Efficiency in Mental Health Randomised Trials

Yifeng Liu

University College London

Thesis submitted for the degree of Doctor of Philosophy

Acknowledgements

There is an endless list of people to whom I am grateful for that have given me an amaranthine of help along my journey over the past 48 months – I could not have longed nor wished for an experience or environment that could have provided better support during the course of my study.

First and foremost, my sincere gratitude goes to my supervisors, Nick Freemantle, Rachael Hunter and Caroline Clarke, who have been avidly supportive throughout. I appreciate both their willingness and enthusiasm to share their knowledge and experience, as well as their patience and efforts in guiding me in the right direction.

Besides my supervisors, I am also grateful to whom I have worked with over the past 4 years, Drs. Emma Pencheon, Julie Fricke and Joanna Moncrieff for helping me with the systematic review, April Slee for introducing me to an array of statistical analysis methods, and Dr. Anna Heath, whom helped me enormously with the Value of Information chapter, as well as Drs. Tra My Pham and Tao Ding for proofreading some of the chapters.

I would also like to thank my fellow PhD students – Andrea, Fabian, Dan, Hamad, Emma, Paulina, Deepani, Sonia, Manuj, JC and Shoba; as well as the colleagues at Health Economics Analysis and Research methods Team. It would not have been possible without their support and help throughout the 4 years and I have learnt many things from each of them.

Certainly, this would not have happened without the wholehearted support mentally and financially, from my family, to whom I owe a debt of gratitude. Making up my mind to take this journey has been somewhat the most audacious decision I have made and it was certainly not a plain sailing. Thanks to my mother who has always been there for me during this period; rain or shine.

I, Yifeng Liu, confirm that the work presented in this thesis is my own. Where information has been derived from other sources, I confirm that this has been indicated in the thesis.

I created the draft of the thesis, with comments from my supervisory team including Dr. Caroline S Clarke, Rachael M Hunter and Prof. Nick Freemantle, who were involved in the majority of the thesis unless specified otherwise.

Chapter 2 contains collaboration with Dr. Emma Pencheon, who screened the titles and abstracts, and extracted the data, as well as Dr. Joanna Moncrieff, who commented on the draft of the published systematic review, which makes up most of the chapter. In addition, the search update screening was in collaboration with Dr. Julie Fricke, whom I would also like to thank.

Chapter 4 contains collaboration with Dr. Joanna Moncrieff and the RADAR operation team, who gave some very helpful feedback onto the design of the decision model.

Chapter 5 contains collaboration with Anna Heath, who very kindly offered some advice on the calculation of the EVSI and EVPPI quantities. I also took some suggestions from Prof. Nicky Welton (University of Bristol), who reviewed and critiqued the chapter as a conference paper in the 6th EuHEA PhD student - supervisor and early career researcher conference in September 2019.

Abstract

Randomised controlled trials (RCTs) are perceived as the gold standard for assessing therapeutic interventions. Modern randomised trials often involve extensive collaboration of different areas of expertise and can result in high cost in design and conduct.

This thesis discusses different aspects in which RCTs in mental health could be improved with better efficiency, in line with a case of research into antipsychotic discontinuation and reduction intervention, which is an RCT funded by the National Institute of Health Research (NIHR).

To begin with, the thesis investigates the effectiveness and cost-effectiveness of different strategies which may improve recruitment into and retention in randomised trials in mental health in a systematic review. There is not enough evidence on “Study within a Trial” to compare the effectiveness of recruitment and retention intervention in mental health trials. Further research in this area could highlight the possibilities of using strategies that are most suitable to trial design and population characteristics.

Secondly, this thesis takes a “real world evidence” perspective to inform RCTs using a decision analytic model which compares the cost-effectiveness of antipsychotic reduction and discontinuation strategy and the standard maintenance strategy, under the current £20,000 – £30,000/QALY willingness to pay threshold.

Thirdly, following the decision analytic model, this thesis discusses to what extent decision uncertainty under certain willingness to pay threshold could be reduced by further research, and the expected value of reducing uncertainties on certain parameters in the decision model, using the Value of Information analysis framework. It could help trialists with the design of future RCTs in line with the standard sample size calculation to determine the economic benefit of the RCT. Decision then can be made by the worthiness of conducting an RCT to reduce the uncertainty regarding the adoption of new intervention.

Impact statement

The potential impacts of the work presented in thesis are in two aspects. Firstly, I have summarised and discussed the strategies to improve recruitment and retention into mental health randomised trials. Highlighting the current available strategies, as well as the evidence gaps in the literature, could help trialists and trial managers identify the appropriate strategies in planning and conducting randomised trials for relevant populations. Subsequent recruitment and retention studies may also be informative in enlarging the current literature. Also, through value of information analysis using previous evidence may be suggestive in future research design, by considering the economic benefits of research programmes and determining the number of participants which results in maximum economic benefit. This could be used in accordance with the traditional sample size calculation during the design stage of a randomised trial, with the question of what the scientifically reliable and economically efficient experimental design could be.

On the other hand, this thesis discusses the efficiency in mental health randomised trials based on a case of antipsychotic reduction and discontinuation strategy, which has clinical impact on patients with severe mental illnesses. Evidence for and against this intervention could inform the current clinical practice in psychiatry, either advocating the intervention or suggesting its use in combination with other early preventative interventions.

Table of Contents

List of Tables

List of Figures

Glossary

AE – adverse event

AMI – amisulpride

ANCOVA – analysis of covariance

ARI – aripiprazole

AUD – Australian dollar

BCEAweb – Bayesian Cost-effectiveness analysis web

BD – bis die, twice daily

BMI – body mass index

BNF – British National Formulary

BPRS – Brief Psychiatric Rating Scale

CATIE – Clinical Antipsychotic Trials for Intervention Effectiveness

CBT – cognitive behavioural therapy

CEA – cost-effectiveness analysis

CEAC – cost effectiveness acceptability curve

CHD – coronary heart disease

CI – confidence interval

CJS – criminal justice system

COI – cost of illness

CRT – clustered randomised trial

CVA – cerebrovascular accident

CVD – cardiovascular disease

DALY – Disability-adjusted life year

DES – discrete event simulation

DSM-5 – Diagnostic and Statistical Manual of Mental Disorders (5th Edition)

DSMC – data safety and monitoring committee

EDE – eating disorder examination

EMR – electronic medical record

ENGS – expected net gain of sampling

EPS – extrapyramidal symptom

EVPI – expected value of perfect information

EVPPI – expected value of partial perfect information

EVSI – expected value of sample information

FDA – Food and Drug Administration

FGA – first generation antipsychotic

FWER – family-wise error rate

GABA – gamma-aminobutyric acid

GAM – generalised additive model

GBP – British pound sterling

GP – general practitioner; Gaussian Process (only in Chapter 5)

HAL – haloperidol

ICC – intra-cluster correlation coefficient

ICD-10 – International Statistical Classification of Diseases and Related Health Problems (10^{th} Version)

ICECAP – ICEpop CAPability measure for Adults

ICER – Incremental cost effectiveness ratio

LAI – long-acting injectable

LD – lethal dose

MAR – missing at random

MCAR – missing completely at random

MI – myocardial infarction

MMRM – mixed model for repeated measures

MNAR – missing not at random

MTD – maximum tolerable dose

NHS – National Health Service

NICE – National Institute for Health and Care Excellence

NIHR – National Institute of Health Research

NMB – net monetary benefit

OD – omni die, every day

OLA – olanzapine

PAL – paliperidone

PANSS – Positive and Negative Syndrome Scale

PCT – Primary Care Trust

PPI – patient and public involvement

PPIR – patient and public involvement research

PSA – probabilistic sensitivity analysis

PSSRU – Personal Social Services Research Unit

QALY – quality adjusted life year

RADAR – Research into Antipsychotic Discontinuation and Reduction

RCT – randomised controlled trial

RIS – risperidone

SAVI – Sheffield Accelerated Value of Information

SD – standard deviation

SE – standard error

SFS – social functioning score

SG – standard gamble

SGA – second generation antipsychotic

SMI – severe mental illnesses

SOP – standard operating procedure

SPDE-INLA – Integrated Nested Laplace Approximation with Stochastic Partial Differential Equation

SWAT – study within a trial

TAU – treatment as usual

TD – ter die, three times daily

THIN – The health improvement network

TIA – transient ischaemic attack

TTO – time trade off

USD – United States Dollar

VoI – Value of Information

WTP – willingness to pay

ZOT – zotepine

Outline of the Thesis

This thesis will discuss several aspects of the efficiency of randomised controlled trial (RCT), with a particular focus on mental health and the case of Research into Antipsychotic Discontinuation and Reduction (RADAR).

The aim is to investigate different ways in which mental health trials may be designed and conducted with more efficiency, by looking at recruitment and retention into mental health RCTs, as well as how real world evidence may be used to inform study design to improve efficiency. The objectives are to use previous evidence into the cost and effectiveness of strategies to improve recruitment and retention; then to see how simulation methods using real world evidence may inform clinical trial design, to achieve efficiency in planning and conduct. In Chapter 1 I will introduce the background and rationale of my thesis, including RCT, economic evaluation, mental health, antipsychotics and the RADAR trial. In Chapter 2 I will review the existing strategies to improve recruitment and retention in mental health RCTs. In Chapter 3, I will briefly review and describe the design features of decision models for patients with schizophrenia, followed by Chapter 4, in which I will describe and discuss a decision analytic model for the cost-effectiveness of the antipsychotic reduction and discontinuation strategy, which will be investigated in the RADAR trial. Chapter 5 will discuss the value of conducting future research for the intervention, in order to reduce the decision uncertainty under the current cost-effectiveness threshold, by a series of value of information (VoI) analyses. Chapter 6 will discuss the main results of the thesis and their implications.

1 Chapter 1 – Introduction

In this chapter I will describe the context and background from which my studies in the following chapters are drawn. I will describe some of the fundamentals of clinical trials, trial-based and non-trial-based economic evaluation, and the decision process for recommendation of reimbursement for therapeutic interventions in England. I will also describe and discuss the therapeutic background on which my thesis is based. This includes the current pharmacological treatments for patients with schizophrenia, primarily antipsychotics, and the pitfall of such interventions. Finally, I will describe a research project which seeks to tackle the problem of antipsychotic treatment, which will act as the case study throughout the thesis. The aim of this chapter is to introduce the theoretical background and the current research environment in related area, on which the rest of the chapters are based.

1.1 Schizophrenia & Antipsychotics

1.1.1 Mental illness

Mental illnesses include a wide range of health conditions with different symptoms. In general, mental illnesses are clinically manifested by some degrees of combination of abnormal emotions, thoughts and behaviours. Some conditions can also affect an individual's social functions. (1) Mental illnesses mainly include:

- Anxiety disorders, including panic disorder, obsessive-compulsive disorder, post-traumatic stress disorder & phobias;
- Bipolar disorder;

- Depression;
- Mood disorders;
- Personality disorders;
- Psychotic disorders, including schizophrenia.

The aetiology behind most of these conditions is not thoroughly understood. Most mental illnesses are a combination of biological, genetic, psychological and environmental factors, which can trigger the illness at some point. (2) A number of current pharmacological treatments try to ameliorate the brain chemistry, mainly the neurotransmitters in some particular brain regions. Also, heredity is considered another factor affecting some mental illness such as schizophrenia and depression. (3) It is suggested that susceptibility may be passed down over generations within family. However, in most cases there are many genes suggested to be associated with a mental illness to the extent that onset of a condition may result from a particular gene, interaction with different genes associated with mental illness, or even some non-inheritable factors, such as environmental and experiential triggers. Besides, cerebral infections, defects or injury, substance abuse and other factors are also known to be associated with mental illnesses.(4)

1.1.2 Schizophrenia

Of the mental illnesses listed in either the International Statistical Classification (10th edition, ICD-10) or the Diagnosis and Statistical Manual of Mental Disorders (5th edition, DSM-5), schizophrenia is considered a severe condition. Manifested by abnormal social behaviour and problems with perceiving the real

world, schizophrenia often starts during one's early 20s. (5) Symptoms of schizophrenia come into 3 categories:

- “Positive” symptoms, psychotic behaviours rarely observed in healthy individuals, such as hallucinations, delusions, thought disorders; (6)
- “Negative” symptoms, that cause diminished emotions, e.g., little emotion, difficulty in experiencing pleasure, reduce verbal communications; (6)
- Cognitive dysfunction, often taken as an important feature of schizophrenia, in most cases represented as diminished working memory, difficulty in executive functioning. (7)

The aetiology of schizophrenia is a complex interaction of biological, psychological and environmental factors, which is yet to be understood. (8) There has been a number of theories that try to explain how functional changes in brain activities are associated with schizophrenia. Current schizophrenia research still remains under the physiology model, and despite efforts from different disciplines, which have achieved some progress, the whole perspective is still fragmented.

The two existing widely accepted theories that attempt to explain the pathology of schizophrenia are the dopamine and glutamate hypotheses. The dopamine hypothesis was first postulated when phenothiazines, a kind of sedative, were found to reduce psychotic symptoms effectively apparently due to their blockage of dopamine function. This influential hypothesis suggested that the positive symptoms of schizophrenia are caused by a malfunction relating to

dopamine pathways, or more specifically, the over-activation of D2 receptors, which is the pharmacological rationale common to all antipsychotics. (9) In fact, correlation is suggested between the pharmacological efficacy of antipsychotics and their affinity to dopamine D2 receptors. (10)

The glutamate hypothesis was proposed after cases in which hypo-function of glutamate receptors was identified in previously diagnosed schizophrenia patients. While glutamate is responsible for activating neurons as well as other brain cells, nearly 60% of neurons use it as neurotransmitter. Schizophrenia symptoms are relevant to multiple areas which are connected by a glutamate regulated circuit. Irregular glutamate activity may produce symptoms, along with its interactions with dopamine and gamma-aminobutyric acid (GABA). (11) Although the glutamate hypothesis was proposed during 1980s and was considered more proximal to the real aetiology, it does not contradict with the earlier dopamine hypothesis. It is likely that both glutamatergic and dopaminergic irregularities are implicated in schizophrenia in terms of dysfunction of the production of neurotransmitters.

1.1.3 Antipsychotics

The current pharmacology treatment strategy of schizophrenia is antipsychotics, which are also used for mental illness such as bipolar disorder, dementia, and major depressive disorder. Since they were first introduced in 1950s, two generations of antipsychotics have been developed for the treatment of mental illness. First generation antipsychotics (FGAs, also known as typical antipsychotics) were developed in 1950s, mainly dopamine antagonists. Commonly prescribed FGAs include: chlorpromazine, perphenazine,

mesoridazine, fluphenazine, thiothixene, molindone, loxapine, trifluoperazine and haloperidol. In recent years, some new medications have been introduced for schizophrenia treatment. These second-generation antipsychotics (SGA) block dopamine to a relatively moderate degree. Aripiprazole, clozapine, ziprasidone, risperidone, quetiapine and olanzapine are the most prescribed SGAs. (12)

The mechanism of most antipsychotics is to block dopamine receptors (mainly D2 receptor), while some atypical antipsychotics (SGAs) also function on other different neurotransmitters such as serotonin. There is no clear evidence that suggests superiority of SGAs in schizophrenia treatment. Some antipsychotics may appear to be more effective but have more side effects. (13) Although antipsychotics can effectively reduce positive symptoms of schizophrenia, side effects of these medications are also significant. FGAs share a group of similar side effects including stiffness and shakiness, drowsiness, or dry mouth. Tardive dyskinesia, a side effect that presents symptom of stiff and jerky movements of face and body, can occur in some patients who are under FGA treatment, but are less seen in cases with patients under SGAs treatment. SGAs may have some common side effects with FGAs, but they are less likely to cause shakiness and stiffness or tardive dyskinesia. Other side effects of SGAs include weight gain, sleepiness, risk of developing diabetes and cardiovascular disease (CVD). (14) Although these effects seem to have been reduced in SGAs, they still entail a great level of risk in treatment. In an 18-month double-blinded randomised trial comparing the effectiveness several FGAs and SGAs, the majority of patients discontinued their assigned treatment of either SGAs or FGAs owing to inefficacy or severe side effects. (15)

1.1.4 Antipsychotics' dilemma

Antipsychotic treatment has long been the standard of care in patients with schizophrenia. However, the uncertainty around the basic aetiology of most of mental illnesses makes their treatment less likely to succeed. The current long-term antipsychotic treatment strategy only aims to control some of the positive symptoms of schizophrenia and reduce the risk of symptom exacerbation. Meanwhile, emerging evidence also suggest that the long-term antipsychotic treatment may have negative impacts on social functioning and other health conditions. (16) There have been existing evidence showing that long-term antipsychotic treatment may increase the risk of metabolic comorbidity, and antipsychotic use is associated with increased risk of CVD. (17) Alongside the adverse effects of antipsychotics such as extrapyramidal symptoms (EPS) and weight gain, antipsychotic treatment renders a choice of stabilising patients with schizophrenia, but in exchange of sacrificing quality of life.

