 2020). To meet the eligibility criteria for the HERO trial patients needed to be aged 18 or over and have OA of the first carpometacarpal (CMC) joint and symptomatic OA affecting other hand joints (Kingsbury et al., 2013). Neither Kingsbury et al. (2013, 2018) nor Ronaldson et al. (2021) provide an estimate of the rate at which patients in the United Kingdom report with such symptoms. Using the NHS Hospital Episode Statistics database (Department of Health and Social Care, 2020), Morgan et al. (2019) found 88,178 diagnoses of OA of the first CMC joint between 2000/2001 and 2017/2018 in England, which equates to an incidence rate of approximately 4,900 per

year. Absent information about what proportion of these individuals would have met the study's inclusion criteria, we assumed that half of them would have done. This gives an estimate of $I_t = 2450$. For the time horizon and discount rate, we followed the approach taken by Forster et al. (2021) and set $T = 10$ and $\delta = 0$, so that $P = 24500$.

4. Estimate of c, the marginal cost per pairwise allocation, is calculated using the financial records from the trial (those used to produce Figure 3.1). Approximately £90,216 was spent prior to recruiting the first patients. This is classified as the fixed set-up cost of the trial. An estimated 50% of the £409,161 of costs incurred between the start of patient recruitment and the finish of follow-up is taken to be the variable cost of the trial, giving an estimate of the marginal cost of adding one pairwise allocation to be £204,580/124 = £1,650. The remaining 50% is taken to be a cost (such as overheads) which would have been incurred during the recruitment phase even if no patients were being recruited. Finally, costs of £336,042 were incurred post follow-up. This gives a total spend of £835,419.

5. Estimate of τ, the delay (measured by the number of pairwise allocations) in observing cost-effectiveness at one year. The trial recruited 248 patients (124 patient pairs) over a period of 611 days (between 24 September 2012 and 27 May 2014). So one pair of patients was randomised approximately every 5 days. This equates to a value of $\tau \approx 74$ pairwise allocations in one year. Comparison of this estimate with the recruitment profile in Figure 3.A.1 shows reasonable agreement: Figure 3.A.1 shows that the actual number recruited in the first year was approximately 149.

Parameter	Definition	Value	Source
	Estimated number of patients per annum	2450	Morgan et al. (2019)
	Time horizon for post-decision population	10 years	
P	Number expected to benefit from adoption decision	24500	Defined from above parameters
σ_X	Standard deviation of incremental net monetary benefit in population	£7,615	Multiple imputation data sets
n_0	Effective sample size of the prior distribution	2 pairwise allocations	Assumption
μ_0	Prior mean for the expected value of incremental net monetary benefit	0	Assumption
Δ	Delay for observing EQ-5D-5L endpoint (in years)	1	Kingsbury et al. (2018)
	Estimated annual rate of recruitment to trial	74 pairwise allocations	Kingsbury et al. (2018)
τ	Delay for observing EQ-5D-5L endpoint (in pairwise allocations)	74 pairwise allocations	Annual rate of recruitment
	Time horizon of trial	611 days	Kingsbury et al. (2018)
I_{N}	Fixed cost of adopting hydroxychloroquine	£0	HERO team advice
	Estimated spend on fixed costs prior to starting trial	£90,216	HERO trial's accounts
	Estimated spend on fixed costs during trial	£204,581	HERO trial's accounts
	Estimated spend on variable costs	£204,581	HERO trial's accounts
	Estimated spend on fixed costs post follow-up	£336,042	HERO trial's accounts
c_{fixed}	Total spend on fixed costs	£630,839	HERO trial's accounts
	Total spend	£835,419	HERO trial's accounts
c	Estimated cost per pairwise allocation	£1,650	HERO trial's accounts
λ	Maximum willingness to pay for one QALY	£30,000	NICE (2013)

Table 3.D.1: Parameter values for the HERO case study

Chapter 4

Guidance material: Practical considerations for a clinical trials unit interested in using a value-based sequential design in a future clinical trial

Martin Forster, Laura Flight, Alan Brennan and Stephen Chick

In this chapter we present some practical matters for a clinical trials unit (CTU) to consider if it is interested in using a value-based sequential approach in a future clinical trial. The considerations are based on the experience that we have gained from carrying out the ProFHER, CACTUS and HERO case studies and running the 'mini course' for CTU ENACT collaborators from the University of York and The University of Sheffield between March and August 2020.

As a precursor to the discussion, we recall the comments in the proof-of-concept application of the model to the ProFHER pragmatic trial, also noted in section 1.2 of this report, in the context of the model's potential applicability to aid decision-making during a clinical trial:

> '... we are interested in whether such a [value-based sequential] rule could complement [existing clinical trial] designs, by providing additional information to trials teams about whether interim evidence suggests that the benefit of randomising further patients into the trial is worth the cost ...' (Forster et al., 2021, Section 1)

We stress that, owing to the nascent nature of this literature, our suggestion is to consider using the value-based sequential design to complement, rather than replace, existing clinical trial designs. Section 5.1 describes some of types of trial which may benefit from use of the design (see, in particular, points 5 and 6).

4.1 Access to appropriate expertise

We recommended that a CTU considering using the value-based sequential design has access to expertise in the following areas:

1. adaptive clinical trial designs in general, both theory and application (Pallmann et al., 2018; Dimairo et al., 2020);

2. Bayesian decision theory, in particular the theory behind sample size determination for the value-based one stage design using a normal prior–normal likelihood approach (Claxton and Posnett, 1996; Claxton, 1999) and the recommendations of the 'Emerging Good Practices Task Force' for Value of Information analysis (Fenwick et al., 2020; Rothery et al., 2020);

3. the theory and application of dynamic programming (Puterman, 1994; Bertsekas, 2005);

4. the value-based sequential design that is presented in this report (Chick et al., 2017a; Alban et al., 2020) and the relevant software (Chick et al., 2017b).

Expertise may be available in-house, from external sources (such as the collaborators on the ENACT project and other researchers working in the area of value of information analysis and adaptive designs) and/or training courses.

We make these recommendations in the light of the important work of Jaki (2013) and Dimairo et al. (2015), who considered why application of adaptive approaches to clinical trial design lags behind their methodological development. Jaki (2013) studied the use of adaptive and Bayesian methods in early stage public research studies in the UK. Using the results of questionnaires sent to 39 clinical trials units, barriers to uptake included: lack of expertise and training (at both degree level and 'on the job'); lack of software; the length of time required to plan such a design and understand its operating characteristics; inadequate funding and difficulties convincing an investigator to change method. Considering the world of public sector confirmatory trials, Dimairo et al. (2015) interviewed 27 stakeholders in clinical trials research, including CTU directors, members of independent data monitoring committees, chief investigators, statisticians and health economists. Barriers hampering the uptake of adaptive designs included: lack of understanding of what is meant by an adaptive design; a conservative outlook (e.g. fear of making the wrong decision, concerns about premature stopping); lack of knowledge and experience; complexity of the design; concerns about trial credibility, integrity and validity and lack of bridge funding to support design work. Similar findings have been reported for the use of value of information methods in real-world decision-making (Fenwick et al., 2020).[1]

[1]Training needs in the areas of Bayes' theorem, dynamic programming, Bayesian decision theory and Matlab

We note that various ongoing initiatives may help to increase awareness and application of adaptive and value-based methods. These include the Collaborative Network for Value of Information (ConVOI, https://www.convoi-group.org/members/), A Practical Adaptive & Novel Designs and Analysis toolkit (PANDA, https://www.sheffield.ac.uk/scharr/research/centres/ctru/panda), Rapid Assessment of the Need for Evidence (RANE, https://shiny.york.ac.uk/rane/), the Adaptive Designs Consort Extension (ACE, https://www.sheffield.ac.uk/scharr/research/centres/ctru/ace), the Sheffield Accelerated Value of Information tool (SAVI, http://savi.shef.ac.uk/SAVI/) and the Costing Adaptive Trials project (CAT, 2020).