On the other hand, it is often posited that antipsychotic withdrawal may increase the risk of relapse, although the existing evidence base has no agreed definition on relapse. (18) The effects such as agitation and hostility, may to an extent be exaggerated by the physiological symptoms alongside the antipsychotic withdrawal. It still remains conflicted in the literature as to whether gradually reducing antipsychotic dose towards withdrawal can help reduce the risk of relapse. (19, 20)

It seems, in the antipsychotic withdrawal case, promising to investigate the efficacy and cost effectiveness of such an intervention.

1.2 What is a clinical trial?

A clinical trial is a prospective experiment to assess the efficacy of therapeutic interventions in medical research. As a practice often carried out in the pharmaceutical industry, the clinical trial framework has been also widely adopted amongst public health researchers. It ensures scientific rigour throughout the development of novel therapeutic interventions, particularly pharmacologically.

1.3 Clinical trial design

During the development of a novel therapeutic drug, 4 stages of clinical trials are often involved. These are often classified as Phase I, II, III and IV trials. (21) These terms have been so widely used that trials for other interventions often refer to the same terminology. (21) These different stages follow from an exploratory to confirmatory nature, where increasing numbers of research participants are involved.

1.3.1 Phase I trials

The purpose of a Phase I trial is to establish the basic dose and safety profile of a novel drug. To explore the tolerability by establishing the maximum tolerable dose (MTD), Phase I trials often introduce a series of dose steps in order to observe the optimal dose starting from the result of pre-clinical data, e.g. the dose which, given all at once, causes the death of 10% of a group of test model animals, lethal dose LD_{10}. Consequently, dose escalation using the modified "Fibonacci" sequence may be used to find the MTD. A "3+3" design has historically often used in determining the MTD in combination with the dose

escalation sequence, although has now partly been replaced by more flexible and efficient Bayesian Continuous Reassessment Model designs. (22)

1.3.2 Phase II trials

Phase II trials aim to establish the safety and efficacy of a novel intervention. More information is collected at this stage on a larger cohort, in order to investigate the feasibility of the intervention, for instance considering adverse effects, safety or cost. Based on the results, it can be advised whether a consequent large-scale comparative study is worthwhile. Randomisation is normally employed at Phase II trials. Although there are no formal defined sub-categories for Phase II trials, generally Phase IIa are 'proof of concept' studies to demonstrate clinical efficacy or biological activity, and IIb are dose finding studies that determine the optimum dose for efficacy and minimum side-effects. In some cases, for instance, in a rare condition where patient population is limited, Phases IIa and IIb may be combined, and on occasions phase II and III are rolled together.

1.3.3 Phase III trials

Phase III trials are larger trials that are confirmatory of the efficacy of the new treatment relative to the control. It is most commonly compared against a concurrent group, being either (most cases for regulatory approval) placebo or in some cases standard of care. Randomisation is the standard practice in this stage of the trial and hence the term RCT is often used interchangeably. Phase III trials generally require many more participants in order to establish the precision of the treatment effects and ensure the power of the study. It is phase

III trials which provide the evidence require to introduce treatments into regular practice.

1.3.4 Phase IV trials

The need for a larger scale, and longer period expanded safety Phase IV trial may be found in chronic conditions or to study the potential occurrence of rare but serious side effect. As pivotal Phase III trials usually only enrol a few hundreds of patients and last for a relatively shorter term, Phase IV trials, often useful in post-marketing surveillance, can help investigating the long-term side effects for the intervention.

1.4 Randomised Controlled Trial

By randomising patients into different treatment groups, RCTs can provide a sound basis for treatment effect estimation. It directly deals with bias from confounders, and attributes both observed and unobserved baseline differences between groups to chance, the effect of which could be estimated by statistics. (21) Anecdotal success alone cannot be the real guide and can be attributed to factors such as “regression to the mean”, selection bias, or non-specific treatment effect. (23)

1.4.1 Randomisation

Randomisation is effective in reducing bias in treatment selection so that investigators cannot assign patients with better prognosis to a particular trial arm. As selection bias can be as influential as many treatment effects in terms of clinical outcomes, randomisation provides benefit in preventing it. Also, randomisation prevents confounding, providing an instrument for direct and credible statistical adjustments. In a classic parallel group RCT, participants are

randomly allocated into either the intervention of interest, or the existing standard intervention (also known as control, or placebo). The aim of randomisation is to ensure that the two groups of participants have similar characteristics so that the outcomes could only be attributed to the treatment effect, or chance. (24)

In practice, there are several methods to carry out randomisation. The most elementary form is to allocate patients by tables of computer-generated random numbers. Then treatments are allocated by telephone or online allocation tools such as “Sealed Envelope” (https://www.sealedenvelope.com/). Simple randomisation may often causes a size difference between trial arms, which may have substantial effects in small trials. To prevent this, other methods such as block randomisation can guarantee similar numbers of participants between groups. (25) For instance, for interventions A and B, a block of 4 can be chosen for randomisation. Sequences of AABB, ABAB, ABBA, BBAA, BABA and BAAB are randomly chosen for each block of participants. Block size can also vary, creating a complex but balanced randomisation scheme. (26) This will help to achieve a close balance in group numbers. Furthermore, differences between the groups at baseline may still exist by using these two methods, due to the possible chance imbalance in prognostic factors. Stratification and minimisation are effective ways to reduce the imbalance in prognostic factors, at least for those where the relationship with outcome are known such as markers of increased severity, and provide efficient, unbiased estimates of the treatment effect. (27)

1.4.2 RCTs have complex framework of planning, conducting and monitoring

RCTs need careful planning before launching. The determination of study design type and sample size is a decision process based on a series of factors, such as the intervention and patient population characteristics, clinical settings and choices of clinical outcomes. Also, a series of Standard Operating Procedures (SOPs) ensures the quality and scientific integrity of RCTs are not to be compromised when one is conducting an RCT, because in clinical trials, which involve humans as experiment participants, patient safety is another essential component. Monitoring the safety of enrolled patients promptly is not only a matter of clinical importance, but also an ethical consideration. The data safety and monitoring committee (DSMC), which can undertake interim analyses during the progress of the trial, is to ensure that any further hazardous effect of the investigational intervention can be prevented by terminating the trial before more patients in the trial get affected.

1.4.3 Sample size

Amongst the various designs for RCTs, the two-group parallel design is the most fundamental, and the most prevalent design. By randomising patients into 2 parallel groups, one with the investigational intervention and the other with a control, the difference, or lack of difference in the outcome can only be attributed to the treatment effect, or simply chance. To determine the sample size of a two-group parallel RCT, a set of parameters on the minimum treatment effect (mean Δ and standard deviation σ, often hypothesised and obtained from the literature), type I error rate α (the probability of rejecting the null hypothesis given it is true), and II error rate β (the probability of rejecting the null hypothesis when it is false), are required. Conventionally, a two-sided α level of 0.05 and

80% or 90% power (1-β) are adopted. Then the sample size per arm for a 1:1 parallel group RCT can be determined by the following formula

$$n = \frac{2\,(z_{1-\alpha}+z_{1-\beta})^2\,\sigma^2}{\Delta^2},(28)$$

where z is the Z value for the corresponding error rates, for instance $z_{1-\alpha} = 1.96$ for two sided α = 0.05 and $z_{1-\beta} = 1.282$ for β = 0.1.

According to the clinical question (hypothesis), sometimes an RCT may not have to show advantages of the treatment effect for the investigational intervention, hence an equivalence design, or a non-inferiority design may be appropriate. Non-inferiority design has become prevalent in oncology and diabetes trials. (29) Non-inferiority hypothesises that the investigational intervention is no worse than the control by “a small amount” – the non-inferiority margin. (30)

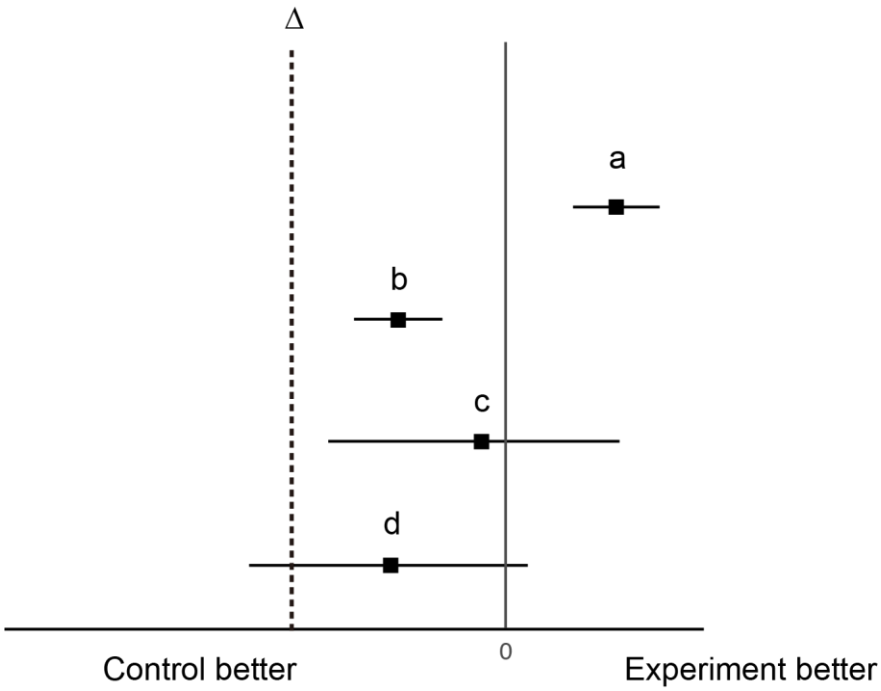

Figure 1-1 Non-inferiority design confidence intervals

Figure 1-1 illustrates the treatment effect estimation in relation to the non-inferiority margin, Δ, noted by the dashed line. In order to show superiority of the experiment intervention, the 95% confidence interval (CI) of the treatment effect should fall entirely to the right of the solid black line, which indicates no difference between two treatments, as shown by estimate a. If the 95% CI crosses or falls to the left of the non-inferiority margin, then the hypothesis of the experiment being worse than the control cannot be rejected, as shown by estimate d. If the 95% CI falls to the right of the non-inferiority margin, then the hypothesis of experiment being worse than the control can be rejected. (estimates b and c).

To determine the sample size for a 1:1 parallel group non-inferiority trial, the number of participants per arm is given by:

$$n = \frac{2\,\sigma^2(z_{1-\alpha}+z_{1-\beta})^2}{(\Delta - \text{NI margin})^2}$$.(31, 32)

When evaluating the effects of non-drug interventions, for instance, policy and service delivery interventions, a clustered randomised trial (CRT) design is sometimes used.(33) Instead of randomising each individual participant, the CRT design randomises each cluster unit, e.g. each ward in the hospital sites, or each general practice unit, and all the participants within the cluster receive the same intervention. This is mostly when it is impractical to randomise individual participants such as where there are group based behaviours. (34) However it will require a larger sample size compared to individual randomisation due to the intra-cluster correlation (ICC) as patients are not completely independent within clusters. This needs to be considered when calculating the sample size, often by multiplying the sample size which would be required under individual randomisation by the inflation factor:

$$1 + (n - 1) \times ICC,$$

where n is the average cluster size.(28)

1.4.4 Challenges and hurdles often seen in conducting randomised trials

Although being the gold standard of clinical research, RCTs are often faced with some hurdles and challenges. These can be operational, analytical, managerial or ethical, as it often involves the participation of numerous stakeholders with different perspectives. Recruitment is a common issue in RCT. Delayed recruitment can give rise to a series of issues, such as additional costs or extension of the study period. Inadequate or ineffective recruitment may often result in reduced power of the study, or even premature termination of a trial, hence the failure in answering an important clinical question. (1) This is more common in publicly funded RCTs. Only slightly over 50% of the publicly funded RCTs managed to reach their target sample size.(35) There have been several strides to improve the recruitment into RCTs but little evidence was found as to the efficacy of ways to improve recruitment. (36)

Similar to recruitment, retention can often be a major issue when conducting an RCT. As RCTs often follow patients up for a period of time and collecting longitudinal data, there are very often loss to follow-up issues and more seriously, patients dropping out from the RCTs results in missing data. Loss to follow-up and patient dropouts can result in reduced study power, hindering the ability to detect potential differences between trial arms if they were to exist, as well as undermining the internal validity of a randomised trial. More importantly, it could introduce systematic bias as the protection of randomisation becomes lost. (37) Although an increasing amount of research has contributed to dealing

with missing data in clinical trials, the risk of bias due to missing data cannot be fully avoided through the application of statistical techniques for missingness, such as multiple imputation, as these techniques require additional assumptions which may not be valid. Low retention rates result in missing data that may lead to bias in the estimation of treatment effect. Missing completely at random (MCAR) occurs on very rare occasions and can lead to wider CIs of effect estimation. (38) Complete case analysis is still considered sensible under MCAR, but may lead to biased estimates when data are missing at random (MAR, unless conditional on fully observed baseline covariates), in which case a few other analytic choices are available, such as Analysis of Covariance (ANCOVA), or mixed model for repeated measures (MMRM).(39) It is challenging to estimate the effect when data are missing not at random (MNAR) and sensitivity analysis is crucial in assessing the impact of potential departures from the MAR assumption. Nevertheless, these analytic techniques can be helpful in pointing out directions of departure for sensitivity analysis. It has been suggested that less than 5% loss to follow-up may lead to an unimportant level of bias, while 20% or greater loss to follow-up poses a substantial threat to a trial's internal validity. (37) Some modern trials aim to reduce this risk by increasing the sample size by 20%, which addresses precision but not internal validity, and poses a further challenge to recruitment.

1.4.5 Monitoring an RCT – when and why a trial needs to stop

Clinical trials are designed to answer a clinical question with the appropriate components. The standard practice of trial design is to determine these components *a priori* based on the existing clinical evidence, for instance, the choice of clinical outcomes (endpoints), the choice of control intervention (active

or placebo), whether to use a parallel group or a cross-over design, the number of arms, and data to be collected. Information relating to these is often included in a pre-specified trial protocol. A target sample size is often determined before the recruitment starts, but patients do not always enter the trial simultaneously. Rather, recruitment often lasts for a long period and patients enter the trial sequentially. Cases in which patient A just enters the trial while patient B has already finished all the follow-up assessments (if not dropping out) often occur, which creates an opportunity for assessing the results on those who finish earlier to have a glimpse of the trial results.

However, one must be aware that sample size depends on the maximum acceptable values for the type I and II error rates. One may observe “random high” in the outcome for the first few patients entering the trial thus incorrectly reject the null hypothesis. As recruitment continues and outcomes on more patients are obtained, the “regression towards the mean” ensures a more stable estimate of the treatment effects.

It is beneficial and necessary to ensure the safety of patients to monitor the emerging data as the trial goes on. Data and DSMCs are vehicles for reviewing pharmacovigilance and other safety aspects, hence enabling the reaction to the potential safety hazards in time. On the other hand, information on the comparison between interventions should not be disclosed to the investigators and sponsors in order to avoid biases or misconducts.

1.4.6 Data and Safety Monitoring Committee

The DSMC often consists of clinical experts in the investigational disease area, independent statisticians for the interim analysis and experts in potential specific adverse effects. Sometimes a patient representative can also be a member of the DSMC. The size of a DSMC is often flexible and can vary between 3 (2 clinicians and 1 statistician) and 10, depending on the size of the trial or the complexity of the clinical questions. No financial interest or other conflicts of interests should be shared between members of the DSMC and the trial's sponsor. (40)

In general, blinding is not recommended for members of the DSMC as the point of the DSMC is to speculate and interpret any important emerging result promptly. Blinding may therefore impede a holistic view of the trial. On the other hand, the emerging comparative results of the treatment effects need to be analysed with the minimum bias possible. In this case, external subgroup-blind (someone who sees the treatment allocation tabulations without knowing the actual comparative interventions) statisticians are required to assist with the comparative reports.

Besides looking at safety and efficacy of the trial, the DSMC also reviews other aspect of the trial's progress, for instance, recruitment and retention status, or the overall timeliness of the study. A trial may be subject to an early stop if patient recruitment is challenging. Data quality is another important item that the DSMC needs to closely monitor. In order to ensure interim analyses are performed promptly and accurately, data should be collected, cleaned and presented with high quality. It requires the assiduous effort from the research assistants who regularly collect trial data and data managers who work on data

quality and provide up-to-date data. Missing data can occur during the data collection process, either due to participants' nonresponse or other incidents. Missing data at this stage may in some respects reflect the characteristics of the investigational intervention, for instance, patients may find that the new drug does not work as well as the control, and therefore stop responding to the trial staff rather than not adhere to the assigned treatment. It is important to observe such incidences before missing data accumulate to imperil the trial's validity.