4.2 Value-based adaptive designs which were not the focus of this report

The ENACT project has focused on one particular kind of value-based sequential design: the Bayesian decision-theoretic model of a sequential experiment proposed by Chick et al. (2017a) and Alban et al. (2020). This is only one of a number of approaches to running an adaptive clinical trial which is informed by value-based considerations. Table 4.1 outlines a range of other approaches which have been proposed.

4.3 The value-based sequential design that is the focus of this report

4.3.1 Questions to ask when considering the value-based sequential design

We anticipate that the value-based sequential design could be used within a framework of existing decision-making machinery for a clinical trial, that is, by complementing a planned fixed sample size or a sequential/adaptive design which uses frequentist principles (Hampson and Jennison, 2013; Zagoraiou, 2017; Baldi Antognini and Zagoraiou, 2014). Figure 4.1 presents a list of questions that a CTU interested in using the value-based sequential design could consider. Answering 'yes' to each question suggests that the value-based sequential design could be used.

were also identified for the members of the ENACT project who were charged with delivering the two case studies. These were addressed in the three day 'mini course' which ran before work started on the CACTUS and HERO case studies. We note that the concerns discussed here are not new: Chernoff and Petkau (1981) noted that the 'formidable' nature of the methods required to solve Bayesian, sequential, decision-theoretic models – the research of Chernoff and his collaborators provides the foundations for the value-based model that is presented in this report – could hamper their uptake.

Topic of interest	Description
Value-adaptive multi-arm clinical trials intended to assess different therapies, combinations of therapies or Phase II/III dose finding.	Multi-arm trials may have correlated expected value of incremental net monetary benefit for different arms. For example, in a dose-finding trial, similar dose levels are likely to have a more similar expected value of incremental net monetary benefit than very different dose levels. Adaptive allocation policies for highly sequential, value-based, multi-arm trials have been proposed and show promise in identifying optimal doses (Chick et al., 2019, 2021).
Accounting for the health outcomes accruing to participants in the trial.	The ProFHER, CACTUS and HERO case studies do not consider the benefits which accrue to participants in the sequential trial. Chick et al. (2017a) and Ryzhov et al. (2012) discuss how such benefits can be incorporated into a value-based, sequential, model.
Adapting the fraction of participants allocated to each arm of a clinical trial	A dynamic programming approach to a group sequential trial has been proposed for Bernoulli (0-1) outcomes to vary the fraction of participants assigned to each arm, based on accumulating data (Ahuja and Birge, 2016). The model aims to improve outcomes for participants in the trial, as well as the probability of correctly selecting the best treatment. This model is not explicitly value-based, but could be adapted to be so if the 0-1 outcomes are weighted by estimates of net monetary benefit.
A value-based approach to inform the optimal rate of recruitment.	The value-based sequential design considered in the ENACT report assumes a constant recruitment rate to the trial. If the recruitment rate is nonlinear – for example, a trial manager may prioritise sites for opening because they are more likely to have a higher rate of successfully enrolled and retained participants – this information can be used to optimise the recruitment rate. (Alban et al., 2020, Section 5.2) discuss one approach to making the recruitment rate adaptive.
Using batch allocations of participants.	In contrast to the 'pairwise allocation' approach which characterises the ProFHER, CACTUS and HERO case studies, batch adaptive trials permit multiple participants to be assigned to multiple treatments in one batch. Several methods have been proposed to allocate batches to a finite number of arms using Bayesian expected value of sample information methods (Wu and Frazier, 2016; Villar and Rosenberger, 2018).

Table 4.1: A summary of value-based approaches to sequential/adaptive clinical trial design which were not the focus of this report.

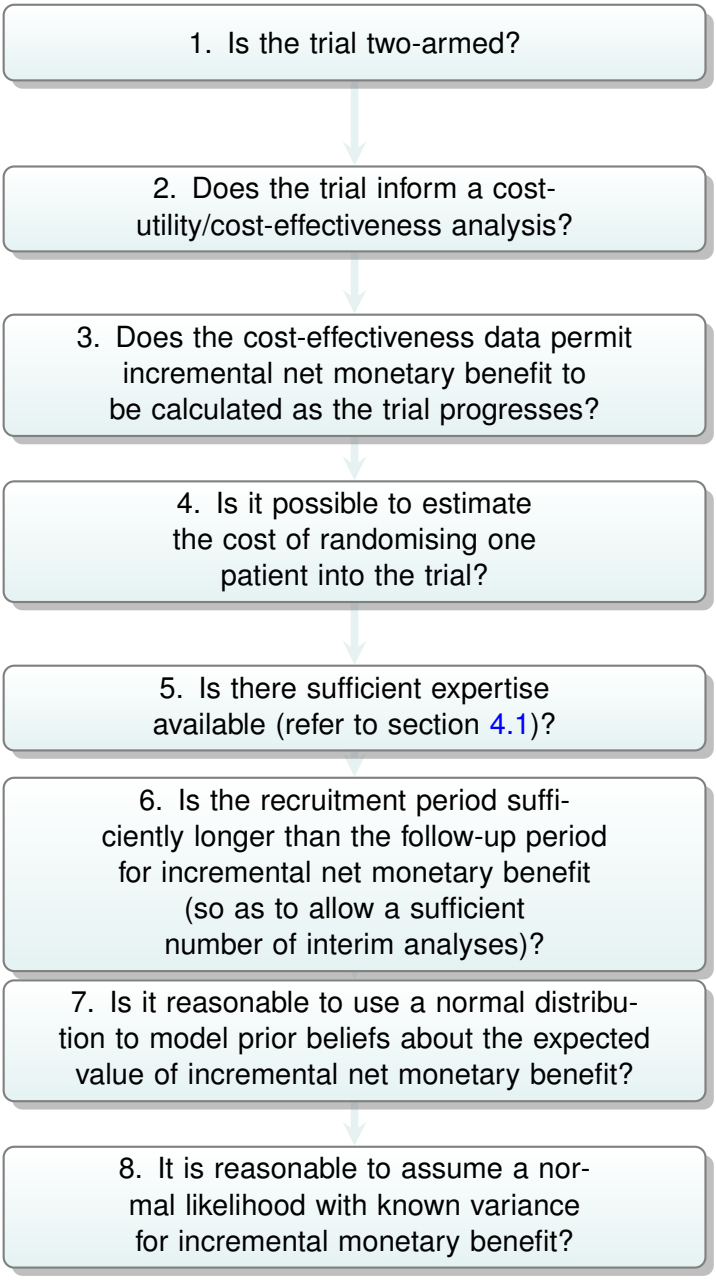


Figure 4.1: Flow chart presenting a set of questions for a CTU to answer when considering whether to use the value-based sequential design that is presented in this report.

Notes:
Answering 'yes' to each of the questions suggests that the value-based sequential design could be used. There is some flexibility permitted when considering these questions. For example, the Big CACTUS trial had three arms, but the Big CACTUS case study presented in Chapter 2 assumes only two arms; variable research costs may not be known for sure, but their variability may be investigated using sensitivity analysis.

There is some flexibility when answering some of these questions. For example, as discussed in Chapter 2, the Big CACTUS trial had three treatment arms, but the analysis presented in this report focuses on only two arms. Estimation of research costs is likely to be subject to uncertainty, and this is something which can be investigated using sensitivity analysis (discussed in the next section). Although the normal prior–normal likelihood assumption may not be realistic, we consider this as a reasonable approximation and have flagged further research on the matter in Chapter 5.

4.3.2 Practical considerations when developing a value-based sequential design

Fenwick et al. (2020) suggest the following approach to carrying out a one-stage value of information analysis:

1. conceptualise a health economic model;
2. parameterise with evidence;
3. generate a probabilistic sensitivity analysis;
4. identify uncertainty;
5. establish whether further research is worthwhile;
6. estimate the value of specific research;
7. iterate with new evidence.