1.4.7 Monitoring an RCT probabilistically

1.4.7.1 Safety stopping

Apart from monitoring early stopping for efficacy, the DSMC also monitor the safety profile of the trial. Serious adverse events (AE) are often scrutinised to suggest whether the trial should stop due to its potential hazardous effect on a patient group. Safety stopping should also consider specific patient characteristics. Oncology trials may often see many toxicity events despite that patients may benefit from treatment overall. Serious consideration should be given when weighing between benefits and risks, the so called benefit / risk ratio.

1.4.7.2 Efficacy stopping

Statistical methodologies for recommending when trials should stop recruiting have been evolving since the 1950s. Such tools, often predefined in the statistical analysis plans or DSMC Charter, can help the DSMC with the decision of whether the trial should continue or stop. It first appeared as 2 specific plans by Bross. (41) The problem of repetitive testing on accumulating data is the inflation of type I error rate. That is, if treatment effects between

groups are repeatedly tested for statistical significance, the probability of finding a significant result becomes higher, but purely by chance. Bonferroni correction may be used in the adjustment of family-wise error rate (FWER), however it inevitably increases the type II error rate. (42)

Table 1-1 Commonly used group sequential stopping boundaries with Z scores and significance levels for different numbers of planned interim analyses

Planned Interim analysis number		O'Brien-Fleming		Haybittle-Peto		Pocock	
		Z	p	Z	p	Z	p
2 interim looks							
	1	2.782	0.0054	2.576	0.0100	2.178	0.294
	2	1.967	0.0492	1.960	0.0500	2.178	0.294
3 interim looks							
	1	3.438	0.0006	2.576	0.0100	2.289	0.0221
	2	2.431	0.0151	2.576	0.0100	2.289	0.0221
	3	1.985	0.0471	1.960	0.0500	2.289	0.0221
4 interim looks							
	1	4.084	0.00005	3.291	0.0010	2.361	0.0158
	2	2.888	0.0039	3.291	0.0010	2.361	0.0158
	3	2.358	0.184	3.291	0.0010	2.361	0.0158
	4	2.042	0.0412	1.960	0.050	2.361	0.0158
5 interim looks							
	1	4.555	0.000005	3.291	0.0010	2.413	0.0158
	2	3.221	0.0013	3.291	0.0010	2.413	0.0158
	3	2.630	0.0085	3.291	0.0010	2.413	0.0158
	4	2.277	0.0228	3.291	0.0010	2.413	0.0158
	5	2.037	0.0417	1.960	0.0500	2.413	0.0158

Table 1-1 summarises the 3 general rules of sequential stopping proposed by Pocock, O'Brien-Fleming and Haybittle-Peto. The widely accepted group sequential methods were introduced by Pocock, after a series of improvements to the classical "closed" sequential methods. (43) By dividing participants into K equally sized groups, the test statistic of each analysis when outcomes for each new group of participants are obtained is compared with the predefined "stopping boundary", which is calculated to give an overall type I error rate (in most cases a two-sided 0.05 level) and is identical for each analysis. However, unifying the significance level for each analysis while keeping the overall type I error rate can result in a problem. The same trial result would be not significant

in this setting whereas it would have been significant in a non-sequential design, as the multiple testing leads to an increase in the significant level. The Haybittle-Peto approach (44, 45) and the O'Brien-Fleming approach avoid this issue and the level of significance in the final analysis is close to the original overall type I error rate. The O'Brien-Fleming boundary has more demanding critical values at the earlier stage of the interim analysis, which is particularly appealing to many DMSCs due to its conservativeness of interpreting the result when the study sample size is limited. At the early stage of a trial additional events can change the overall results substantially simply due to the play of chance and the decision to stop early should be made with considerable caution. A less steep alpha spending function such as the Power Family can be considered attractive where the DMSC is considering safety outcomes, again reflecting a conservative approach when considering patient safety. (2)

Different clinical scenarios may require different approaches in terms of recommending early stopping. In particular, when recommending to stop early for benefit, the Haybittle-Peto boundary has some merit and requires the p-value of less than 0.001 for early stopping, where stringent statistical thresholds are necessary. (46)

The group sequential design provides a solution for the inflation of type I error rate. The pre-specified assumptions of this strategy, one being the number of interim looks needs to be planned in advance and another being equal numbers of participants must be recruited between each look, however, restrict its use in some occasions in which more flexibility is called for. The alpha spending functions allow for flexible addition to the scheduled interim look plan.(47)

During the course of the trial the investigators are allowed to decide how much they wish to allocate the type I error rate based on a spending function. In this method it is not requisite to plan the number or time of interim looks as long as the pre-specified α has not been “spent up”. However, the spending function itself cannot be changed during the trial.

In most cases, a RCT is designed to test for superiority of the intervention under investigation even though the hypothesis is two-sided. It therefore calls for a different standard during the interim look. Moreover, it may be unethical for the trial to continue if it is not showing any major sign of advancement during the interim look, even if it is not showing inferiority. Asymmetrical boundaries can provide a less conservative threshold for futility, which means the study may be suggested to stop if the interim look shows sufficient evidence, although not necessarily statistically significant, for a futile or even harmful effect of the intervention. This approach tries to reflect the decision process of the DSMC, which should always prioritise patients’ best interests. As the data emerge during an RCT, one may also speculate the likelihood of future results based on observed trial data. The question of what the probability of observing a different treatment effect is when there is a difference in treatment effect, based on the observed data, is often asked. Conditional power is then used to answer the question of when there is enough evidence to recommend terminating the trial even if the data remaining to be collected would have suggested otherwise. The decision of stopping a trial early boils down to the balance of the risk of drawing a wrong conclusion and risk of missing the opportunity to offer a beneficial treatment to the public. If the trial is stopped too early, we might run into a chance of drawing an incorrect conclusion of the comparative intervention

by making imprecise or incorrect treatment effect estimates. On the other hand, if the trial is stopped too late, ethical issues may arise as too many patients may be treated with an inferior treatment. (48)

Statistical tools to inform the opportunity of stopping early are indicative during the DSMC meeting. However, they should not be used as the sole basis in the decision.

1.4.7.3 Other considerations

The DSMC can also provide recommendations based on information outside of the trial. For instance, if any DSMC member in trial A is aware of a similar study B that is ongoing, then the result from the study B could inform the likelihood of success in A.

1.4.8 Efficiency in RCT

Conducting an RCT can cost a substantial amount, but the reporting of costs and resource use of RCTs has not been straightforward. In a recent systematic review, overall costs per patient in an RCT could reach more than $103,254 United States Dollars (USD). (49) To ensure RCTs are conducted in an economical fashion without compromising its scientific integrity, RCTs need to be more efficiently designed, conducted and managed. The National Institute of Health Research (NIHR) has been setting up the Annual Efficient studies funding calls for Clinical Trial Units projects to support the design, development and delivery of more efficient, faster, innovative studies to provide robust evidence to inform clinical practice and policy since 2016. (53)

One way to improve the efficiency is by encouraging recruitment into RCTs. However, the nature of a large number of RCTs requires blinding (mainly drug trials), in which patients are unaware of the treatment they are receiving. Summer *et al.* discussed reasons for the preferences, participation and non-participation for psychological therapy in severe mental illness (SMI), suggesting that a number of interviewees expressed preference on treatment arms or disliked the nature of randomisation. (3) Concerns of being randomised into an unfavourable arm (often the placebo arm) or worries of symptoms worsening during the washout phase are also important reasons for non-participation. (50) Hemminki *et al.* compared recruitment to an open design with recruitment to a blinded placebo-controlled design and found that the open design helped increase recruitment by avoiding resentful demoralization. (51) However, the pitfall of an open design is that it would inevitably introduce a certain degree of bias as patients know beforehand which treatment they will be receiving.

Evidence on the efficacy of ways to improve recruitment into RCTs is however, scarce. In practice, recruitment is often carried out in an *ad hoc* manner. That is, trialists often test out different strategies sequentially. RCTs, as the gold standard, also an instrument for assessing the efficacy of different interventions, are not always applicable to recruitment interventions. Although there have been strides into introducing the “Studies Within A Trial” (SWAT) concept, the effect of many recruitment strategies still remains unclear. (36, 52)

Also, the efficiency of RCTs can also be improved with the previous evidence. With previous evidence, using methods such as simulation and decision model

can help to guess what could happen in the RCT, therefore prevent issues before they happen.

1.5 RCT-based economic evaluation

RCTs can provide unbiased estimation of the effect of a health care intervention. However, economic outcomes are also important when health providers plan to apply a new intervention for general use. Treatment costs could, to an extent, limit the choice of available interventions due to the increasing burden on the budget. For instance, a certain budget of X with a cost of Y per patient treated will result in a number of X/Y patients getting treated. To increase the value of X/Y, the choices are either to increase the budget X, or to reduce Y. However, with the limited healthcare resource, it is not the case that X can increase considerably for everyone to have access to the expensive treatment. It is therefore more sensible to reduce Y. Trial-based economic evaluation has the potential to provide estimates of benefit associated with new treatments that may be more expensive than the current one. Cost-effectiveness analysis (CEA) aims to compare treatment strategies by calculating incremental cost-effectiveness ratio (ICER), defined as follows,

$$ICER = \frac{\Delta C}{\Delta E} = \frac{C_T - C_C}{E_T - E_C},$$

where C_T=cost of new treatment, C_C=cost of controlled (standard) treatment, E_T=effectiveness of new treatment, E_C=effectiveness of controlled (standard) treatment. It is often the case that quality adjusted life years (QALYs), or changes (gains or losses) in QALYs, are used as a measure of effectiveness and can be calculated using questionnaires such as EQ-5D and associated

preference tariffs. (4) EQ-5D is a standard and validated generic instrument that is widely used in many patient populations. Although criticism of EQ-5D has been received on its inadequacy of capturing certain domains of interest, for instance, vision, hearing or SMI. It is highly useful as it provides a universal currency of communicating between different health states and is efficient for comparison and synthesis. (54)

Meanwhile, the ICER may also be interpreted as Net Monetary Benefit (NMB), in which a willingness to pay (WTP), λ is taken into account, together with the incremental effectiveness ΔE and costs ΔC.

$$NMB = \lambda \times \Delta E - \Delta C$$

In the case in which cost-effectiveness is the primary study objective and cost-effectiveness outcome is defined as primary outcome of a study, sample size and power calculation should include relations to the cost-effectiveness results rather than the effectiveness results alone. NMB based on certain WTP threshold may be used for power calculation for cost-effectiveness outcome. (55)

Similar to other types of outcome data, missing data may often be an issue in trial-based economic evaluations for both costs and outcomes, and the issue of inappropriately handling missing data in trial-based economic evaluations has been widely acknowledged.(56) Complete case analysis has been used most frequently across studies whereas multiple imputation has seen a rise in the three years. (57) Since 2018, there has been comprehensive guidance available on handling missing data in trial-based economic evaluations that consider the different assumptions and missing mechanisms for both costs and outcomes. (58)

According to the National Institute for Health and Care Excellence (NICE) guide to the methods of technology appraisal, QALYs are considered the most appropriate generic measure of health benefit that reflects both mortality and health related quality of life effects. (59) As for NICE, an intervention that can cost below around £20,000 to £30,000 per QALY gained would likely be considered cost-effective compared with the current recommended intervention. (60) However, heated debate has been sparked as to "how much is too much" for the National Health Service (NHS) to pay for a new drug or intervention. Claxton *et al.* gave an estimation about just under £13,000 for the opportunity cost of a QALY in the English NHS in 2008/09 primary care trusts (PCT) data, suggesting that a decrease in PCT spending of £13,000 will lead to the loss of 1 QALY and NICE threshold should be no higher than this. (61) Although the cost-effectiveness threshold should equal to the budget decrease which would cause the NHS to lose 1 QALY, others criticise that the analysis may require too strong assumptions and the results are too sensitive to alternative assumptions. (62, 63) While NICE responded that setting the cost-effectiveness threshold too low may discourage innovation of new technology and the £20,000 – £30,000 threshold represents reasonable compromise between the fairness of public access to the NHS and encouraging access to innovative therapeutic strategies. (64)

1.6 Decision analytic models and Value of Information

Decision analytic modelling approaches are prevalent in appraising health technologies for public use and are now common practice in economic evaluations of health care technologies. As randomised trials often do not necessarily compare all the available options, health economic evaluation

alongside one randomised trial may fail to include evidence from other trials, meta-analyses or observational studies. Besides, RCTs have their limitation in providing long term evidence because of relatively shorter follow-up time. Decision analytical modelling in this case, provides another option for economic evaluation.(60)

1.6.1 Decision trees

The decision tree approach is a simple form of decision model. The key features include:

- A square decision node, indicating a decision point between alternative options, often seen at the start of a tree;
- A circular chance node showing a point where more than one event for a participant is possible; the alternatives are shown as multifurcation from the node. It represents the uncertainty of events which an individual patient may experience;
- Paths are mutually exclusive;
- Probabilities of the occurrence for each particular event path.

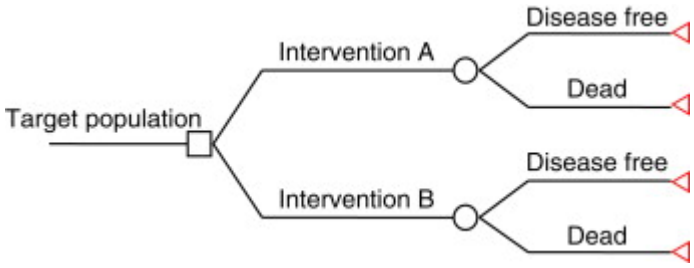

Figure 1-2 An illustrative example of a decision tree

Figure 1-2 illustrates the structure of a decision tree. The expected costs and outcomes are then calculated based on the summation of each path, weighted by probabilities of occurrence for each path. The decision tree features simplicity and transparency which are excellent in explaining different decisions.

However, the assumption of non-recurrence or looping could complicate the model considerably.

1.6.2 Markov models

Markov models represent a series of events connected by a more complex form of transition. It allows for a flexible sequencing of outcomes including recurring outcomes, through time with complexity. Patients are assumed to remain in any health states which were embedded in the model or can transfer to another one with a predefined transition probability, over a series of discrete time intervals or cycles. The number of cycles or durations, and health states will depend on the decision problem and can vary on different degrees.

1.6.3 Other modelling approaches

Markov models and decision trees, or a combination of the two, are widely used models in economic evaluation, while some other methods are available. Patient level simulation models, discrete event simulations (DES) or dynamic models are also sometimes used in decision modelling. The choice of model largely depends on the decision question, but also on the availability of relevant data required for each model.

1.6.4 Value of Information

Value of Information analysis (VoI) plays an important role in health technology assessment. It provides estimates on returns from investing in research projects and can inform different aspects of the decision. Funding bodies often find it informative to rank projects regarding expected return on investment amongst competing projects. Moreover, trialists can use this method to identify the

efficient sample size of a trial as an alternative to the traditional power calculation. Along health economic evaluation, VoI can also be used as a quantitative adjunct to the proposition of any future projects. (65)

Expected value of perfect information (EVPI) is the price value that a decision maker would be willing to pay for perfect information on all factors that influence the treatment of preference from a CEA. Using the EVPI we could answer the question of whether the decision of adopting a technology under the current WTP threshold, should be made on currently available information or whether it is worth investing in additional information to reduce uncertainty. In other words, we wonder whether the value of additional information outweighs its cost. (65) In health technology assessment (HTA), the question is whether we need to obtain additional information by conducting future data collecting research to reduce uncertainty on the proposed new intervention. The possible options include delaying the adoption decision and conducting future research; adopting the new intervention and also conducting a future research to reduce uncertainty; or adopting the new intervention without any further information. (66) EVPI may provide an overview of the value of future research in reducing uncertainty. Expected value of partial perfect information (EVPPI) is the price value that a decision maker would be willing to pay for perfect information on one or a selection of factors that influence the treatment of preference from a CEA. EVPPI may inform what parameters in the decision model on which we may find worth investing future research to reduce the decision uncertainty. Further, using the expected value of sample information (EVSI) and expected net gain of sampling (ENGS) could inform what the optimal sample size would be if we would like to take future research for a certain group of parameters

based on the results of EVPPI. EVSI values the decision to collect additional sample to reduce uncertainty where as ENGS is the difference between EVSI and the cost of conducting the data-collecting practice.