It should be noted that the value-based sequential design is adaptive in nature. That is, it permits the trial team to learn about the accumulating cost-effectiveness evidence by updating the posterior mean for incremental net monetary benefit as evidence accumulates. Then, it informs the stopping decision through a comparison of the posterior mean with the model's stopping boundary. This process does not take place in the value-based one stage design. Hence the above list, while useful, is not exhaustive. We develop this point in the following sections.

1. Conceptualise the health economic model

The health economic model that is considered in this report is the value-based sequential design of Chick et al. (2017a) and Alban et al. (2020). section 1.4 provides an overview. Deployment of the model proceeds in two stages:

- prior to running the trial, best estimates of parameter values for the model are chosen, so that a stopping boundary can be obtained for Stage II of the trial (see point 1 below);

- as Stage II proceeds, accumulating evidence about cost-effectiveness that arrives from pipeline patients is used to update the prior/posterior mean for the expected value of incremental net monetary benefit. During Stage II, at a defined set of interim analyses, the posterior mean for the expected value of incremental net monetary benefit is compared with the stopping boundary to establish whether the trial should continue or stop.

2. Parameterise with evidence

Column (2) of Table 4.2 lists the parameter values that are required to solve the value-based sequential design. Column (1) presents the parameter values that are required to solve the value-based one stage design. The value-based sequential design requires the same set of parameter values as the value-based one stage design, together with: (1) parameters which define the starting time for Stage II of the value-based sequential trial (equal to the delay in following up the estimates of cost-effectiveness); (2) the number of participants expected to be in the pipeline when Stage II starts (this information helps determine the expected reward from stopping the trial at the start of Stage II – further details may be found in Chick et al. (2017a, Section 2.1.1)); (3) the number of interim analyses that take place during Stage II.

Section 2.B (CACTUS case study) and section 3.D (HERO case study) present examples of how parameter values were chosen for the two case studies presented in this report. Choice of parameter values may be informed by the research proposal itself, data which are available before the trial begins (such as data from a pilot trial, as in the CACTUS case study), discussions with clinical researchers and/or following existing guidelines for good practice (NICE, 2014; Drummond et al., 2015; Fenwick et al., 2020). The rest of this section discusses in further detail some of the issues surrounding the choice of the parameter values that are listed in Table 4.2. We do not attempt to provide a full account of matters for consideration, rather the focus is on anything that is special about the value-based sequential model, noting that there is already a lot of guidance for good practice in the literature.

1. *Size of the population to benefit from the technology adoption recommendation*. Both case studies estimate the size of the population to benefit from the result of the trial by multiplying the estimated incidence rate for the condition of interest by the time horizon over which the adoption decision is deemed to apply. An alternative approach involves discounting the sum of the incidence rate over the relevant time horizon to obtain a 'net present value' for the population size (Claxton, 1999, Eq. (6b)) or (Rothery et al., 2020, Eq. (15)).

2. *Standard deviation for incremental net monetary benefit*. The ProFHER and HERO case studies estimated the sampling standard deviation for incremental net monetary benefit from the data available at the end of the trial. This was deemed a reasonable approach,

Parameter	Required for the value-based: (1) one stage model?	(2) sequential model?
Size of the population to benefit	Yes	Yes
Standard deviation for incremental net monetary benefit in population	Yes	Yes
Prior mean for expected incremental net monetary benefit	Yes	Yes
Prior variance for expected incremental net monetary benefit	Yes	Yes
Estimate of mean number of pipeline patients	No	Yes
Delay in observing cost-effectiveness data	No	Yes
Fixed cost of adopting new treatment	Yes	Yes
Total spend on fixed costs	Yes	Yes
Cost per pairwise allocation	Yes	Yes
Maximum willingness to pay for one unit of effectiveness	Yes	Yes
Number/timing of interim analyses during Stage II	No	Yes

Table 4.2: Parameter values required to solve the value-based one stage and sequential models which are the focus of this report.

Notes:
The table shows that the value-based sequential approach uses all of the parameter values that are required for the value-based one stage approach, plus those which relate to the length of the recruitment period, the rate of recruitment and the follow-up period for cost-effectiveness data.

owing to the retrospective nature of the case studies and their 'proof of concept' status. The CACTUS case study used the pilot study data to estimate the standard deviation (refer to section 2.3.8). Other approaches to estimating the standard deviation could include using data from existing published studies or using information from the sample size calculation for the frequentist design (see e.g. Handoll et al. (2015, pages 11–12) for the ProFHER pragmatic trial). However, these approaches may be of limited use owing to the fact that the standard deviation is required for incremental net monetary benefit, and this is unlikely to be the same as the standard deviation of the primary clinical outcome in the trial.

3. *Prior mean and variance*. For the HERO case study, it was assumed that there was little evidence available to inform the choice of prior mean for the expected value of incre-

mental net monetary benefit prior to the trial commencing. Hence the prior mean was set equal to zero. The variance of the prior distribution was set to a value which implied the information contained within the prior distribution (the 'effective sample size' of the prior information) was low (equal to 2). Since the effective sample size is the ratio of the sampling variance (the square of the standard deviation for incremental net monetary benefit that was discussed in the previous point) to the prior variance (discussed in this point), it is straightforward to obtain the prior variance by rearranging this expression.

The CACTUS case study used the pilot study to inform the choice of parameter values for the prior distribution (section 2.3.8).

4. *Estimate of the expected number of pipeline patients present at the start of Stage II and the delay in observing the cost-effectiveness data*. These parameters, which are required for the value-based sequential design but not the value-based one stage design, are used to calculate the expected reward from stopping the trial during Stage II. The estimate of the expected number of pipeline patients present at the start of Stage II can be calculated by multiplying the expected rate of recruitment into the trial by the time to follow-up the cost-effectiveness data. Estimating the likely recruitment rate should be a standard part of clinical trial design (see e.g. Handoll et al. (2015, pages 12–13) for the ProFHER trial). Estimating the time to follow-up can be informed by the research proposal, in particular, the points in time at which generic health outcome data such as the QALY, and treatment costs, are measured. It is possible that trial participants are followed up on more than one occasion during the trial. On this point, the results from the ProFHER and HERO case studies are instructive: the longer is the delay that is chosen for following up cost-effectiveness data, the less is the value that can be delivered by the value-based sequential design.

5. *Research costs*. We note that treatment costs will already be included in the estimate of incremental net monetary benefit and so should not be considered as part of the costs of the trial, to avoid counting these costs multiple times. For an NIHR-funded trial, there will be three main categories of cost:

   (a) costs incurred by the funder. These typically include higher education institute (HEI) costs such as staff and estate costs, as well as costs associated with conducting the research, such as recruiting and randomising participants. Funding bodies such as the NIHR typically cover 80% of HEI costs;

   (b) National Health Service support costs. These may include costs associated with employing National Health Service staff, such as research nurses to conduct the research through, for example, the NIHR Clinical Research Network;

(c) National Health Service treatment costs/savings requested from the National Health Service. These will include the costs associated with delivering the intervention on the National Health Service.

The precise set of costs to be included in the design of the study will depend on the study perspective. A societal perspective may consider all research costs, regardless of where they are incurred and who would pay them (for example, 100% of HEI costs). A funder perspective, such as that adopted in the CACTUS and HERO case studies, may only consider the costs incurred by the funder itself. We note that a sister project, the CAT project (CAT, 2020), is investigating the measurement of research costs for adaptive designs in more detail and might provide valuable information for informing cost calculation.

Absent definitive guidance about how to measure the costs of research, we note that the split of research costs into fixed (costs which are incurred independently of the number of patients recruited) and variable (costs which vary according to the number of patients recruited) elements was approached differently in the CACTUS and HERO case studies:

(a) In the CACTUS case study, research costs were allocated at the patient level. They were separated into costs incurred during the recruitment and follow-up periods, which permitted the calculation of the potential savings of staff and meeting costs, should the trial finish early (see section 3.4.1).