For a standard NICE technology appraisal, the adoption decision can be made often at between £20,000 - £30,000/QALY WTP threshold for the new intervention. In the case that the decision may subject to large uncertainty, VoI analysis may help in deciding whether it is appropriate to adopt immediately, or to conduct further research before making the adoption decision.

Different from statistical inference, where decision making is based upon hypothesis testing for a false error rate defined *a priori*, decision making for economic evaluation is often made on the probability of cost-effectiveness under a WTP threshold. This is generally due to the nature of economic evaluation outcome often being secondary in RCT hence the issue of alpha inflation can arise as a result of multiple hypothesis testing, especially for ICER which had two dimensions. Moreover, hypothesis testing often informs the decision of the existence of a clinical benefit (difference) through rejecting the null, which is often “no difference”, whereas economic evaluation is not for the purpose of demonstrating “no difference” in incremental cost per QALY, rather, the value of ICER is usually acceptable within a range for which probability is presented.

1.7 RCT, decision modelling and VoI in the decision process

RCTs answer the question of the efficacy of therapeutic interventions. They can provide unbiased estimates of the treatment effects upon which the decision of whether approving interventions for public access by regulatory agencies, such

as the US Food and Drug Administration (FDA), is based. Evidence from RCTs suffices to answer the decision questions of approving new interventions based on its safety and efficacy profile. On the other hand, when it comes to investing in new technology which often has higher price, the decision question involves maximising the health benefit under a fixed budget. Hence, effectiveness becomes not the only criterion for adopting new interventions, costs should also be considered. Furthermore, utility is introduced to help compare the “value for money”, not only within the population with the same condition, but also across different disease areas. QALYs create a universal measurement on which the effectiveness of different interventions across different disease areas can be compared.

Although RCTs can provide an unbiased estimate of the treatment effect, they often have a relatively short timeframe and therefore are inadequate to answer the long-term cost-effectiveness questions. Of course, post-marketing surveillance studies can help to answer whether any long-term toxicity or adverse effects are associated with the intervention. Large cohort observational studies are the main study design for this issue, despite risk of unknown confounders that could lead to biased results. Hence, these studies may fail to give a reliable answer, particularly when adverse effects appear in the longer term, where many other unknown factors can affect the outcomes. Phase IV trials can also provide unbiased estimates. However, they often employ much larger sample sizes and incur much more cost, hence the decision of conducting phase IV trials depends on how certain it is to confirm the hazardous effects. In this respect, there may also be a lack of incentive to

sponsor such trials, where a large amount of money is spent to confirm the intervention is not appropriate for public access.

Decision models for economic evaluations, on the other hand, have the advantage of making use of all the relevant evidence (RCTs, observational studies, expert opinions, etc.) and incorporating a time horizon of interest. In this context, decision models may be seen as complementary to RCTs, where RCTs can provide estimates for the parameters needed in the decision model. Decision models may also inform future research by synthesising all available information in the past and providing an “educated guess” as to what might happen to a clinical question, therefore can lead to the design properties and assumptions in a proposed RCT.

The aim of HTA, on the other hand, is to bring innovation that can treat health conditions and therefore improve quality of life. Attention must be given to not get the priorities mixed up. Improvement of a specific condition, or the effectiveness of an intervention, is characterised by the changes (for the better) in relevant clinical measurements, which should be chosen as the primary outcome of a trial. QALYs, although for easy comparison across disease areas, should not be used as the sole consideration in measuring the improvement in health. One might easily manipulate improvement in QALYs in an RCT with weekly post with a £20 note in an envelope versus weekly post with a blank sheet of paper in an envelope amongst patients with alcohol-use disorders, and it is almost certainly going to be more cost-effective than cognitive behaviour therapy. But does the weekly £20 payment really solve the alcohol-use

disorders problem? This may be an multi-attribute utility instrument issue rather than a QALY issue.

VoI analysis helps with the decision question as to whether an intervention should be adopted with the current level of uncertainty, or further research is worthwhile to reduce the uncertainty of the parameters in order to reduce the cost of making a wrong decision. If the expected cost of a project exceeds the expected benefit, then further research should not be carried out and decisions are made based on the current level of uncertainty, on the other hand research may be worthwhile if the expected benefit is great than the expected cost of a new project.

In the following sections and chapters, I will discuss and explore the ways in which the efficiency of RCT in mental health may be improved using some of the concepts and methods described previously, in an RCT example that compares the efficacy and safety of an antipsychotic reduction and discontinuation strategy for patients with schizophrenia.

1.8 RADAR

The Research into Antipsychotic Discontinuation And Reduction (RADAR) study is a RCT which aims to estimate the potential benefits and harms of gradually reducing antipsychotic use in people with schizophrenia and non-affective psychosis. Long-term antipsychotic treatment is the benchmark for patients with psychotic episodes or schizophrenia diagnosis. It is also recommended by the NICE that patients remain on lifelong antipsychotic treatment. This has impacted the NHS in many respects. First of all, many patients hope to see whether they can manage without antipsychotic medications, as it may be

discouraging for patients to accept that they have to take medication for life. With the side effects such as diabetes, cardiovascular conditions or EPS caused by antipsychotics, patients are faced with the risk of unsettling experiences for the rest of their life. Also, health services have devoted substantial efforts into not only patients' adherence to antipsychotics, but also maintaining the financial capacity to fund for these medications. Previous research suggested that it is possible that long-term antipsychotics may lead to a worse state of a patient, and patients who tried to reduce and stop antipsychotic, no matter successfully or not, functioned better than the ones who were continuously on treatment. (67) This project has the potential to assess whether it is feasible to reduce and discontinue antipsychotic medication on patients who already remain in a stable state, thus lightening the burden of both patients and health care providers.

RADAR is a phase IV, open, parallel group, multi-centre RCT that aims compare a flexible and gradual antipsychotic reduction programme with its counterpart maintenance treatment. Participants will be individually randomised to the two treatments strategies, which will be delivered by clinicians. The follow-up period will be 2 years, and follow-up assessments will be conducted at 6 months, 12 months and 24 months for all participants. Wellbeing, as measured by the social functioning score (SFS), will be assessed as the primary outcome, whereas serious relapse, symptoms, side effects, employment and adherence will be assessed as secondary outcomes. RADAR employed a non-inferiority design with the non-inferiority margin of 10 % on the severe relapse outcome. Estimated treatment effects are obtained from a meta-analysis on antipsychotics versus placebo for relapse prevention in

schizophrenia. (20) The sample size of 372 has 90% power to exclude a difference of 10%. An extra 15% is added to take into account attrition, which makes the final sample size of 402. This also has 98% power of detecting 2 points of difference in the primary outcome SFS, and 90% power for 1.6 points of difference.

The estimated total duration is 54 months. Main inclusion criteria include age (over 18 years); diagnosis (schizophrenia, schizoaffective disorder, delusional disorder or other non-affective psychosis); more than one previous episode or a single episode lasting for more than a year; under ongoing antipsychotic prescription, and no hospital admission for at least 3 months. Exclusion criteria include the lack of consent capacity; insufficient command of spoken English language; under Community Treatment Order that requires antipsychotics medication; serious risk of harm to self or others; females who are breastfeeding or pregnant. RADAR will investigate a strategy of supported and flexible antipsychotic discontinuation with regard to social functioning improvement and risk of severe relapse, providing rigorous evidence for the efficacy of long-term antipsychotic maintenance compared with reduction and discontinuation. It will also provide potential evidence on which recovery can be justified.

In this chapter I have introduced the background and rationale of my thesis, including RCT, economic evaluation, mental health, antipsychotics and the RADAR trial. In the next chapter I will start discussing the efficiency of mental health RCTs, by reviewing the existing studies on the effectiveness and cost-

effectiveness of different strategies to improve recruitment and retention in mental health RCTs.

2 Chapter 2 – A systematic review on strategies to improve recruitment and retention in mental health RCTs

Recruitment and retention are now major challenges in conducting RCTs, and they are crucial to a clinical trial's efficiency. In this chapter I systematically review current studies on recruitment and retention strategies in mental health RCTs, and compare the effectiveness and cost effectiveness of different recruitment and retention strategies. A major part of this chapter has been published in PLoS ONE. (68) The aim of this chapter is to evaluate the evidence base for strategies to improve the recruitment and retention of patients to clinical trials in mental health. A secondary aim is to evaluate the cost and effectiveness of different recruitment and retention strategies, reported as the cost per patient recruited, or cost per patient retained.

2.1 Introduction

As the gold standard of clinical research, RCTs are often faced with the challenge of recruiting enough participants and it may be more common in mental health. Often requiring extensive collaboration between clinical researchers, patients, clinical professionals and research institutions, each party in a trial has their unique perspective, expectations and concern in a trial. (69) Concerns about mental health patients' vulnerability and reduced decision-making ability often make recruitment more difficult. (70) Particularly for patients, doubts of participating in a trial primarily centre on the risks or benefits to their own health. During the consent process, in which potential participants are introduced to trial's information, they could be easily put off if anything inconvenient, abstruse or irrelevant occur.

The fundamental biological aetiology for some mental health conditions is still not well understood, and often the effects of psychiatric treatments are small and uncertain. Hence there may be scepticism that new treatments will be very helpful, which might make psychiatric trials less appealing.(71-73) High placebo response rates also highlight the importance of randomised trials in providing unbiased estimates of treatment effects. Patients with mental health problems often still consider their conditions as stigmatised (sadly often for good reason) and conceal their condition and treatment from public attention. Also, for some mental illnesses, there are ethical concerns when involving patients who are at high risk or have a history of aggression or self-harm. These concerns make recruitment to mental health clinical trials challenging. (74)

Retention is another pivotal component to a trial's scientific success as it is key to a trial's validity. Attrition may happen in drug trials because of side effects of the medication. Some patients may experience deterioration of their health during the follow-up period, making them reluctant to continue with treatment or the trial. In trials of complex interventions, such as cognitive behavioural therapy (CBT) or early supported discharge, the absence of blinding in the control arm means that the participants know that they have not received the intervention, which may reduce engagement with trial follow-up or increase the risk of drop-out. It has been suggested that high drop-out rates are associated with larger sample sizes in antipsychotic trials, more specifically trials with multi-centre design. (75) However in modern trials, the sample size required often necessitates a multi-centre design as a single site would not provide enough participants and may not provide sufficiently generalisable results. This requirement for a multi-centre design might result in retention issues.

Effectiveness means whether a given intervention works in a meaningful way in real world. It can be defined as the extent to which an intervention achieves its intended effect in the usual clinical setting, in terms of changes in the relevant outcomes. It is often evaluated in real-world studies (for instance, observational studies, pragmatic trials, etc.). Cost-effectiveness evaluates the effectiveness of a new intervention relative to the difference in cost of offering it or not, often summarised using the ICER. The efficiency of trial design explores the possibilities in which resource (patient numbers, conduct procedures etc.) may be used more efficiently without compromising the integrity of scientific objectives of an RCT. An efficient trial design could take a perspective from patient recruitment. On one hand the key to a successful trial is to recruit enough patients to ensure an RCT has enough power. However, it is also important to avoid recruiting an exceeding amount of patients where resources could have been used on other research projects.

There has been extensive effort in exploring efficient trial design, such as introducing master protocols for a group of trials in oncology. Investigators and trialists have been seeking ways in which patients may be recruited more efficiently, including less resource use and fast recruitment process. Strategies that facilitate recruitment could benefit recruitment process, potentially save resources. (76)

Previous systematic reviews by Treweek *et al.* and Brueton *et al.* investigated the efficacy of different strategies to improve recruitment and retention to randomised trials.(77, 78) However, evidence in the mental health trial population remains to be thoroughly understood. The review by Treweek *et al.* summarised the efficacy of different strategies to improve recruitment into

randomised trials, but only included 3 eligible studies in mental health. No mental health studies were included in the systematic review by Brueton *et al*., which investigated the efficacy of different strategies to improve retention. Treweek *et al.* found that open trial design, and telephone reminders to people who do not respond to postal invitations may improve recruitment, whereas bespoke participant information materials helped little in recruitment.(36) Offering a small financial incentive for completing follow-up questionnaires appeared to help retain patients in the trials, as suggested by Brueton *et al*. (78) An increasing number of studies employ the use of a SWAT method to assess the impact of technical or design innovations on a trial's efficiency.(52) To date, most different recruitment strategies are usually employed in an *ad hoc* manner. Evidence on comparing recruitment strategies retrospectively and observationally can also provide some insight before SWATs are planned.

2.2 Methods

2.2.1 Criteria for considering studies for this review

2.2.1.1 Type of studies

Two reviewers independently screened titles and abstracts and any disagreements in selection were resolved through discussion. Studies that used randomised or observational methods to compare different recruitment strategies designed to recruit participants to RCTs of interventions for mental health problems were considered. Embedded randomised studies of different recruitment strategies were identified, but given the small number of such studies, RCTs of mental health interventions which reported the effectiveness a range of strategies used in recruitment retrospectively (e.g. without randomising to different recruitment strategies) were also included. For retention,

randomised trials of different retention strategies that were embedded in a randomised clinical trial (host trial), or within epidemiological studies such as cross-sectional surveys were included. A full description of the study protocol is available in Appendix A Protocol for the systematic review in Chapter 2.

2.2.1.2 Types of data

Studies comparing recruitment or retention that involved adult participants with mental health problems, regardless of gender, ethnicity or geographic location, were included. Of particular interest were trials including patients with serious mental illnesses (SMI), such as schizophrenia, but given the expectation of finding only a small number of studies involving these patients, the criteria were broadened to include common mental health problems such as depression and anxiety. Dementia and other organic mental health conditions were excluded, given the different context in which these trials are likely to be conducted. Studies on substance misuse were also excluded as this group of patients is likely to present different recruitment and retention challenges. Studies which did not report outcomes on recruitment or retention strategies for RCTs, studies in which mental illness was comorbid with other physical medical conditions (e.g. CVD) because patients with severe co-morbid physical conditions tend to be excluded from severe mental health trials. Studies not involving adults (e.g. children or adolescents) were also excluded.

2.2.1.3 Types of methods

Strategies aimed at enhancing recruitment and retention included, but were not limited to:

- Incentives for either or both of patients and clinicians

- Advertising
- Periodic phone call follow-up
- Mailshots and newsletters
- Customised or optimised consent materials
- Amendments to protocol
- Presentations to appropriate groups
- Presentations at conferences
- Trial material customised to specific sites
- Resource manual for recruiters

2.2.2 Types of outcome measures

2.2.2.1 Primary outcome

For recruitment, the main outcomes of interest were the type of strategies employed in different studies and the number of patients recruited using each individual strategy. I and the second reviewer also extracted data on how many potential participants were approached, if available, using each different strategy in each study. For studies comparing different retention strategies, the primary outcome was 'response', defined as the percentage of participants who were successfully engaged in follow-up assessments via each strategy out of the total number of people initially randomised to that strategy.

2.2.2.2 Secondary outcomes

Secondary outcomes are the cost of each patient recruited/retained through a specific strategy (if any mentioned), defined as the mean cost per patient recruited or mean cost per patient retained, respectively.

2.2.3 Search strategy

I designed a search method for identifying published randomised trials that focused on improving recruitment and retention in mental health randomised trials. No language restrictions apart from English language abstracts were applied.

The search method was comprised of 4 components, each of which included both free-text terms and subject headings. The Boolean operator OR combined terms related to enhancing recruitment and improving retention. This was then combined using the AND operator with terms related to mental health conditions and RCTs. A brief search strategy is described as follows:

(informed consent OR recruit OR particip) OR (retention OR attrit OR retain)

AND

Randomi#ed controlled trials

AND

Mental health condition filters

Electronic databases searched included:

- MEDLINE, Ovid (1946 to date of search, searched on 28 July 2016);
- EMBASE, Ovid (1980 to date of search, searched on 28 July 2016);
- PsycINFO, Ovid (1806 to date of search, searched on 28 July 2016);
- Cochrane Methodology Review Group Specialised Register (CMR) (from inception until July 2012, searched on 28 July 2016).

An updated search was conducted on 25 February 2020 for MEDLINE, EMBASE and PsycINFO. CMR database ceased to update since 2012 therefore no update search was run for CMR.

The full search strategies for all of the 4 databases are included in Appendix B Search strategies for the systematic review in Chapter 2.

2.2.4 Data extraction and analysis

2.2.4.1 Data extraction

The second reviewer extracted the data from eligible studies.

Data extracted for the recruitment trials and their corresponding main trials included:

- For host trials: country, disease area, design, sample size, setting, primary outcomes, funding body;
- For embedded randomised recruitment trials: strategies to which participants were randomised, number of participants in each arm who were recruited to the main trial;
- For studies that compared recruitment strategies retrospectively: strategies used for recruitment, number of participants recruited and approached via each strategy.

For retention trials and their host trials, data extracted included:

- For host trials: country, disease area, design, sample size, setting, primary outcomes, follow up period, funding body;

For retention trials: strategies to improve retention; retention rates for each strategy.

2.2.4.2 Assessment of risk of bias

For each eligible study, risk of bias was assessed by Critical Appraisal Checklist for RCTs developed by the Joanna Briggs Institute (JBI). I calculated the overall score based on the number of items checked for each assessment. Details of the risk of bias assessments are included in Appendix C Risk of bias assessment for the included studies in Chapter 2.

2.2.4.3 Data analysis

For randomised comparative studies, relative risk was used to describe the effect of each recruitment strategy. Non-randomised studies were categorised according to similarity of strategies, for instance, by combining optimised consent materials and incentives. Strategies were ranked based on the numbers of patients recruited and identified strategies that recruited most participants in each study. I also calculated the total number of patients randomised through each recruitment strategy for recruitment studies and the number of responses in each retention strategy for retention studies. Cost and effectiveness of strategies where cost data were available was measured by average cost per patient randomised or average cost per response, respectively. Cost information was first converted to the equivalent monetary value in 2016 using relevant inflation rates of the study country, and subsequently into British pound sterling (GBP) based on average exchange rates between each currency and GBP in 2016. (http://www.ukforex.co.uk/forex-tools/historical-rate-tools/yearly-average-rates). The average costs of each category of recruitment strategy was calculated using a weighted average approach, where the mean costs were weighted by the sample size.

For studies where cost data were not reported, relevant information on the processes were extracted and I generated the cost using available reference cost information (from e.g. Personal Social Services Research Unit, PSSRU). Given the considerable uncertainty in this approach due to insufficient information on resources used, I also performed sensitivity analysis under different scenarios. For instance, the cost of recruiting using health care providers, or research staff, depends on the number of hours spent on recruitment. The assumptions made were: part-time (e.g. 3 hours/day) versus full-time (e.g. 7.5 hours/day). The costing mainly employed a bottom-up approach from a UK NHS/personal social services (PSS) perspective. The unit cost of each component mentioned during the recruitment process was multiplied by the recruitment duration, or study duration otherwise, before all the relevant cost components were summed to make the total recruitment cost. The cost-effectiveness of each strategy was obtained using the total cost divided by the total number of participants for each strategy.

2.3 Results

2.3.1 Results of the search

A total number of 5157 records were identified from electronic databases from inception of records until July 2016. Of these, 116 were identified for full-text screening and 12 were found to be eligible for inclusion. One additional study (Hughes-Morley 2016) identified by one of the reviewers during the peer-review process of PLoS ONE publication. This article was also included. Out of the 13 included studies, 11 are studies on recruitment strategies in the mental health clinical trial context, and 2 focused on retention. Five out of 13 were randomised comparative trials of different recruitment or retention strategies, and 8 were

observational comparisons of recruitment strategies embedded within a randomised trial. Figure 2-1 shows the flow chart of the article selection process. Figure 2-2 shows the results of the updated search and study selection process. Amongst the 2 included studies, Hughes-Morley 2016, which was originally identified by one of the reviewer during the publication process, was identified during the update search and was consequently included.

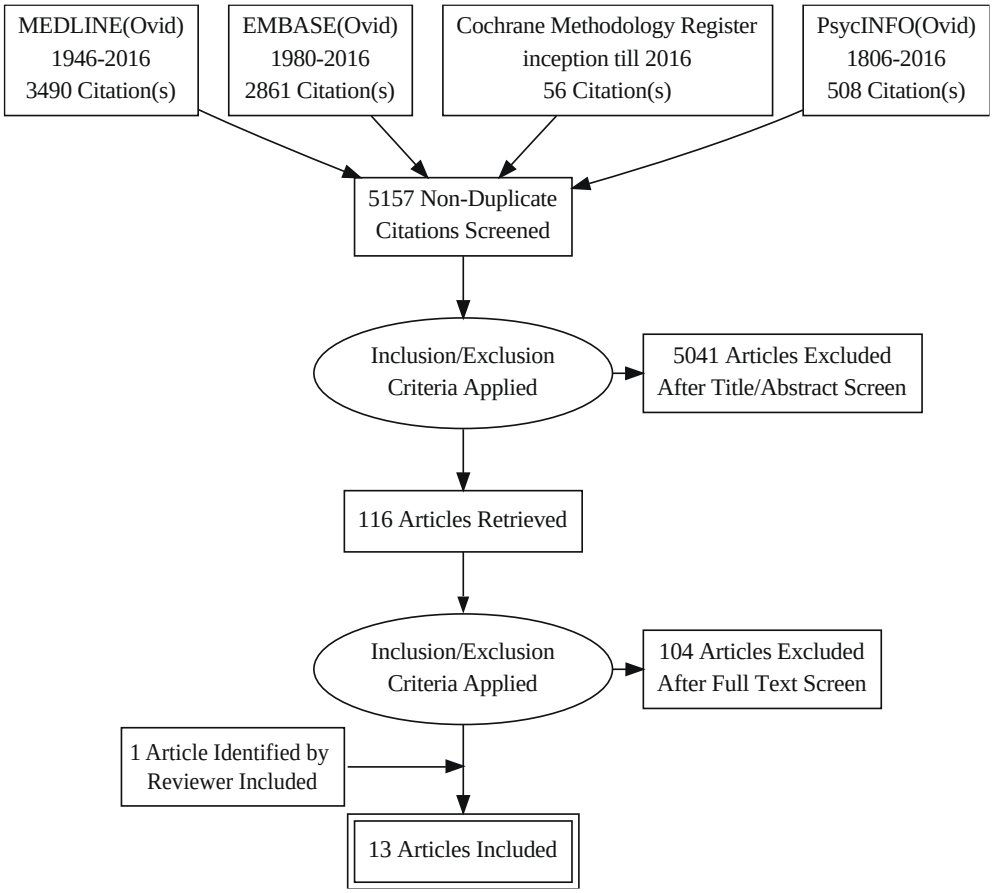

Figure 2-1 the flowchart of study selection process

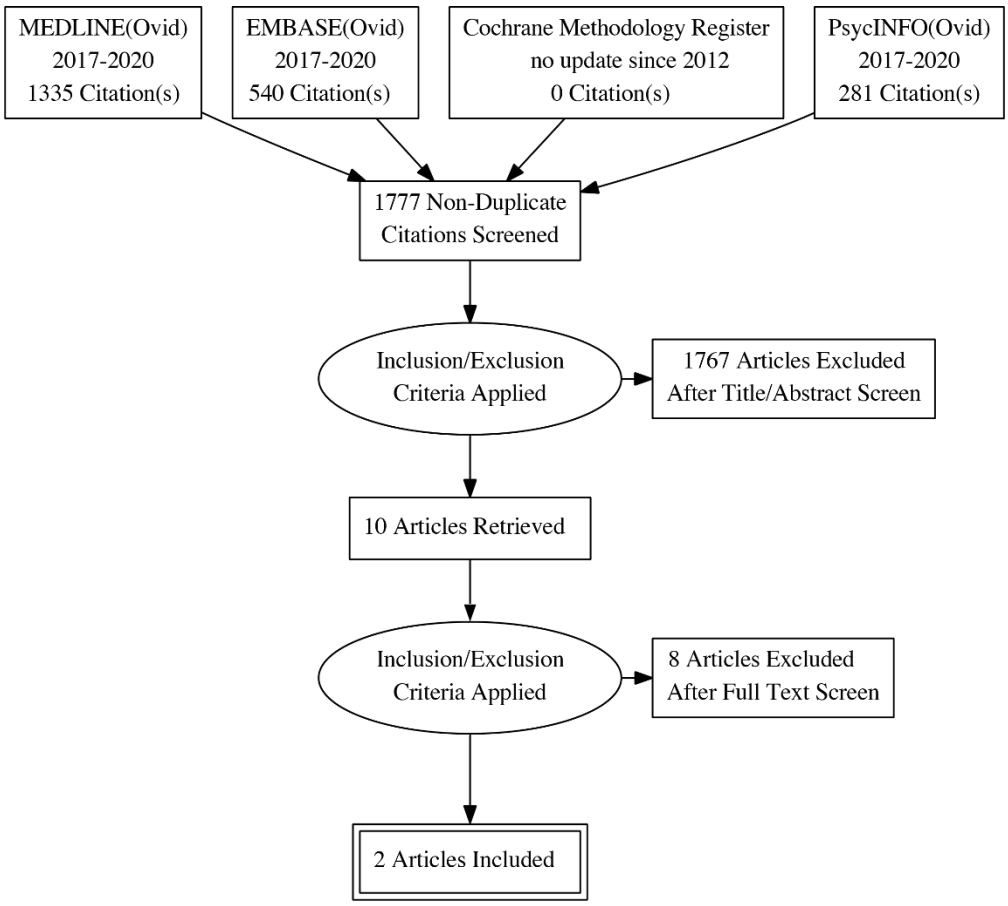

Figure 2-2 the flowchart for the update search and selection since 2017

2.3.2 Description of the included studies

- Man 2015 (79)

  A recruitment trial embedded in a large primary care research programme of two multi-centre trials. 1364 patients with depression were recruited to either original patient information materials or optimised version of the material (trial booklet). Outcome was percentage of participants randomised. 27 out of 682 patients were recruited using original patient material and 43 out of 682 patients were recruited using optimised version of the material.

- Jeste 2009 (80)

  A recruitment randomised trial for a hypothetical drug trial for healthy individuals and schizophrenia patients. 128 patients with

schizophrenia and 60 healthy participants were randomised to either a multimedia educational consent procedure or general routine consent procedure. Outcomes were numbers of schizophrenia patients who were willing to participate in the hypothetical trial. A total of 128 patients with schizophrenia took part in the study. 41 out of 62 patients in the multimedia group and 44 out of 66 patients were willing to participate in the hypothetical trial.

- Krusche 2014 (81)

  An evaluation on 10 recruitment strategies utilised to recruit participants with a history of recurrent depression into a RCT. Recruitment strategies include newspaper advertising, web-based advertising, advertising at exhibitions, radio advertising, advertising on buses, poster advertising, GP referrals, mental health care referrals, word of mouth, and charitable organisations referrals. Outcomes include numbers of participants who were randomised into trial through each recruitment strategy and cost effectiveness of each strategy.

  Table 2-1 describes the recruitment information detail and the cost incurred for different strategies.

Table 2-1 Summary of recruitment strategy of Krusche 2014

Strategy	Number of patients recruited	Number of patients approached	Cost per patient recruited (in GBP)
Word of mouth	16	46	0
Information from charity	2	8	0
Posters	30	123	69
Web-based adverts	37	300	105
Mental health care referral	8	32	178
Radio adverts	26	412	241

GP referral	18	116	396
Bus adverts	2	4	571
Newspaper adverts	11	101	805
Exhibition	3	11	2562

- Morgan 2013 (82)

  A study on internet-based recruitment to a depression prevention trial. Different online recruitment sources were used. Outcomes were numbers of participants who entered the trial. Available cost information was also collected for cost effectiveness evaluation. 755 patients were recruited via Google Advertising and 35 via Facebook Ad, with cost per patient recruited being $14.71 and $19.89 (AUD) respectively.

- Rollman 2008 (83)

  A comparative study between recruiting by primary care physicians' referral to electronic medical records (EMR) and recruiting by research assistants' outreach in practice waiting rooms. Outcomes were numbers of patients who were enrolled in the trial through each different strategy. EMR – prompted primary care physicians referred 794 patients and 176 were recruited, whereas clinical waiting-room recruitment approached 8095 patients and recruited 193 participants.

- Daley 2008 (84)(5)

  A feasibility RCT on an exercise intervention for women with postnatal depression. Four recruitment strategies were used in the trial, including GP referral, special care unit referral, self-referral and health visitor referral. Outcomes were numbers of patients being

randomised via different strategies. 24 out of 96 patients were recruited by general practitioners (GP); 12 out of 28 were recruited by “mother and baby unit”; 8 out of 10 were recruited from health visitors and 3 out of 4 were recruited by self-referral.

- Woolhouse 2014 (85)

  A pilot RCT using mindfulness interventions to reduce antenatal depression, anxiety and stress. Three recruitment channels were used: recruiting at clinic waiting rooms, recruiting via mail-out from hospitals, recruiting via specialist care unit (physiotherapy and childbirth education classes). Outcomes were numbers of patients being randomised via different strategies. 14 patients were recruited at clinic waiting room, 16 were recruited via hospital mail-out, 2 were recruited at education classes.

- Debar 2009 (86)

  A study that discussed recruitment for a guided self-help binge eating prevention RCT. Recruitment strategies used include invitation to comprehensive eating disorder examination (EDE) assessment with $5 incentive for completing an online questionnaire and $50 for baseline assessment, invitation to abbreviated EDE assessment with incentive of $25 for baseline assessment, and self-referral. Primary outcome was the numbers of patients who entered the randomisation of the main trial. Secondary outcomes of cost of incentives were also collected. Of the 11984 patients approached in the comprehensive EDE recruitment wave, 70 were randomised. 154 patients were recruited via the abbreviated EDE recruitment wave, in which 20810 patients were invited to take part.

- Le 2008 (87)

  A study that discussed recruiting Latino women in the U.S. and women in Mexico into a multisite postpartum depression prevention trial. The study was conducted with immigrant Latinas in Washington, DC, U.S.; the other site was with women in Mexico City, Mexico. Recruitment in the U.S. used different strategies that compared outreach with potential participants by community health centre staff with recruitment by clinical research staff at a prenatal care clinic. 217 patients were recruited in the U.S. site. Primary outcome were numbers of patients who entered the randomisation. (87)

- Schlernitzauer 1998 (88)

  A study that compared various strategies for recruiting elderly with bereavement-related depression in to a randomised placebo-controlled trial which tested the efficacy of nortriptyline and interpersonal psychotherapy for the acute and continuation treatment of bereavement-related depression in Pittsburgh, U.S. Recruitment period lasted for 5 years. A total of 65 patients were recruited. Media advertisement was most successful strategy and recruited 54% participants (n=35). Other strategies used included friend or acquaintance, obituary letter, psychiatric referral, and so on.

- Hughes-Morley 2016 (89) (identified in update)

  A recruitment randomised trial for the EQUIP host trial – a clustered RCT of a new user-led training package to increase user and carer involvement in care planning for patients with a diagnosis of SMI under community mental health teams. Patients with different cluster

pairs were randomised to either invitation with patient and public involvement research (PPIR) invitation leaflet or ordinary recruitment procedure. 216 out of 5382 patients were recruited in the PPIR group and 148 out of 2800 were recruited in the ordinary recruitment group. Telephone follow-up of non-respondents recruited a total of 221 patients. However, PPIR did not show statistically significant difference for improving recruitment.

- Brown 2019 (90) (identified in update)

  A study that reported their strategy of using non-diagnostic title and self-referral to a brief “self-confident” workshops intervention RCT for depression. Overall 459 patients were recruited within 12 months via self-referral. It was suggested that self-referral to a brief intervention for depression with a non-diagnostic title could be an effective strategy of recruiting depressed people into RCTs.

- Dirmaier 2007 (91)

  A randomised study with a 2×2 factorial design that investigated monetary incentives and shortening the questionnaire in relation to response rates in a mailed follow-up survey 1 year after psychotherapeutic treatment. Partial nonresponse and self-report of treatment outcome were also assessed. 3825 patients were randomised to (1) receiving prepaid small bill incentive or none; and (2) getting abridged or normal questionnaire. It showed that response rates were significantly increased by 7.3% (95%CI 2.6 to 11.9%) when using incentives, and 3.7% (95%CI 0.9 to 8.3%) higher when using a short version of questionnaire.

- McLean 2014 (92)

  A randomised study that investigates the effect of pre-notification or envelop teaser on response rates in a bulimia nervosa mental health literacy survey. A 2 (pre-notification present; absent) by 2 (teaser present; absent) design was used. Questionnaires were mailed to 3010 adults, and significantly higher response rates were found for the use of pre-notification.

2.3.3 Recruitment strategies

2.3.3.1 Characteristics of the included studies

Table 2-2 describes the characteristics of the studies included which looked at recruitment strategies. Overall, three studies employed a randomised design for comparing recruitment strategies (Man 2015, Jeste 2009 and Hughes-Morley 2016). The other studies compared different recruitment strategies retrospectively without randomisation. Five studies were carried out and funded in the UK, 5 in the US and 2 in Australia. One study involved recruitment to a preventive programme for depression, and one involved a relapse prevention trial in women with a history of post-partum depression. Two of the studies were conducted with people with SMIs. Five were carried out in a primary care setting. Four involved female participants only. Except for one RCT which was a study of recruitment to a hypothetical trial, the studies involved recruitment to randomised trials involving a range of interventions including mindfulness CBT, health promotion via email, telehealth intervention, exercise, antidepressants, interpersonal therapy and psycho-education.