(b) In the HERO case study, research costs incurred during the recruitment phase were split into fixed and variable elements according to a predefined ratio (50:50), following the approach taken by Forster et al. (2021) for the ProFHER trial. This means that only variable costs are saved should the trial stop early. While the 50:50 split represents a simple approach to the problem of distinguishing fixed costs from variable, it is not supported by evidence. As a result, it might under-estimate both the savings that could be achieved due to early stopping (because some fixed costs associated with the trial design might also be saved), as well as the additional costs that could be incurred should the trial run late (because additional fixed costs might be incurred).

Finally, both the CACTUS and HERO case studies assume that the fixed costs of the trial do not vary according to the type of design (one stage or sequential). In practice, there are likely to be differences. For example, the value-based sequential design is likely to incur additional research costs to cover design, carrying out interim analyses, and providing support. As described by Flight (2020), interim analyses will require a number of additional costed tasks, including data extraction and cleaning, data analysis

and report writing. Research teams may therefore wish to consider allocating different fixed costs to each design.

6. *Willingness to pay for health*. Health benefits are valued at a defined willingness to pay for one unit of effectiveness (sometimes referred to as the 'cost-effectiveness threshold'). For example, in the National Health Service, the willingness to pay for one quality-adjusted life year is often taken to be between £20,000 and £30,000 (NICE, 2013). Choice of this parameter is common to both the fixed and sequential value-based designs. For further details see Drummond et al. (2015).

7. *Number/timing of interim analyses in Stage II*. The solution to the value-based sequential design permits the posterior mean to be compared with the stopping boundary at any point during Stage II. This opens up the possibility of monitoring data from the trial continuously. In practice, this is unlikely to be practical owing to the large demands on the time of the research team. As a result, it is necessary to make a choice about the number and frequency of the interim analyses which take place during Stage II. In the two applications presented in this report, it was assumed that interim analyses would take place in blocks of five (the CACTUS case study) ten (the HERO case study) pairwise allocations at a time. Conduct of interim analyses poses at least three challenges for the researcher: the randomisation process used in trials does not deliver patients to the arms in a pairwise manner; the choice of the block size for the interim analyses is a function of the number of interim analyses chosen; based on evidence from the ProFHER, Big CACTUS and HERO case studies, there are likely to be missing data in the cost-effectiveness data.

   The smaller the block size, the greater the scope for stopping the trial but (most likely) the higher the costs of administering it. In addition, it will be important to consider what can be meaningfully concluded if the study stops at an early interim analysis, resulting in a low overall sample size. This is a matter that is discussed in the context of the ProFHER by Forster et al. (2021).

   Regarding the assumption of pairwise allocation to the two arms of the trial and the handling of missing data, an approach is required which provides a reasonable way of 'fitting the data to the model' in the absence of randomisation using a block size equal to two. For the ProFHER case study, the randomisation order of patients was amended so that patients could be analysed in pairs, maintaining as closely as possible the original chronological ordering of the data. Once re-ordered, the estimate of the expected value of incremental net monetary benefit was obtained for blocks of ten pairwise allocations (that is, 20 patients) at a time and used to obtain the posterior mean for expected value of incremental net monetary benefit. The missing data mechanism was assumed to be

'completely at random', meaning that the estimates used were assumed not to be biased. This approach was believed to be a reasonable assumption for a 'proof of principle' study such as the ProFHER case study, but is unlikely to be realistic.

For the Big CACTUS and HERO applications, handling the pairwise allocation assumption and missing data used more sophisticated methods. The approaches are best understood by reviewing the methodology that was used – see section 2.3.8 (for Big CACTUS) and section 3.B.3 (for HERO).

3. Generate a probabilistic sensitivity analysis

To date, in the three case studies that have applied the value-based sequential design, the parameter values that are used to obtain the stopping boundaries are treated as being fixed and not subject to any uncertainty. However, reflecting the advice of Fenwick et al. (2020), prior to the trial taking place, it would be feasible to investigate how sensitive the Stage II stopping boundary is to variations in the values of parameters which are, it is believed, subject to the highest levels of uncertainty. For example, the stopping boundaries could be run under the assumption that the variable research costs and/or the standard deviation for incremental net monetary benefit are double, or half, their estimated values. This is an under-explored area.

4. Identification of uncertainty to 7. Iterating with new evidence

These points are less relevant for the value-based sequential design, which incorporates the idea of learning to reduce the level of uncertainty, and stopping when the benefits of further research are deemed to be not worth the costs.

4.4 Using the value-based sequential design model prospectively

The above points all refer to the choice of parameter values prior to the trial starting, and using them to define a stopping boundary for Stage II of the sequential trial once it has started. We note that the only case studies to apply the value-based sequential designto date have all been retrospective, so there is little guidance available for how monitoring could take prospectively. Nevertheless, the case studies have provided some insights into two matters that could be encountered in a prospective deployment: the use of a model-based analysis (CACTUS) and the handling of missing data (HERO). We briefly discuss these matters here, noting that further detail is provided in the case studies themselves.

4.4.1 Using a model-based analysis during Stage II

The CACTUS case study used a model-based analysis to compare the value of the posterior mean with the stopping boundary during Stage II. Key issues to consider for a model-based analysis include the development of the model and the complexity of analysing it on multiple occasions as the sequential trial progresses. In the CACTUS case study, we have the advantage that a simple Markov model had been developed as part of the CACTUS pilot trial. This provided a solid starting point for developing the more complex model required for the Big CACTUS trial. In a prospective application of the value-based sequential design, this model would need to be developed at the design stage of the trial, possibly even before the trial is funded, to provide sufficient detail in the grant application. As Flight et al. (2020) highlight, this represents a large investment of (potentially unfunded) time and so may not be feasible when there are no existing health economic models to work from and a lack of 'bridge' funding for researchers to draw on for development work (Dimairo et al., 2015). Research teams may consider using a simplified model to illustrate the value-based design at the grant writing stage, but then include funding to develop the model further at the start of the funded period, before the trial is under way.

An additional consideration is the requirement to analyse the model at multiple interim analyses over the course of the trial. For the CACTUS case study, the model had a manageable number of parameters estimated from the trial data and the computational burden of running the model was not too high. Research teams will need to consider whether it is feasible to repeat their model-based analyses multiple times during the trial and whether their choice of software facilitates this.

4.4.2 Handling of missing data

For the HERO trial, the bootstrap analysis that was carried out provides an illustration of how missing data could be handled, so as to obtain a posterior mean to compare with the stopping rule at multiple interim analyses during Stage II. This matter has already been discussed in detail in section 3.C.

Chapter 5

Discussion of the main results and directions for future research

Martin Forster, Laura Flight, Stephen Chick and Alan Brennan

Chapters 2 and 3 of this document have reported the results of an application of the Bayesian decision-theoretic model of a sequential experiment proposed by Chick et al. (2017a) and Alban et al. (2020) to two new case studies from the UK. The case studies include a comparison of the performance of the value-based sequential design with two comparator designs: (a) the trial itself and (b) a fixed sample size design which maximises the expected net benefit of sampling for a Bayesian decision-theoretic model using a normal prior and a normal likelihood (the 'value-based one stage design', see Claxton and Posnett (1996); Claxton (1999)). In addition, the case studies have added to existing knowledge by exploiting special features of the Big CACTUS and HERO trials, including the use of a model-based health economic analysis, pilot trial data and missing data. All of these objectives were part of Deliverable 3 of the ENACT project. Chapter 4 of this report has presented a summary of procedures that a clinical trials unit could follow if it is considering using the value-based sequential design in a future clinical trial.