Table 2-2 A summary of the characteristics of included studies on recruitment strategies

Study ID	Trial design & intervention	Method of recruitment strategy comparison	Sample size (N)[1]	Study duration	Recruitment strategies	No. Patients recruited/No. Patients approached or where contact was attempted[2]	Country
Woolhouse 2014 (85)	RCT of mindfulness vs treatment as usual (TAU) in women of depression, anxiety or stress	retrospective	32	6 weeks	a. researcher recruiting at clinic waiting room	14/50	Australia
					b. mailed-out brochures	16/2500	
					c. recruitment via physiotherapy and childbirth education classes	2	
Krusche 2014 (81)	RCT of mindfulness-based CBT vs TAU in preventing relapse in people with recurrent depression conducted in primary care	retrospective	153	8 weeks	a. word of mouth	16/46	UK
					b. information from charity	2/8	
					c. posters	30/123	
					d. web-based adverts	37/300	
					e. mental health care referral	8/32	
					f. radio adverts	26/412	
					g. GP referral	18/116	
					h. bus adverts	2/4	
					i. newspaper adverts	11/101	
					j. exhibition	3/11	
		retrospective	1699	6 weeks	a. Google advertising	755	Australia

[1] For randomised recruitment trials, N = sample size of its host trial. For non-randomised studies, we assume that the sample size is the sum of number of patients recruited via each strategy.
[2] Containing numbers of patients approached if data are available.

Morgan 2013 (82)	RCT of email delivered self-help health promotion intervention for adults with subthreshold depression symptoms to prevent depression (patients were screened online using PHQ-9)				b. Facebook adverts	35	
					c. online forums	unknown[3]	
					d. links from mental health websites	unknown	
					e. online community noticeboards	unknown	
					f. group emails	unknown	
Man 2015 (79)	RCT of a telephone support and computer-based self-management intervention vs. usual care in patients with depression in primary care	RCT	60	12 months	a. optimised written patient information material	43/682	UK
					b. original patient information material	27/682	
Rollman 2008 (83)	RCT of telephone-based collaborative care for treating patients with DSM-IV panic and anxiety disorders	retrospective	369	Not reported	a. EMR reminder to primary care clinicians to approach eligible patients	176/794	US
					b. waiting room recruitment by research staff	193/8095	
Jeste 2009 (80)	Hypothetical RCT of a cognition-enhancing drug vs. placebo in patients	RCT	248	14 weeks	a. multimedia enhanced consent procedure	31/62	US
					b. ordinary consent procedure	29/66	

[3] According to the study, there were a total number of 94,808 approaches made in the study.

	with DSM-IV schizophrenia						
Daley 2007 (84)	RCT of an exercise intervention for women with postnatal depression	retrospective	38	12 weeks	a. recruitment via GP	19/96	UK
					b. recruitment via specialised “mother and baby” unit	9/28	
					c. recruitment by health visitors	7/10	
					d. self-referral	3/4	
Le 2008 (87)	RCT of an antenatal psycho-educational group intervention to prevent postpartum depression in patients with high risk[4]	retrospective	310	8 weeks	a. recruitment by community health centre staff	276/553	US
					b. recruitment by clinical research staff at hospital-based clinic	34/1349	
Debar 2009 (86)	RCT of a CBT-based guided self-help program on patients with DSM-IV Binge Eating Disorder	retrospective	249	not reported	a. mail invitation to comprehensive Eating Disorders Examination (EDE) assessment, $5 incentive for completing online screening questionnaire and $50 for baseline assessment	70/11984	US
					b. mail invitation to abbreviated EDE assessment + telephone interview, $25 for baseline assessment (no payment for screening)	154/20810	

[4] High risk = Epidemiologic Studies Depression Scale (CES-D) ≥ 16; all patients were self-reported.

					c. self-referral	25/87	
Schlernit-zauer 1998 (88)	RCT of nortriptyline and interpersonal psychotherapy in elderly patients (age ≥ 65) with bereavement-related major depression (screened using HAM-D scale).	retrospective	65	Not specified	a. adverts	35/194	US
					b. obituary letter	9/99	
					c. acquaintance/friend	9/54	
					d. outpatient/in-house psychiatric referral	7/47	
					e. non-specific resources	2/20	
					f. non-mental health physicians	3/11	
					g. letters sent to medical community/health professionals	0/7	
					h. inpatient psychiatric referral	0/5	
					i. private mental health practitioner	0/3	
					j. other mental health facilities	0/1	
Hughes-Morley 2016 (89)	EQUIP host trial – clustered RCT of a new user led training package to increase user and carer involvement in care planning for patients with a diagnosis of severe mental illness under	RCT and retrospective[5]	480	30 months	a. leaflet sent to advertise patient and public involvement in research (PPIR)	216/5382	UK
					b. control (without leaflet)	148/2800	
					c. leaflet sent to advertise PPIR + telephone follow up for non-responders	129/4988	
					d. control + telephone follow up for non-responders	92/2580	

[5] Patients who were enrolled during telephone follow up (strategy c & d) were not included in the primary outcome as this was not the intervention for which this trial was designed to find evidence.

	community mental health teams						
Brown 2019	CLASSIC host trial – a brief intervention of self-confidence workshops for depression	retrospective	420	12 months	Self-referral	459/1042	UK

Abbreviations: RCT, randomised controlled trial; TAU, treatment as usual; CBT, cognitive behavioural therapy; GP, general practitioner; DSM, Diagnostic and statistical manual of mental disorders; EMR, electronic medical records HAM-D, Hamilton depression rating scale

2.3.3.2 Randomised comparative studies

Of the included studies, Jeste 2009, Man 2015 and Hughes-Morley 2016 used a randomised approach to compare alternative recruitment strategies. (Table 2-3) Jeste et al. compared a multimedia consent process using a DVD to present key information from the consent form, with routine consent procedure plus a 10 min 'control' DVD giving general information about research. Man et al. used an 'optimised' version of the trial information sheet, with contrasting colour, larger fonts, bulleted lists, and accessible wording, compared with the original 8-page A5 patient information booklet. Hughes-Morley et al. investigated the impact of a strategy of providing a leaflet describing the patient and public involvement (PPI) in the trial on recruitment of people who had a diagnosis of SMIs. Using multimedia during the consent process did not significantly improve recruitment in patients with schizophrenia, whereas optimised written patient information material was superior to non-optimised information for recruitment of patients with depression in primary care, but this result may have occurred by chance. Finally, offering information on PPI collaboration on the trial was not found to have a positive impact on trial recruitment.

Table 2-3 Summary of randomised comparative studies on recruitment strategies

Study ID	Strategy comparison (intervention vs. control)	No. Patients recruited / No. Patients attempted (intervention)	No. Patients recruited / No. Patients attempted (control)	Relative Risk (with 95% CI)
Jeste 2009	DVD multimedia consent with key information from consent form vs. routine consent procedure + 10 min control DVD on general	41/62	44/66	0.9919 (0.7751 to 1.2694, p=0.9487)

	information on the research			
Man 2015	optimised written patient information material vs. original patient information material	43/682	27/682	1.5926 (0.9960 to 2.5467, p=0.0520)
Hughes-Morley 2016	Leaflet invitations sent to advertise PPIR vs. no leaflet invitations	216/5382	148/2800	Odds Ratio =0.75[6]

Abbreviations: DVD, digital video disc; PPIR patient and public involvement research

2.3.3.3 Non-randomised studies

Krusche et al. suggested that recruiting by adverts and posters showed no less efficacy than recruiting from GP referrals. In contrast, a study using electronic health records to remind GPs to approach potentially eligible patients was more efficient than recruitment by research staff in the clinic waiting room. The latter involved considerably more effort (more than 8000 patients were approached) (Rollman 2008). Le *et al.* also suggested that being contacted by clinical staff was more successful than being contacted by research staff. Among trials involving people with common mental illnesss, GP referrals and contact by clinical staff were the most efficient and successful recruitment strategies and both resulted in an adequate number of patients for the size of a modern trial. Financial incentives are commonly used in commercially funded trials. The study done by DeBar suggested that neither different levels of financial incentives nor different lengths of assessment substantially affected recruitment rates. (86)

In an RCT of mindfulness versus treatment as usual (TAU) in women with depression, anxiety or stress, Woolhouse *et al.* used both more active

[6] The study reported ORs and used a random effects logistic regression, which yielded OR = 0.75, 95 CI: 0.53 to 1.07, p=0.013

(researcher approaching patients in clinic waiting room) and less active (invitations sent to potential participants) strategies. (85) The numbers of patients recruited were similar, despite 2,500 mailshots being sent compared with the researcher approaching 50 patients. A study comparing various forms of online recruitment for a preventive intervention for people with subthreshold depressive symptoms (as assessed by an online questionnaire) found that Google adverts recruited the highest number of participants (755 patients recruited). However, it was indicated that a total of 94,808 potential participants were approached, echoing findings by Krusche, suggesting that lower success rates may often be the case in recruitment via online advertisements. (81)

2.3.3.4 Cost effectiveness of recruitment strategies

The results of the average cost per patients of recruitment strategies are reported in Table 2-4. The strategy with the lowest cost per patient recruited was web-based advertisement (£13.41 per patient), followed by recruiting via specialised care (£183.24 per patient), non-web-based adverts (£372.03 per patient) and recruitment via primary care (£407.65 per patient). The sensitivity analysis considered the variation in cost according to the different strategies used. For instance, the cost of recruiting using health care providers depends on how much time is spent on recruitment. The two assumed levels of time commitment were part-time (3 hours/day) versus full-time (7.5 hours/day). Table 2-5 shows the results of the costing and sensitivity analyses, in comparison with the costs reported with original data. As each study reported different information on costing the recruitment, even for similar strategies across different studies, costs obtained from available sources showed considerable variation. Shown below is an example of how recruitment cost was obtained, using a study by Morgan *et al.*

(82). Further details and sensitivity analysis are given in Appendix D Costing and sensitivity analysis for the recruitment strategies in Chapter 2.

Table 2-4 Average costs per patients recruited of different kinds of recruitment strategies across studies

Recruitment strategy	Number of studies where strategy was used	Average cost per randomised participant (in GBP) with original data [7]	Number of times recruiting the most within study
Web-based adverts[8]	2	£13.41	2
Via specialised care	4	£183.24	1
Via secondary care	2	not reported	1
Non-web-based adverts	3	£372.03	1
Financial incentives[9]	1	not reported	1
Via primary care	4	£407.65	2
Others	4	not reported	0

[7] Results account for the average exchange rates GBP/Australian dollar (AUD) and GBP/USD in year 2016, and inflation rates of
the countries of publication from year of publication until 2016. (http://www.ukforex.co.uk/forex-tools/historical-rate-tools/yearly-average-rates; http://ination.stephenmorley.org/; U.S. Internal Revenue Service)

[8] Including Morgan 2013, a study which used a number of different online resources to recruit patients.

[9] DeBar 2009, a study which used different incentives to recruit patients.

Table 2-5 An example of the detailed costing a strategy used in Morgan 2013

Study ID	Recruitment strategy	Description in original text	Resource used for costing	Calculation	Notes	Min. Cost (in GBP)	Max. Cost (in GBP)
Morgan 2013	links from webpage	"A new page of supporters was created to accommodate this requirement. This page thanked each organization or website that had helped promote the study to participants. Some websites were generous and included a link and blurb on their home page; others listed the website within a section of their site that contained links to other interesting websites."	Clicking on the webpage link, assumed cost zero.	assuming from 2hrs/day to full time responsible for mailing and posting, salary Band 7 £38,786. (£52/hr)	recruitment from Feb2010 to March 2011(13 months). However, no information on how many hours dedicated to such strategy.	25,740	38,786

2.3.4 Retention strategies

Table 2-6 summarises two studies identified that compared different strategies to improve postal response in surveys. On joining the trials, participants were randomised to be followed up via different methods, and their response rates at follow-up were compared as a proxy for retention rates using the different methods. (6) (7) McLean *et al*. investigated the effects of pre-notification (e.g. notifying participants in advance that they would be asked for information) and envelope 'teaser' (placement of a short message on the survey envelope) on increasing postal response rates in a bulimia nervosa mental health literacy survey. Dirmaier *et al.* conducted a randomised trial to find out whether small cash incentives and a shortened questionnaire helped increase postal response rates in a mailed follow-up survey one year after inpatient psychotherapeutic treatment for mental health patients. Both studies used a 2×2 factorial design to investigate the impact of strategies on postal response rates. Financial incentives, abridged questionnaire and pre-notification were suggested to be effective to increase postal response rates, but the effects were small.

Table 2-6 A summary of retention strategies

Study ID	Retention strategy	Study period	Numbers approached	Numbers responded	Response rates	Cost information	Relative risk
McLean 2014 (92)	Prenotification (+), envelope teaser (-)	not reported	762	190	25%	$23.68/response	Marginal Prenotification RR = 1.165 (p = 0.027) Marginal Envelope Teaser RR = 0.955 (p = 0.508)
	Prenotification (+), envelope teaser (+)	not reported	747	167	22%	Not reported	
	Prenotification (-), envelope teaser (-)	not reported	750	150	20%	$26.25/response	
	Prenotification (-), envelope teaser (+)	not reported	747	154	21%	Not reported	
Dirmaier 2007 (91)	Financial incentive (+), abridged questionnaire (-)	1 year	832	458	55%	Not reported	Marginal Incentive RR = 1.146 (p < 0.0001) Marginal abridged questionnaire RR = 1.073 (p = 0.021)
	Financial incentive (+), abridged questionnaire (+)	1 year	845	500	59%	Not reported	
	Financial incentive (-), abridged questionnaire (-)	1 year	1045	502	48%	Not reported	
	Financial incentive (-), abridged questionnaire (+)	1 year	1103	569	52%	Not reported	

Abbreviations: RR, risk ratio

2.4 Discussion

The review identified only 3 eligible randomised comparative studies of alternative recruitment strategies in mental health clinical trials. None showed a statistically significant difference between using standard and optimised patient consent and information materials. Our findings were consistent with those of Treweek *et al.* The difference approached significance in one trial of recruitment using optimised patient information material compared with original patient material (Man 2015), although the effect was small. The 8 other studies included in the recruitment section of the review consisted of non-randomised, retrospective comparisons of different recruitment strategies. It is difficult to know whether the different strategies were employed in comparable ways in these studies, or for the same duration. Given the small number of randomised comparative studies identified, and the inconclusive results, this review suggests further research in this area may benefit trial recruitment. Two randomised studies comparing different retention strategies in mental health were identified (McLean 2014, Dirmaier 2007). Both involved different ways of maximising response rates to postal assessments. As follow-up assessment in RCTs is often carried out in the form of a questionnaire, the response rate to this type of assessment may be appropriate as a proxy for retention.

Prior to this review, I also piloted a search strategy that encompassed informed consent, recruitment, antipsychotics and randomised trials, attempting to review recruitment strategies in antipsychotic randomised trials. It generated approximately 2,000 records from MEDLINE, EMBASE, and CMR. However, after screening there was only 1 study (Jeste 2009) which met our criteria. Little attention has been paid to such methodological trials (e.g. using SWATs to

increase the evidence base for trial decision-making) that endeavour to tackle some of the most common issues in mental health clinical trials.

The included studies showed substantial differences in strategies used, but also in clinical settings, mental health conditions and study design. Therefore a pooled estimate of recruitment efficacy of these strategies was difficult to obtained due to the non-randomised designs used, and the choice of analysis which could be used to assess the relative efficacy of different strategies was limited. It is therefore difficult to estimate the efficacy from beyond an individual level. Also, some included studies did not report numbers of potential participants approached by each strategy, e.g. the denominator for the efficiency measure of recruitment strategies (number recruited divided by number approached), and comparison between numerators should be made with caution, as some strategies have broader reach to the population and some studies required larger sample sizes. There were some interesting insights from the result of some recruitment studies, nevertheless. For instance, although clinical staff and GPs are often thought to be helpful in recruiting patients into randomised trials, here it was shown that they recruited no better than advertisements. The comparisons made were *ad hoc*, however, and in the absence of randomised controlled experiments, the area needs more rigorous investigation.

In this review, I also considered costs for each strategy based on numbers of participants recruited and cost incurred. It provided some useful information for public funded trials, which often work on a limited budget. Outcome of interest was average cost per patient recruited/retained (responded). Conventionally a

CEA often adopts an incremental approach, with incremental cost per outcome gain used as outcome. This is often carried out in a randomised design setting to avoid potential bias of the estimates. This approach, had it been adopted, would have had its limitations as majority of the included studies made comparisons in an observational fashion. It should be highlighted that the difference may result in different messages that can inform the funding bodies. For the case of recruitment, the average costs per patient recruitment illustrate nicely of how the recruitment situation will be informed by different recruitment strategies, whereas a CEA that discusses “how much more to pay in order to recruit extra N patients” may appear to be of less meaning as more costs would naturally lead to more patients recruited. Further, if the recruitment outcome is not confined as the number of patients recruited, rather, for instance, a separate dimension on retention may be incorporated, a CEA approach may have its merit in help the decision. In absence of robust estimation on incremental costs per additional patient recruited based on randomised comparison, the average cost per patient recruited measure can provide preliminary estimation based on such naïve comparison upon which further decision should be made. Funders or sponsors should have expectation or approximation of the cost of recruiting patients in the trial design stage, based on patient population, clinical settings etc. – very often this is used to approximate the sample size which is then used in power calculation to see the study specificity and sensitivity. However, it is worth noting that the choice of recruitment strategy should consider not solely cost, but also the study design, types of intervention and more importantly, population characteristics. For instance, I found that although using web-based advertisements showed merit in terms of efficacy and cost in recruitment, however the loss to follow-up in the population recruited via this method cannot

be ignored. It is essential also to consider whether certain recruitment methods may identify a selected and unrepresentative population. I also considered the uncertainty due to the inadequately reported cost information in the included studies, and performed sensitivity analysis of the costs obtained. The lack of a standard and transparent methodological framework for reporting the costs or resource use during recruitment has engendered considerable variations in the analysis and has led to challenges in interpreting the results. For instance, strategies that involved research assistants recruiting in clinic waiting rooms did not specify the total hours spent, therefore it was necessary to make assumptions regarding the numbers of hours spent per day on the recruitment task. Even for studies which employed similar recruitment strategies, reporting on resources used during recruitment varied substantially, leading to considerable differences in costs obtained. Speich *et al.* also found in their systematic review that none of the included studies provided empirical resource use and cost data for all aspects of an RCT, and for trials that reported costs of recruitment, even similar recruitment strategies could cost different amounts across studies. Within a given category of recruitment strategies, for instance, the median cost of a mailed invitation was 228 USD, ranging from 15 to 1,116 USD per patient. (49)

Recently the ORRCA project (Online Resource for Recruitment research in Clinical triAls, www.orrca.org.uk) has attempted to bring together all the studies on recruitment into randomised trials by creating a searchable database. This initiative may help to inform trialists and recruiters of better ways to recruit patients into trials.(93) Also, Madurasinghe *et al.* provided guidelines for reporting embedded recruitment trials, for which a checklist based on the Consolidated Standards for Reporting Trials (CONSORT) statement 2010 was

developed and several examples were listed. (94) Unlike the existing literature, this review has a focus on recruitment via different channels used as strategies described in the included studies, partly because of the inadequacy of the evidence available for mental health trials. It provides some evidence from a different perspective and makes suggestions regarding possible future research in this area. For instance, SWATs may be designed to compare the efficacy of recruitment by research staff with recruitment by clinical staff. Promoting the guidelines by Madurasinghe *et al.* will help to improve the quality of reporting for these methodological trials. Furthermore, it is also worth investigating the performance of different recruitment strategies with respect to other aspects of the trial, such as the population characteristics or adherence to the trial intervention, as these features also can determine a trial's precision and efficiency. Some strategies may recruit a biased sample. For instance, using web-based adverts as a recruitment method in mental health trials may inadvertently recruit the "worried well" or those who do not sufficiently resemble real-world patients.

This study has the following limitations. Firstly, I only identified 3 randomised comparative studies of recruitment and two of retention. The rest compared different strategies without randomisation and this diminishes the internal validity of their findings. Secondly, out of 13 identified studies, the majority were in depression-related illnesses. The limited number of studies involving people with diagnoses of SMIs such as bipolar disorders or schizophrenia, reduces the generalisability of the review. It highlights the need for more research in this area, since there are many challenges to recruitment within this group of people. Moreover, there were no eligible RCTs aimed at improving retention

within RCTs. We included 2 studies which focus on improving postal response rates in follow-up, despite the fact they were not set within a randomised clinical trial. However, since they used a randomised design to assess methods to enhance response rates, we recognised that they contribute useful information, although clearly more studies are needed to address retention issues in randomised trials and in studies that use face to face assessments rather than postal questionnaires. Lastly, lack of reported information on costs in many of the included studies means there is considerable uncertainty in our findings on cost.

The recruitment and retention challenge in RCT are often considered carefully during the design stage of the trial. Discussion on the potential recruitment difficulties and whether the projected sample size of the trial should illuminate the most appropriate design and avoid the chance of failure. Identifying the appropriate group of population and their behavioural characteristics could help investigators and trial operation team with the optimal recruitment strategies to improve recruitment efficiency.

From this systematic review it could be concluded that paying attention to the accessibility of information and consent materials (optimisation) may help improve recruitment. Recruitment by clinical staff and non-web-based adverts showed some efficiency and success in certain circumstances. Pre-notification, abridged questionnaires and financial incentives have small positive effects on retention rates in postal surveys. Moreover, more futures SWAT research may benefit the evidence base for recruitment and retention strategies so that more robust comparison of effectiveness and cost-effectiveness may be made.

In this chapter I discussed the possibilities of improving mental health RCT's efficiency through improving recruitment and retention success, via a systematic review. In the next two chapters, I will look at how previous real world evidence, including clinical trials and meta-analyses, may be used to construct simulated scenarios to mimic trial results using a decision modelling approach. I will review the existing decision modelling on patients with schizophrenia and propose a decision model for the case of RADAR, comparing the cost-effectiveness of the antipsychotic reduction and discontinuation strategy with the maintenance strategy. It has the potential to inform us as to what may be present in the real trial itself.

3 Chapter 3 – A Review of current decision models of antipsychotic medications and patients with schizophrenia

In the previous chapter I discussed how the efficiency of mental health RCTs could be improved by a variety of recruitment and retention strategies. Improving recruitment and retention could ensure efficient design and conduct of an RCT. Another way of improving the efficiency of a trial design is to make use of the previous research evidence. For the case of RADAR trial, using previous knowledge about the benefit and risk of antipsychotic medication, as well as that of reducing antipsychotics, could be helpful in understanding what may happen in the trial. In Chapters 3 and 4 I develop and discuss a decision model to compare the effectiveness and cost-effectiveness of the antipsychotic reduction and discontinuation strategy with the maintenance strategy, using previous available research evidence. The aim of this chapter is to review the existing published decision models for patients with schizophrenia and discuss their design characteristics, which then lead to inform the design of the decision model itself described in Chapter 4.

3.1 Introduction

The costs of schizophrenia are manifold and not limited to healthcare only. Its enduring impact on patients' health, and on their families and society can result in substantial financial burden.(95) It is estimated that the annual health service costs for schizophrenic disorders in 2019 will reach £3 billion, whereas the projected societal cost will reach £5.5 billion. (96) Medication cost makes up to 18% of the service cost. In this regard, reducing the use of antipsychotic medications in patients with schizophrenia, as suggested by the RADAR trial, will not only have a chance of improving quality of life, but also may help bring

down the cost of service use by reducing antipsychotic prescriptions, or even other service and societal costs if proved a success.

Trial-based economic evaluation is a useful tool in assessing the cost-effectiveness of RADAR's strategy. As a chronic condition with recurring relapse events, schizophrenia and the cost-effectiveness of RADAR's strategy need to be studied in the long run. Hence a within trial CEA in the RADAR trial alone may fall short of sufficient evidence to demonstrate the cost-effectiveness of RADAR's strategy from both the NHS's and public sector's perspectives. Economic evaluation with decision modelling techniques provides a framework to inform decision making under different uncertainties in the long run, and has been increasingly used to inform clinical decisions at both population and individual levels. (60)

Frequently used decision models include decision tree, Markov model, patient level simulation, DES (discrete event simulations) and so on. In the case of schizophrenia, some of these techniques have been used in modelling the cost-effectiveness of pharmacological or complex interventions. The choice of modelling technique does not only depend on the decision problem, but also on the availability of the data. (97) In this chapter I will review the existing published decision analytic models for a range of interventions for patients with schizophrenia in terms of their design characteristics and input parameters. Reviewing these pre-existing models may provide insights into modelling the cost-effectiveness of RADAR's strategy in terms of what may be more important to capture in the model for the case of RADAR. It is anticipated that the potential risks of antipsychotic discontinuation may include an increased chance

of relapse and hospitalisation, whereas the benefit could be that it may reduce the risk of developing antipsychotic-related adverse events such as EPS, diabetes, or CV events. Hence reviewing the literature, particularly for these aspects of the treatment, may provide useful information to the model development.

3.2 Methods

MEDLINE and the Centre for Reviews and Dissemination assessed economic evaluation database (CRD, NHS EED, University of York, last update 2014) were searched using a comprehensive search method in September 2017, combining mental health key words and economic analysis and modelling related key words. Detailed search strategy is provided in Appendix E. I searched decision analytic models for economic evaluation of various pharmacological therapies, or treatment management, in patients with schizophrenia, schizo-affective disorders, or other non-affective psychotic disorders.

An updated search of the MEDLINE database was carried out in April 2020, in addition to a full search of a new database, PsycINFO. I applied the same screening criteria process. The newly identified eligible studies are then appended into the original review database.

3.2.1 Inclusion criteria

- Decision models including decision tree, Markov model, DES, micro-simulation, Monte Carlo simulation or dynamic modelling on cost effectiveness /cost utility of pharmacological treatment options;

- Patients with schizophrenia, schizo-affective disorders, or other non-affective psychotic disorders;
- Publications in English.

3.2.2 Exclusion criteria

- Studies in patients with bipolar or other affective disorders;
- Studies on cost-effectiveness/cost-utility analysis but without decision modelling approach;
- Reviews, comments, editorials, letters, case reports or studies in animals;
- Statistical models developed from observational or epidemiological data to predict risk equations;
- Budget impact models or cost of illness (COI) studies.

3.3 Results

3.3.1 Result of search

A total number of 6734 records were identified from MEDLINE by a comprehensive search method. 267 records were identified from Centre for Reviews and Dissemination NHS EED using a similar strategy customised accordingly, with key words including “psychosis”, “economic evaluation”, “antipsychotic”, “schizophrenia spectrum and other psychotic disorders”. I conducted title and abstract screening separately in 2 databases and removed duplicates. 400 records were eligible for full text screening. Common reasons for excluding include analysis without decision modelling structure, or unsuitable population. 32 studies were included in the final review. A brief description of the screening process is given below in Figure 3-1.

An update search of the MEDLINE and a new search in PsycINFO database were performed in April 2020. In the MEDLINE search update, 54 studies were identified. After removing duplicates from the original MEDLINE search database, there were 17 studies left but no eligible studies were included. From the PsycINFO database search, 5284 studies were identified through the initial search, 3863 studies were left after removing duplicates with MEDLINE original search database and update. After title and abstract screening, 13 studies were left for full-text assessment. 2 studies were eligible to include for final review. Overall, the update identified 2 additional studies in the new PsycINFO database, while the updated search in MEDLINE did not identify any additional studies from the date of the initial search. A flow chart of the search update is illustrated in Figure 3-2.

Quality of the included studies were accessed using the Philips checklist. (98) The checklist consists of a total of 59 questions, covering assessment from structure, data and consistency perspectives of the decision model. The percentage of “Yes” answer to questions were recorded as an overall measure of the study quality.

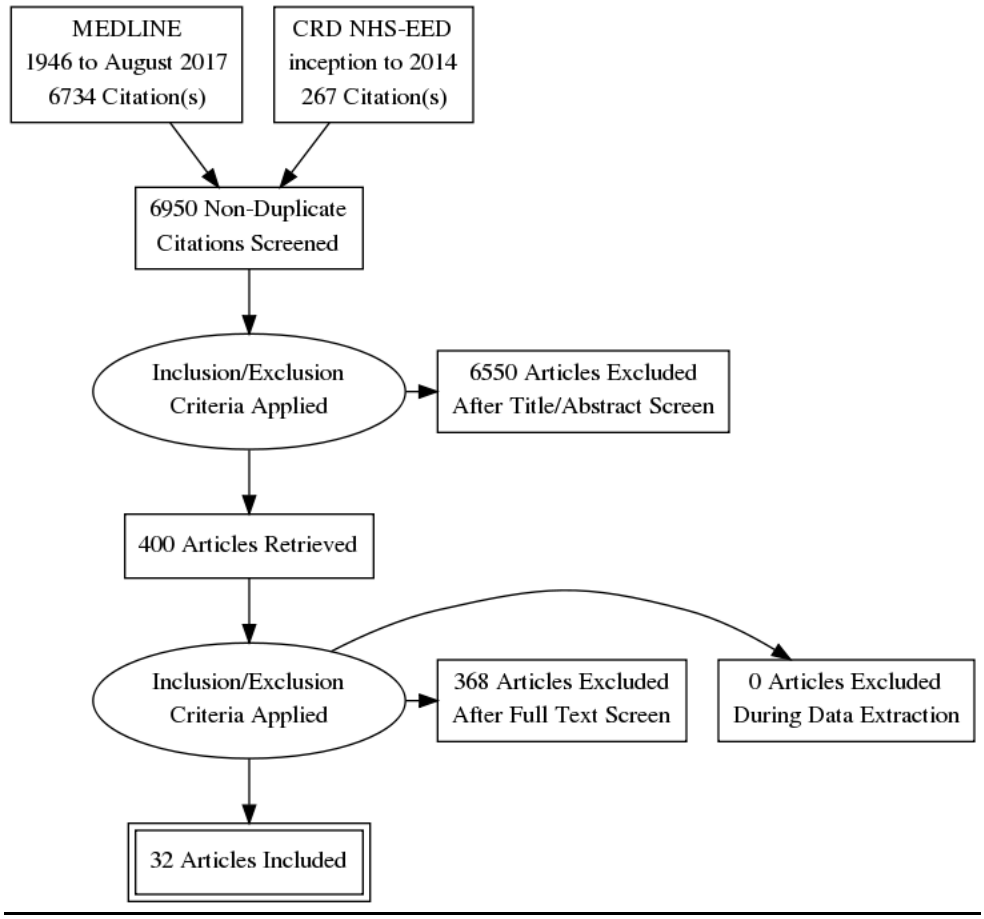

Figure 3-1 A flow chart of the literature screening process

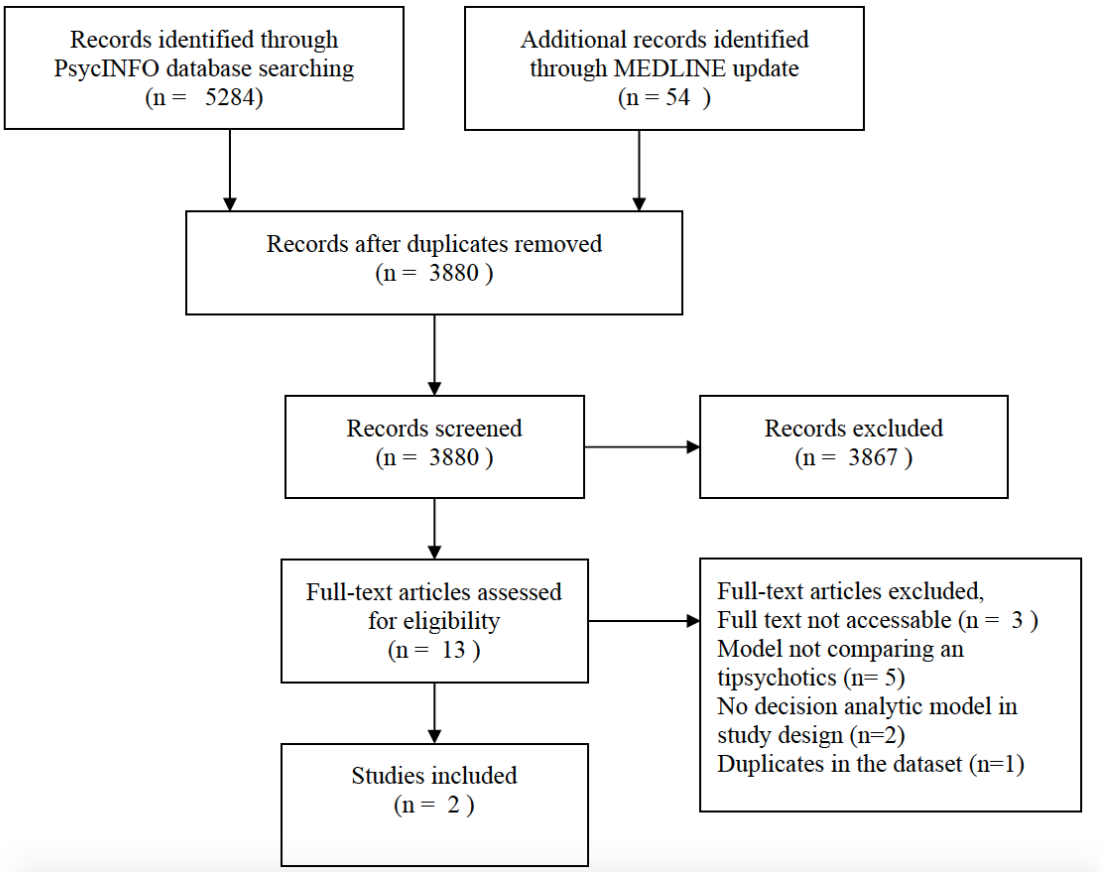

Figure 3-2 A flowchart of the literature review update screening process

3.3.2 Characteristics of included studies

The overall 34 included studies used various modelling techniques, Markov models were used most often (n=15), followed by decision tree (n=10). Two studies used a combination of Markov model and decision tree approach to model the decision process at different stages. Three studies used DES and 5 studies used other miscellaneous model techniques, including patient level simulation, semi-Markov model and micro-simulation. The table below summarises the key characteristics of the included studies including country, model types, time horizon, main outcome measures and whether some of the major adverse effects or comorbidities, such as relapse, cardiovascular events, were considered.