Throughout this report, we have taken care to stress that the results of the case studies are related to the assumptions that have been made to set up and solve the value-based sequential model, together with the assumptions that have been made in the applications themselves, in particular as they relate to the choice of parameter values. The value-based sequential design takes a normal prior–normal likelihood approach to the Bayesian updating for the unknown expected value of incremental net monetary benefit and assumes that the sampling variance for incremental net monetary benefit is known. In the applications, we have assumed that the rate of randomisation into the trial (and the subsequent arrival of outcome data) is constant and that interim analyses take place in blocks of five (CACTUS case study) or ten (HERO case study)

pairwise allocations at a time. Regarding the choice of parameter values, we have had to make assumptions about the size of the population to benefit from the technology adoption decision, the estimate of the fixed costs of the trial and the variable cost of randomising and monitoring patients, the choice of the prior mean and variance for the distribution of the expected value of incremental net monetary benefit, the choice of the maximum sample size of the sequential trial and the maximum willingness to pay for an additional Quality Adjusted Life Year. Finally, we have not applied discounting to any trial. Although making assumptions such as these may appear restrictive, the process brings to the fore the meaning of the term 'value-based', making explicit consideration of the costs of the trial, the size of the population to benefit and the health care system's willingness to pay for health gain in the definition of 'value'.

5.1 Discussion of the main results from the case studies

Table 5.1 summarises some of the main results from the two case studies, together with the ProFHER case study of Forster et al. (2021). The table's 'base case' results for the value-based sequential design refer to the version which uses as its maximum sample size the sample size of the value-based one stage design. The main results to emerge from the two case studies, plus the ProFHER case study, are the following:

1. In all three case studies, the expected value delivered to the health care system by the value-based one stage design is higher than the expected value delivered by the design used in the original trial. This result is not new, and goes back at least as far as Claxton (1999): as long as the sample size of the original trial design is not the same as the sample size which maximises the expected net benefit of sampling, the value-based one stage design delivers more value.

2. However, the quantitative gain of the value-based one stage design over the design of the original trial can be small, even though the sample sizes of the two designs differ by a fair margin. This is clear from Table 5.1 (refer to the entries under *(Expected) sample size* and *Expected net benefit*), where the increase in the expected net benefit delivered by the value-based one stage design is less than one percent for the HERO and ProFHER case studies, despite the value-maximising sample sizes differing by +43% (HERO) and -10% (ProFHER). The result arises because the expected net benefit of sampling function is reasonably flat in the interval defined by the potential sample sizes of the trials. Section 3.4.1 provides further discussion on this point for the HERO trial.

3. In all three case studies, the expected value delivered to the health care system by the 'base case' value-based sequential design is higher than the expected value delivered by

	CACTUS case study	% increase over the trial	HERO case study	% increase over the trial	ProFHER case study	% increase over the trial
(Expected) sample size						
The trial	*95	-	*124	-	*125	-
Value-based one stage design	†132	38.95	†177	42.74	‡112	−10.40
Base case value-based sequential design	†100	5.26	†174	40.32	†73	−41.60
Expected cost (£000000)						
The trial	*1.22	-	*0.84	-	*1.47	-
Value-based one stage design	†1.39	13.93	†0.92	10.5	‡1.42	−3.40
Base case value-based sequential design	†1.24	1.64	†0.92	9.88	†1.25	−14.97
Expected net benefit (£000000)						
The trial	†3.54	-	†52.0	-	‡51.2	-
Value-based one stage design	†3.60	1.69	†52.0	0.00	‡51.2	0.01
Base case value-based sequential design	†3.85	8.76	†52.1	0.19	‡51.2	0.27
Estimated probability of concluding the new technology is cost-effective						
The trial	*0.32	-	*0.39	-	*0.20	-
Value-based one stage design	†0.27	−15.63	†0.44	12.82	‡0.05	−75.00
Base case value-based sequential design	†0.34	6.25	†0.44	11.79	†0.09	−55.00

Table 5.1: A summary of some of the main results from the CACTUS and HERO case studies, along with those of the ProFHER case study.

Notes:
Results for the base case value-based sequential design reported in this table are for designs which use as their maximum sample size the sample size from the value-based one stage design. A higher expected net benefit may be obtained by increasing the maximum sample size. Percentage changes are calculated using the trial results as the base case. *Sources*. For the CACTUS case study: Palmer et al. (2015), Latimer et al. (2020) and related information (∗) and Tables 2.3 and 2.4 of this report (†), assuming a maximum willingness to pay (WTP) for one QALY of £20,000. For the HERO trial: Ronaldson et al. (2021) and related information (∗) and Tables 3.2 and 3.3, assuming a maximum WTP for one QALY of £30,000. For the ProFHER case study: Handoll et al. (2015, Table 85) and related information (∗), Forster et al. (2021, Tables 2 and 3) (†) and new analysis carried out for this report (‡), assuming a maximum WTP for one QALY of £20,000.

the value-based one stage design. This, too, is not a surprising result: when the value-based sequential design has as its maximum sample size the value-maximising sample size of the value-based one stage design, the benefit of operating a trial sequentially is that the trial can be halted when the posterior mean for the expected value of incremental net monetary benefit indicates that the expected benefit of randomising a further pair of patients into the trial is not worth the cost. Such flexibility is not offered by any one stage design, since they offer no opportunity for early stopping.

4. Once more, however, it is instructive to consider the quantitative size of the gain in value when interpreting this result: the largest gain of the value-based sequential design over the value-based one stage design is found in the CACTUS case study (approximately 7%). Results from the ProFHER and HERO case studies show the gain to be negligible (less than 1%). These results are based on an 'all other things being equal' comparison, which assumes that running a sequential trial does not incur any additional costs. If the sequential design incurs higher monitoring costs, for example, this could eliminate any expected gain of sequential over one stage.

5. The time to follow-up cost-effectiveness data in the HERO and CACTUS case studies was relatively long when compared with the total time required to recruit patients to the trial. In such scenarios, there can be limited benefits to using the value-based sequential design. This is because patients have already been recruited to the trial – meaning that research costs have been committed and information obtained, albeit with delay – by the time Stage II of the value-based sequential model starts. Section 3.4.1 provides more discussion of this point, in the context of the HERO trial. This result suggests that some of the greatest potential for the value-based sequential design could lie in deployment in trials with a relatively quick follow-up as a proportion of total trial length.

6. The value-based sequential design may also perform well in trials where one technology is clearly more cost-effective than the other. This result is notable in the ProFHER pragmatic trial, where the bootstrap analysis showed that the value-based sequential design led to an expected sample size which was between 35% and 42% smaller than the two comparator fixed sample size designs. As a result, the analysis suggested a research cost saving of approximately 15%, with a probability of about 91% of concluding that surgery was not cost-effective (which was the result in the original trial). This result was driven by the fact that one technology (surgery) was much more expensive than the comparator (sling immobilisation) with little, if any, evidence to show that it was more effective. In this scenario, the advantage of the sequential design in stopping the trial early is clear.

7. Following on from the previous point, the value-based one stage design and the value-

based sequential design do not discriminate well between technologies when there is an equivocal signal from the trial. This is best illustrated by the HERO case study, where the differences in effectiveness and cost between the two technologies were negligible, resulting in a posterior mean which tracks zero and is (almost) equally likely to conclude in favour of hydroxychloroquine as it is in favour of placebo (refer to Table 3.3).

8. Points 1 to 5 above deal with the expected value delivered to the health care system by the competing trial designs. However, as Table 5.1 and point 6 show, other operating characteristics, such as the expected sample size of the trial, or the expected cost of the research, may be important considerations for the system. For example, in the CACTUS case study, the value-maximising sample size of the value-based one stage design is 39% higher than the sample size of the original Big CACTUS trial, implying that the cost of the value-based one stage design is expected to be about 14% higher. But the value-based sequential design which uses as its maximum sample size the value-maximising sample size of the value-based one stage design has an expected sample size which is roughly equal to the design of the original Big CACTUS trial (it is only about 5% larger), while delivering additional expected net benefit (+10.5% when compared with the design of the trial). In contrast, for the HERO case study, the sample size of the value-based one stage design and the expected sample size of the value-based sequential design are about 42% higher and deliver very little additional expected net benefit. This raises the question of study perspective: which operating characteristic is key, and which are secondary, when considering the performance of a study's design?

9. Results from the CACTUS and HERO case studies show that increasing the maximum sample size of the value-based sequential design increases the expected net benefit, but that (once again), consideration of the quantitative gains is important. For example, for the CACTUS case study, the gain is approximately 13% when the maximum sample size of the value-based sequential design is set to double the optimal sample size from the value-based one stage design. The equivalent gain for the HERO trial is only 0.4%.

5.2 Directions for future research

There are numerous directions for future research:

1. Further work with the existing case studies could investigate how sensitive the results are to the choice of parameter values. This could be done as the case studies are prepared for publication. We note the many methods available to handle parameter value uncertainty for the value-based one stage design (Rothery et al. (2020) provide an overview) and

that equivalent methods are not currently available for value-based sequential designs. As well as exploring sensitivity with respect to standard parameters of interest, such as the cost of sampling and the size of the population to benefit, it would be interesting to explore, in more detail, how the value delivered by the value-based sequential design changes as the maximum sample size changes and the time to follow-up is reduced (developing the analysis of Figure 3.6).

2. Further retrospective case studies could be carried out. In particular, it would be interesting to apply the value-based sequential design to another ProFHER-like trial, in which it is believed that one of the two technologies is substantially more expensive than the other, for the reasons outlined in point 6 of the previous section.

3. Work could explore how the value-based sequential design may be used to carry out subgroup analyses during the trial. The Big CACTUS trial showed that computerised speech and language therapy is potentially cost-effective in patients with mild to moderate word finding abilities (Latimer et al., 2020). Further work could explore how this subgroup analysis is affected by a value-based stopping rule that considers the cost-effectiveness of research from the perspective of the whole population. For example, if the stopping rule is determined by the 'aggregate' signal from the trial, would early stopping provide sufficient data to carry out subgroup analyses? Or could the value-based sequential rule itself be informed by the analysis of subgroups?

4. The CACTUS case study considered only computerised speech and language therapy with usual care. The value-based sequential design could be adapted to consider a three arm trial, that is, comparing computerised speech and language therapy with usual care and attention control. The multi-arm, multi-stage, methods developed by Chick et al. (2021) could be used.

5. Flight (2020) have shown that adaptive designs have the potential to impact a health economic analysis, by biasing point estimates, owing to the adaptive nature of the design. Further work is required to understand the potential for bias in value-adaptive methods and what adjustments are required to maintain an accurate analysis when these designs are used. In addition, analysis could inform the degree to which the value-based sequential design affects the precision of estimates of effectiveness and cost-effectiveness.

6. As explained in point 5 of the previous section, where the time to follow-up comprises a large proportion of the overall recruitment period, there is limited scope for applying the value-based sequential design. In such situations, one could consider using shorter term health economic outcomes which act as a surrogate to inform interim analyses. To date,

no work has been carried out to investigate such a possibility, nor to relate this result to the wider literature on adaptive designs, where our understanding is that this is a well known issue.

7. The case studies assume that recruitment is constant over the duration of the trial and patients arrive in pairs, so that the number of participants is equal between arms at each interim analysis. In reality, recruitment of participants is likely to be non-linear during the trial and at a given interim analysis there may be an unequal number of participants in each arm. Relaxing the assumption of a constant rate of recruitment and pairwise allocations could be explored in further work.

8. As highlighted by Flight et al. (2020), stakeholders in health technology assessments consider clinical effectiveness to be the main focus of a clinical trial. They may be reluctant to consider using a design which considers cost-effectiveness criteria alone. The three case studies do not consider the impact of the value-based sequential design on the clinical effectiveness data, nor do they consider how sequential monitoring of clinical and cost-effectiveness data may be carried out together. Further work with the case studies could consider how evidence of clinical effectiveness accumulated over the course of the trial, and the strength of clinical evidence upon stopping. Further comparisons could be drawn between stopping rules for group sequential trials, such as those proposed by Pocock (1982) and O'Brien and Fleming (1979), and the value-based stopping rule.

9. The model that has been applied to the case studies assumes that the sampling variance is known. Chick et al. (2017a) also provide the solution for the case of unknown sampling variance, and this could be applied to the case studies to extend the analysis.

10. Additional operating characteristics could be considered, such as probability that the new technology is considered to be cost-effective along the stopping boundary for the sequential trial, and the relationship between the probability of cost-effectiveness as estimated in the sequential trial and the more traditional estimates that emerge from existing economic evaluation of health technologies such as the cost-effectiveness acceptability curve (Fenwick et al., 2001).

11. Further work could investigate the relationship between the information that is used to inform the prior distribution for expected incremental net monetary benefit and the eventual result from the trial. For example, the CACTUS pilot study was used to obtain a positive prior mean for expected incremental net monetary benefit. However, the Big CACTUS trial itself suggested that the true value of expected value of incremental net monetary benefit was negative. Work could investigate the influence of choice of prior on the op-

erating characteristics that are used in this report and investigate why the prior/posterior results were so different.

12. None of the models have considered using a non-zero discount rate to calculate the size of the population to benefit from the adoption decision. Nor have they considered the use of discounting during the recruitment and follow-up phases of the trial. This could be investigated further.

13. Further work could investigate how robust results of the case studies are to relaxing the normal prior–normal likelihood assumption that is used to solve the mode.

14. Further work could investigate the performance of the value-based sequential design under a range of assumptions about how to handle missing data, including using covariates.

15. The case studies have taken a conservative approach to measuring research cost savings, restricting them to savings in variable costs. This is necessitated by the lack of good guidance about how to cost adaptive/sequential clinical trials. The 'Costing Adaptive Trials' project (CAT, 2020) that was referred to in chapter 1 might provide helpful additional guidance.

Glossary of terms

- CACTUS: Cost-effectiveness of Aphasia Computer Treatment Compared to Usual Stimulation
- CSLT: Computerised Speech and Language Therapy
- CTU: Clinical Trials Unit
- E[INMB]: The expected value of incremental net monetary benefit
- ENACT: The ENACT project. See https://www.sheffield.ac.uk/scharr/research/centres/ctru/enact
- ENBS: The expected net benefit of sampling (Claxton and Posnett, 1996; Claxton, 1999)
- EVSI: The expected value of sample information (Claxton and Posnett, 1996; Claxton, 1999)
- HERO: Hydroxychloroquine Effectiveness in Reducing symptoms of hand Osteoarthritis
- INMB: incremental net monetary benefit
- NICE: The National Institute for Health and Care Excellence
- NIHR: National Institute for Health Research
- MI: Multiple imputation
- MICE: Multiple imputation by chained equations
- OA: Osteoarthritis
- PANDA: Practical Adaptive & Novel Designs and Analysis. See https://panda.shef.ac.uk/
- QALY: Quality Adjusted Life Year
- UC: Usual Care
- Value-based one stage design: A Bayesian decision-theoretic model for a normal prior–normal likelihood which chooses a fixed sample size to maximise the expected net benefit of sampling (Claxton and Posnett, 1996; Claxton, 1999)
- Value-based sequential design: The Bayesian decision-theoretic model of a sequential experiment proposed by (Chick et al., 2017a; Alban et al., 2020)

References

Ahuja, V. and Birge, J. R. (2016). Response-adaptive designs for clinical trials: Simultaneous learning from multiple patients. *European Journal of Operational Research*, 248(2):619–633. 81

Alban, A., Chick, S. E., and Forster, M. (2018). Extending a Bayesian decision-theoretic approach to value-based sequential clinical trial design. In Rabe, M., Juan, A., Mustafee, N., Skoogh, A., Jain, S., and Johansson, B., editors, *Proc. 2018 Winter Simulation Conference*, pages 2459–2470, Piscataway, NJ. IEEE, Inc. 56

Alban, A., Chick, S. E., and Forster, M. (2020). Value-based clinical trials: selecting trial lengths and recruitment rates in different regulatory contexts. Discussion Papers 1/20, Department of Economics, University of York. 7, 9, 15, 17, 21, 79, 80, 81, 83, 91, 99

Baldi Antognini, A. and Zagoraiou, M. (2014). Balance and randomness in sequential clinical trials: the dominant biased coin design. *Pharmaceutical Statistics*, 13(2):119–127. 80

Bell, M., Fiero, M., and Horton, N. (2014). Handling missing data in rcts; a review of the top medical journals. *BMC Medical Research Methodology*, 14(118). 12

Bertsekas, D. P. (2005). *Dynamic Programming and Stochastic Control: Volume I*. Athena Scientific, Belmont, MA, 3 edition. 79

Carpenter, J. R. and Kenward, M. G. (2012). *Multiple Imputation and its Application*. John Wiley & Sons Ltd., First edition. 52

CAT (2020). Costing Adaptive Trials 2020, Newcastle University Population Health Sciences Institute Biostatistics Research Group. Available online at https://www.newcastle-biostatistics.com/methodology_research/adaptive_designs/ (accessed 5 March 2021). 11, 80, 87, 98

Chernoff, H. and Petkau, A. J. (1981). Sequential medical trials involving paired data. *Biometrika*, 68(1):119–132. 80

Chick, S., Gans, N., and Yapar, O. (2019). Sequential, value-based designs for certain clinical trials with multiple arms having correlated rewards. In *2019 Winter Simulation Conference (WSC)*, pages 1032–1043, National Harbor, MD,. IEEE. 81

Chick, S., Gans, N., and Yapar, O. (2021). Bayesian sequential learning for clinical trials of multiple correlated interventions. *Accepted to appear* in Management Science. 81, 96

Chick, S. E., Forster, M., and Pertile, P. (2017a). A Bayesian decision-theoretic model of sequential experimentation with delayed response. *Journal of the Royal Statistical Society, Series B*, 79(5):1439–1462. 7, 9, 15, 17, 18, 21, 29, 56, 69, 79, 80, 81, 83, 84, 91, 97, 99

Chick, S. E., Forster, M., and Pertile, P. (2017b). htadelay package. https://github.com/sechick/htadelay. 21, 79

Claxton, K. (1999). The irrelevance of inference: a decision-making approach to the stochastic evaluation of health care technologies. *Journal of Health Economics*, 18(3):341–364. 7, 9, 13, 17, 19, 20, 21, 27, 56, 79, 84, 91, 92, 99

Claxton, K., Martin, S., Soares, M., Rice, N., Spackman, E., Hinde, S., Devlin, N., Smith, P. C., and Sculpher, M. (2015). Methods for the estimation of the National Institute for Health and Care Excellence cost-effectiveness threshold. *Health Technology Assessment*, 19(14):1. 27

Claxton, K. and Posnett, J. (1996). An economic approach to clinical trial design and research priority-setting. *Health Economics*, 5:513–524. 7, 9, 17, 19, 20, 21, 27, 56, 75, 79, 91, 99

Cookson, R., Mirelman, A. J., Griffin, S., Asaria, M., Dawkins, B., Norheim, O. F., Verguet, S., and J. Culyer, A. (2017). Using cost-effectiveness analysis to address health equity concerns. *Value in Health*, 20(2):206–212. 13

Cui, L., Zhang, L., and Yang, B. (2017). Optimal adaptive group sequential design with flexible timing of sample size determination. *Contemporary Clinical Trials*, 63:8–12. 13

Department of Health and Social Care (2020). Hospital episode statistics database. http://www.webarchives.nationalarchives.gov.uk. 75

Dimairo, M., Boote, J., Julious, S. A., Nicholl, J. P., and Todd, S. (2015). Missing steps in a staircase: a qualitative study of the perspectives of key stakeholders on the use of adaptive designs in confirmatory trials. *Trials*, 16(1):430. 79, 90

Dimairo, M., Pallmann, P., Wason, J., Todd, S., Jaki, T., Julious, S. A., Mander, A. P., Weir, C. J., Koenig, F., Walton, M. K., et al. (2020). The Adaptive designs CONSORT Extension (ACE) statement: a checklist with explanation and elaboration guideline for reporting randomised trials that use an adaptive design. *BMJ*, 369. 11, 79

Drummond, M., Sculpher, M., Claxton, K., Stoddart, G., and Torrance, G. (2015). *Methods For the Economic Evaluation of Health Care Programmes*. Oxford University Press, New York, fourth edition. 84, 88

Faria, R., Gomes, M., Epstein, D., and White, I. (2014). A guide to handling missing data in cost-effectiveness analysis conducted within randomised controlled trials. *PharmacoEconomics*, 32:1157–1170. 52

Fenwick, E., Claxton, K., and Sculpher, M. (2001). Representing uncertainty: the role of cost-effectiveness acceptability curves. *Health Economics*, 10:779–787. 97

Fenwick, E., Steuten, L., Knies, S., Ghabri, S., Basu, A., Murray, J. F., Koffijberg, H. E., Strong, M., Sanders Schmidler, G. D., and Rothery, C. (2020). 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*, 23(2):139–150. 13, 79, 83, 84, 89

Flight, L. (2020). *The use of health economics in the design and analysis of adaptive clinical trials*. PhD thesis, University of Sheffield. 87, 96

Flight, L., Arshad, F., Barnsey, R., et al. (2019). A review of clinical trials with an adaptive design and health economic analysis. *Value in Health*, 22:391–398. 13

Flight, L., Julious, S., Brennan, A., Todd, S., and Hind, D. (2020). How can health economics be used in the design and analysis of adaptive clinical trials? A qualitative analysis. *Trials*, 21(1):1–12. 90, 97

Forster, M., Brealey, S., Chick, S., Keding, A., Corbacho, B., Alban, A., Pertile, P., and Rangan, A. (2021). Cost-effective clinical trial design: Application of a Bayesian sequential model to the ProFHER pragmatic trial. *Clinical Trials*, 0(0):17407745211032909. PMID: 34407641. 7, 8, 10, 11, 15, 17, 18, 69, 71, 76, 78, 87, 88, 92, 93

Hampson, L. and Jennison, C. (2013). Group sequential tests for delayed responses. *Journal of the Royal Statistical Society, Series B*, 75:3–54. 12, 13, 80

Handoll, H., Brealey, S., Rangan, A., et al. (2015). The ProFHER (PROximal Fracture of the Humerus: Evaluation by Randomisation) trial - a pragmatic multicentre randomised controlled trial evaluating the clinical effectiveness and cost-effectivesness of surgical compared with non-surgical treatment for proximal fracture of the humerus in adults. *Health Technology Assessment*, 19:1–280. 11, 14, 85, 86, 93

Jaki, T. (2013). Uptake of novel statistical methods for early-phase clinical studies in the UK public sector. *Clinical Trials*, 10:344–346. 79

Kingsbury, S. R., Tharmanathan, P., Adamson, J., et al. (2013). Hydroxychloroquine effectiveness in reducing symptoms of hand osteoarthritis (HERO): study protocol for a randomized controlled trial. *Trials*, 14(64). 51, 75

Kingsbury, S. R., Tharmanathan, P., Keding, A., et al. (2018). Hydroxychloroquine effectiveness in reducing symptoms of hand osteoarthritis: a randomized trial. *Annals of Internal Medicine*, 168:385–395. 1, 7, 9, 10, 51, 52, 54, 55, 71, 75, 77

Latimer, N. R., Bhadhuri, A., Alshreef, A., Palmer, R., Cross, E., Dimairo, M., et al. (2020). Self-managed, computerised word finding therapy as an add-on to usual care for chronic aphasia post-stroke: An economic evaluation. *Clinical Rehabilitation*, 0(0):0269215520975348. PMID: 33233972. 7, 9, 10, 23, 25, 26, 93, 96

Latimer, N. R., Dixon, S., and Palmer, R. (2013). Cost-utility of self-managed computer therapy for people with aphasia. *International Journal of Technology Assessment in Health Care*, 29(4):402. 24, 28, 29

Morgan, O. J., Hillstrom, H. J., Ellis, S. J., Golightly, Y. M., Russell, R., Hannan, M. T., et al. (2019). Osteoarthritis in England: Incidence Trends From National Health Service Hospital Episode Statistics. *ACR Open Rheumatology*, 1(8):493–498. 75, 77

National Institute for Health Research (2014). Annual Efficient Studies funding calls for CTU projects. Accessed 4 September 2020 https://www.nihr.ac.uk/documents/ad-hoc-funding-calls-for-ctu-projects/20141. 11

National Institute for Health Research (2020). Delivering Complex and Innovative Trials. Accessed 4 September 2020 https://www.nihr.ac.uk/partners-and-industry/industry/run-your-study-in-the-nhs/complex-innovative-trials.htm. 11

NICE (2013). Guide to the methods of technology appraisal. 16, 25, 28, 77, 88

NICE (2014). Interim methods guide for developing service guidance: Modeling and health economics considerations. UK National Institute for Health and Care Excellence, Accessed March 20, 2018, https://www.nice.org.uk/process/pmg8/chapter/modelling-and-health-economics-considerations. 84

O'Brien, P. C. and Fleming, T. R. (1979). A multiple testing procedure for clinical trials. *Biometrics*, 35:549–556. 97

Pallmann, P. et al. (2018). Adaptive designs in clinical trials: why use them, and how to run and report them. *BMC Medicine*, 16(29). Accessed May 7, 2018, https://doi.org/10.1186/s12916-018-1017-7. 13, 79

Palmer, R., Cooper, C., Enderby, P., Brady, M., Julious, S., Bowen, A., and Latimer, N. (2015). Clinical and cost effectiveness of computer treatment for aphasia post stroke (Big CACTUS): study protocol for a randomised controlled trial. *Trials*, 16(1):1–12. 48, 93

Palmer, R., Dimairo, M., Cooper, C., Enderby, P., Brady, M., Bowen, A., Latimer, N., Julious, S., Cross, E., Alshreef, A., et al. (2019). 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*, 18(9):821–833. 7, 9, 10, 30

Palmer, R., Dimairo, M., Latimer, N., Cross, E., Brady, M., Enderby, P., Bowen, A., Julious, S., Harrison, M., Alshreef, A., et al. (2020). Computerised speech and language therapy or attention control added to usual care for people with long-term post-stroke aphasia: the Big CACTUS three-arm RCT. *Health Technology Assessment*, 24(19):1. 7, 9, 10, 23, 25, 28, 30, 37, 48

Palmer, R., Enderby, P., Cooper, C., Latimer, N., Julious, S., Paterson, G., Dimairo, M., Dixon, S., Mortley, J., Hilton, R., et al. (2012). Computer therapy compared with usual care for people with long-standing aphasia poststroke: a pilot randomized controlled trial. *Stroke*, 43(7):1904–1911. 24, 28

PANDA (2020). Practical Adaptive & Novel Designs and Analysis (PANDA) toolkit. Available online at https://panda.shef.ac.uk/ (accessed 5 March 2021). 11

Philips, Z., Claxton, K., and Palmer, S. (2008). The half-life of truth: what are appropriate time horizons for research decisions? *Medical Decision Making*, 28:287–299. 75

Pocock, S. J. (1982). Interim analysis for randomized clinical trials: the group sequential approach. *Biometrics*, 38:153–162. 97

Puterman, M. L. (1994). *Markov Decision Processes*. Wiley, New York, First edition. 79

Raiffa, H. and Schlaifer, R. (1961). *Applied Statistical Decision Theory*. Harvard University Graduate School of Business Administration, Boston, first edition. 27

Rangan, A., Handoll, H., Brealey, S., et al. (2015). Surgical vs nonsurgical treatment of adults with displaced fratures of the proximal humerus: the PROFHER randomized clinical trial. *Journal of the American Medical Association*, 313(10):1037–1047. 11, 14

Ronaldson, S., Keding, A., Tharmanathan, P., Arundel, C., Kingsbury, S., Conaghan, P., and Torgerson, D. (2021). Cost-effectiveness of hydroxychloroquine versus placebo for hand

osteoarthritis: economic evaluation of the HERO trial [version 1; peer review: awaiting peer review] . *F1000Research*, 10(821). 1, 10, 51, 52, 56, 71, 73, 75, 93

Rothery, C., Strong, M., Koffijberg, H., Basu, A., Ghabri, S., Knies, S., et al. (2020). Value of Information Analytical Methods: Report 2 of the ISPOR Value of Information Analysis Emerging Good Practices Task Force. *Value in Health*, 23(3):277–286. 13, 75, 79, 84, 95

Royston, P. (2004). Multiple imputation of missing values. *Stata Journal*, 4(3):227–241. 51, 52

Royston, P. (2005). Multiple imputation of missing values: update. *Stata Journal*, 5(2):188–201. 51, 52

Ryzhov, I. O., Frazier, P. I., and Powell, W. B. (2012). The knowledge gradient algorithm for a general class of online learning problems. *Operations Research*, 60(1):180–195. 81

Spiegelhalter, D., Freedman, L., and Parmer, M. (1994a). Bayesian approaches to randomised trials. *Journal of the Royal Statistical Society, Series A*, 157:357–416. 16

Spiegelhalter, D. J., Freedman, L. S., and Parmar, M. K. (1994b). Bayesian approaches to randomized trials. *Journal of the Royal Statistical Society: Series A (Statistics in Society)*, 157(3):357–387. 49

StataCorp (2013). Stata: Release 13. Statistical Software, College Station, TX: StataCorp LP. www.stata.com. 51, 52

Sutton, L., Julious, S. A., and Goodacre, S. W. (2012). Influence of adaptive designs on unnecessary patient recruitment: reanalysis of the RATPAC trial. *Annals of Emergency Medicine*, 60:442–448.e1. 13

Swain, S., Sarmanova, A., Mallen, C., Kuo, C., Coupland, C., and Doherty, M. (2020). Trends in incidence and prevalence of osteoarthritis in the United Kingdom: findings from the Clinical Practice Research Datalink (CPRD). *Osteoarthritis and Cartilage*, 28:792–801. 75

The Math Works, Inc. (2019). MATLAB. Version 2019a. The Math Works, Inc., 2020. Computer Software. www.mathworks.com. 51

Villar, S. S. and Rosenberger, W. F. (2018). Covariate-adjusted response-adaptive randomization for multi-arm clinical trials using a modified forward looking Gittins index rule. *Biometrics*, 74(1):49–57. 81

Willan, A. and Kowgier, M. (2008). Determining optimal sample sizes for multi-stage randomized clinical trials using value of information methods. *Clinical Trials*, 5:289–300. 27

Willan, A. R. and Pinto, E. M. (2005). The value of information and optimal clinical trial design. *Stat. Med.*, 24:1791–1806. 50

Wu, J. and Frazier, P. (2016). The parallel knowledge gradient method for batch bayesian optimization. In *Advances in Neural Information Processing Systems*. 81

Yu, D., Jordan, K., Bedson, J., Englund, M., F. Blyth, A. T., et al. (2017). Population trends in the incidence and initial management of osteoarthritis: age-period-cohort analysis of the Clinical Practice Research Datalink, 1992–2013. *Rheumatology*, 56:1902–1917. 75

Yu, D., Peat, G., Bedson, J., and Jordan, K. (2015). Annual consultation incidence of osteoarthritis estimated from population-based health care data in England. *Rheumatology*, 54:2015–2060. 75

Zagoraiou, M. (2017). Choosing a covariate-adaptive randomization procedure in practice. *Journal of Biopharmaceutical Statistics*, 27(5):845–857. PMID: 28166466. 80