Table 3-1 A brief summary of the included studies

Author	Year	Country	Design	Relapse (Y/N)	Relapse evidence	Adherence	Metabolic adverse effects	Cardiovascular adverse effects	Suicide	EPS or others	Time horizon (year)	Cycle length in months	Measure of effectiveness	No. of Yes answers in the Philips checklist (%)
Einarson (99)	2017	Netherlands / Spain	Decision tree	Y	RCT	N	N	N	Y	N	1		QALY	57.89
Phanthuane (100)	2011	Thailand	Decision tree	N		N	Y	N	Y	N	0		DALY[10]	78.95
Furiak (101)	2012	US	Markov model	Y	Previous model	Y	N	N	N	N	5	1	QALY	68.42
Rajagopalan (102)	2013	US	Markov model	Y	RCT	Y	Y	Y	N	N	5	12	relapse avoided	73.68
Davies (103)	2008	UK	Markov model	Y	RCT	Y	Y	N	N	Y	10	4	QALY	73.68
Bounthavong (104)	2007	US	Decision tree	Y	RCT	N	N	N	N	N	0		Efficacy rate (% responders)	66.67
Bobes (105)	2004	Spain	Micro-simulation	N		Y	Y	N	N	y	1		month with psychotic symptoms controlled	64.91
Garcia-Ruiz (106)	2012	Spain	Decision tree	Y	Systematic review	Y	Y	Y	Y	Y	1		QALY	66.67
Frey (107)	2014	Germany	Markov model	N		N	N	N	N	N	10	1	per hospital day avoided	77.19
Beard (108)	2006	Germany	Markov model + decision tree	Y	previous model	N	N	N	N	N	3		QALY + avoided relapse	68.42
Wang (109)	2004	US	Markov model	Y	Systematic review	N	N	N	N	Y	life-time	3	QALY	78.95
Graham (110)	2012	US	Decision tree	Y	RCT	N	Y	Y	Y	Y	0		QALY	70.18

[10] DALY: disability-adjusted life year

Vera-Llonch (111)	2004	US	Markov model	Y	RCT	Y	Y	N	N	Y	1	1	relapse, AE and switching	63.16
Palmer (112)	2002	US / Mexico	Markov model	Y	Case studies	N	N	N	N	Y	5	3	BPRS	70.18
Almond & O'Donnel (113)	1998	UK	Markov model	Y	RCT	Y	N	N	Y	N	5	3	BPRS & non-relapse	45.61
Heeg (114)	2008	UK	DES	Y	RCT	Y	Y	N	N	Y	5		QALY	59.65
Park (115)	2014	US	Markov model	Y	RCT	Y	Y	Y	N	Y	5	4	QALY	78.95
Furiak (101)	2011	US	Micro-simulation	Y	RCT	Y	Y	N	Y	Y	1		QALY	73.68
O'Day (116)	2013	US	Markov model	Y	RCT	Y	Y	Y	N	N	5	12	relapse-related hospitalisation avoided	73.68
Laux (117)	2005	Germany	DES	Y	RCT	Y	Y	N	N	Y	5		QALY + avoided relapse	77.19
Einarson (118)	2013	Czech Republic / Finland / Sweden	Decision tree	Y	Observational	Y	N	N	N	N	1		QALY	68.42
Lachaine (119)	2014	Canada	Markov model + decision tree	N		N	Y	Y	Y	Y	6		QALY	77.19
Dilla (120)	2014	Spain	DES	Y	Observational	N	Y	N	Y	Y	5		QALY	80.7
Kasteng (121)	2011	Sweden	Markov model	N		N	Y	Y	N	N	Life time	12	QALY	85.96
McIntyre (122)	2009	US	Semi-Markov model	Y	RCT	N	Y	Y	N	Y	5		QALY	75.44
Lindner (123)	2009	Brazil	Markov model	Y	previous model	Y	N	N	Y	N	5	3	QALY	64.91
Geitona (124)	2008	Greece	Decision Tree	N		Y	Y	N	N	Y	1		numbers of non-relapse days	52.63
Obradovic (125)	2007	Slovenia	Decision tree	Y	Observational	Y	Y	N	N	Y	1		proportion of patient in remission	50.88

Yang (126)	2005	Taiwan	Decision tree	N		N	N	N	N	Y	2		response	64.91
Tilden (127)	2002	UK	Markov model	Y	Observational	Y	N	N	Y	Y	5	3		63.16
Lecomte (128)	2000	Belgium	Semi-Markov model	N		Y	Y	N	Y	Y	1		time with minimum symptoms and minimum toxicity	59.65
Colombo (129)	2008	Italy	Patient level simulation	N		Y	Y	N	N	N	5		QALY	64.91
Mehnert (130)	2012	Sweden	Markov model	Y	RCT	Y	Y	N	N	Y	5	1	QALY	84.21
Yang (126)	2009	China	Decision tree	Y	Expert opinion	N	N	N	N	Y	2		No. patient treated	36.84

Abbreviations: RCT, randomised controlled trials; DES, discrete event simulation; QALY, quality-adjusted life years; DALY, disability-adjusted life years; BPRS, brief psychiatric rating scale; EPS, extrapyramidal symptom

Table 3-1 summarises the included studies. Twelve models were developed for the US population, 4 each for the UK and Spanish populations and 3 for German population. Seven models were developed for other European populations and 5 for the rest of the world (Canada, Brazil, Thailand and China,. Figure 3-3). Some models were designed and then adapted to other countries .

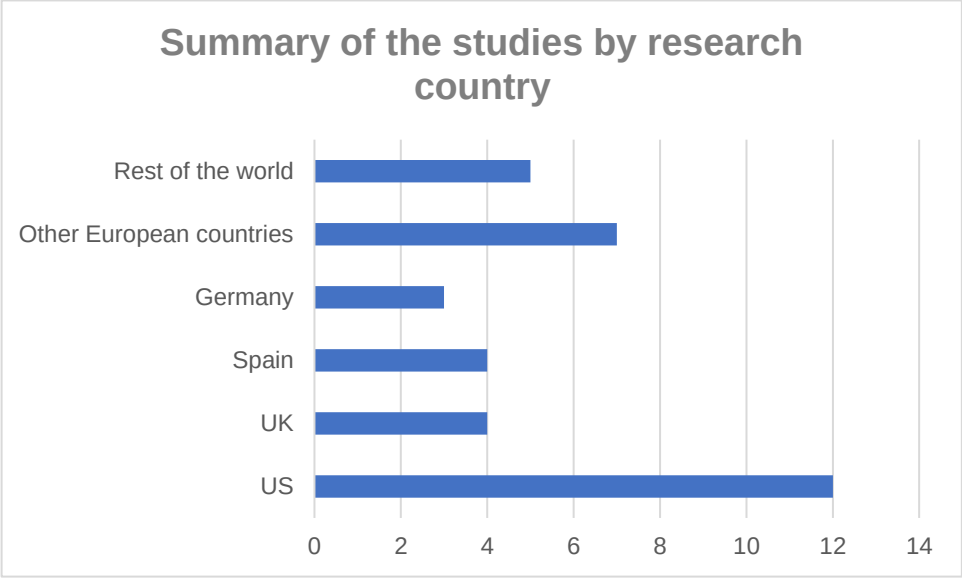

Figure 3-3 A summary of analytic models by country

The majority of the studies investigated the cost-effectiveness of different antipsychotics within different steps of the treatment delivery pathway, such as whether to give a drug in first- or second-line; typical versus atypical antipsychotics; and oral form versus long-acting injection intervention. Other studies have also focused on the common issues in schizophrenia patients, including suicide, adherence with treatment and criminal behaviours.

A wide range of clinical parameters was covered, depending on the research interest and clinical question each decision model addressed. Relapse was included extensively across the included studies (n=25), with evidence collected mainly from related RCTs (n=14), systematic reviews (n=2), observational

studies (n=4) or other sources. Adverse effects such as EPS symptoms (n=19), diabetes and weight gain (n=20) were modelled in studies that compared cost effectiveness of different generations of antipsychotics. QALY was considered in 19 studies as outcome measure for effectiveness whereas other outcome measures such as the numbers of relapse cases prevented, Brief Psychiatric Rating Scale (BPRS), the numbers of non-hospitalisation days, and disability adjusted life years (DALYs) were also used.

There was a clear difference in the amount of decision models published before 2015 and after 2015, with only 1 published in 2017 (Figure 3-4).On average, the included studies had 68% yes answers to the questions from the Philips checklist. 2 studies had under 50% yes answers, which is indicative of lower study quality.

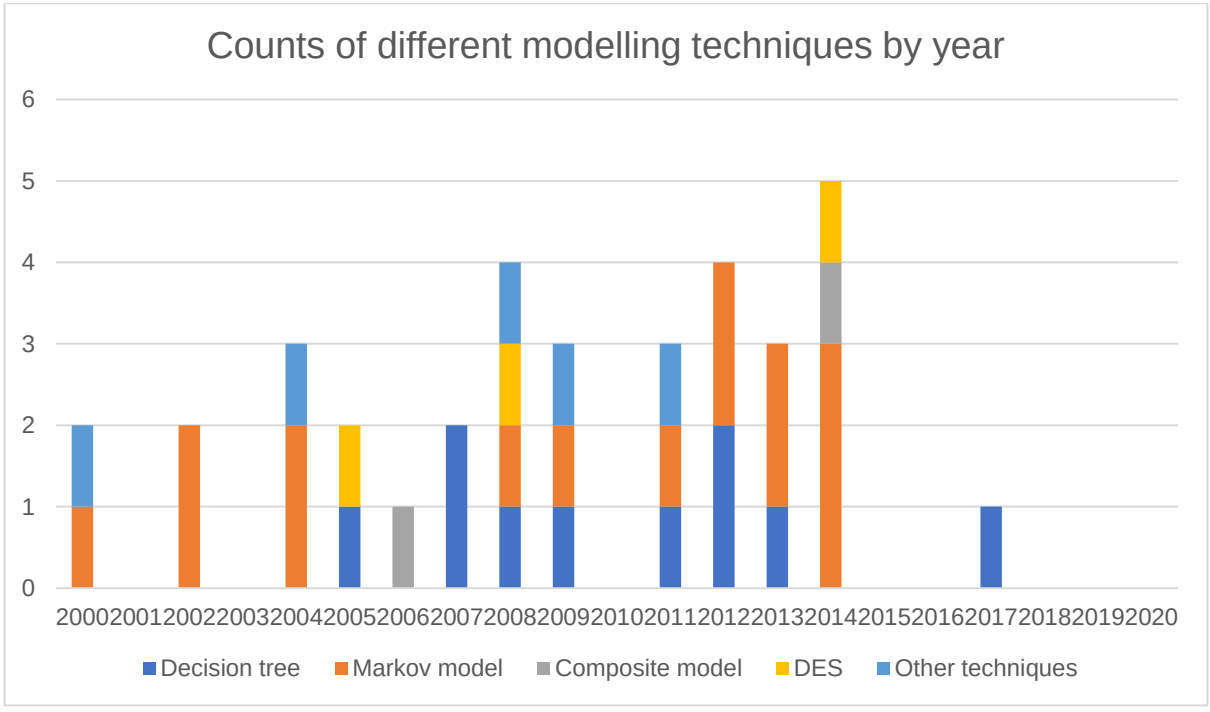

Figure 3-4 Model design count by year

3.3.3 Results regarding design characteristics of the included studies

3.3.3.1 Decision tree design

The decision tree is often considered the simplest and yet most widely used form of decision model. It uses a tree-like model of decision and their possible consequences via a series of mutually exclusive pathways. (see 1.6.1) Amongst the included studies, ten employed a decision tree design. One of the characteristics of this design is that it is not as convenient to capture long term effects of cost-effectiveness of interventions being assessed as state transition model design. Moreover, the time horizon is usually within a year, without discounting. The decision tree design is often suitable for short term decision problems, for instance, adverse effects due to antipsychotic treatment initiation in antipsychotic-naïve patients or treatment switch due to non-response are often modelled in a decision tree. (131)

3.3.3.2 Markov and related models

State-transition models, also known as Markov models, consist of a set of mutually exclusive health states which are evaluated at regular intervals. The flow of patients through the model over time depends on the transition matrixes which include the probability of transferring between each state.(see 1.6.2) Markov models are particularly helpful in modelling a longer time horizon compared with decision tree, which is often used to model short term effects where timing is considered not as important. The discrete time approach requires evaluation at fixed time-points determined by the cycle length and time horizon of the model. This only provides an approximation to the real world case as clinical events can occur at any point. Furthermore, the “Markovian assumption” of state transition models makes it challenging to evaluate scenarios in which the future state transitions depend on previous events. Although this problem can be solved by introducing “tunnel states” to record the

previous event, it may ultimately become difficult to manage as there is always a limit on the number of tunnel states that could be incorporated. (132)

The majority of the included studies employed a Markov state transition model for the economic evaluation (n=15). Commonly used states include relapse (n=8), dead (n=5), stable (n=7) and comorbidities such as diabetes (n=3). Cycle length and time horizon are the main design characteristics of a Markov model. Amongst the included studies which used a Markov design, the 5-year time horizon is the most frequently used design (n=9). Cycle length varies amongst different studies (see Table 3-2). In general, choices included approximately 1 month, 3 months, 4 months and 12 months of cycle length. The difference in cycle lengths depends on the clinical contexts and how often the event of interest that has the smallest time interval occurs. Relapse, or psychosis were usually included as an important state and could be decisive to the cycle length, although the definitions varied between studies, resulting in different cycle lengths. For instance, non-hospitalised relapse tends to have different time intervals to hospitalised relapse, which is often considered more severe. Other events could be decisive to the cycle length, too. For instance, Rajagopalan *et al* looked at long-term costs and outcomes of lurasidone and aripiprazole amongst adults with schizophrenia who previously failed at least 1 atypical antipsychotic. The model described had an annual cycle to assess the chance that patients discontinue either drugs for any cause and switch to clozapine at each cycle. (102) The trade-off between longer and shorter cycle length lies in the balance between the error due to the fixed-time assumption of the transition probabilities and the computational burden. (133)

The choice of key parameter inputs depends on the clinical context of the research question. Studies that compare the cost-effectiveness of different antipsychotics generally take into account a series of adverse effects, and input parameters are obtained from the literature. More specific research questions such as one in Garcia-Luiz *et al.*, which investigated the cost-effectiveness of different antipsychotics in reducing schizophrenia relapses, would require more specific clinical events, for example relapse related parameters. (106)

Table 3-2 An overview of the design features of the included Markov models

Time horizon	Count	Cycle length	Count
1 year	1	**1 month**	5
3 years	1	**3 months**	5
5 years	9	**4 months (18 weeks)**	2
10 years or more	4	**12 months**	3

3.3.3.3 Other models

Apart from Markov models and decision trees, DES is another design that is suggested as appropriate in modelling schizophrenia because of its flexibility in handling patient heterogeneity where there is limited evidence for long-term follow-up research. DES progresses through the times at which events happen to individuals based upon samples from discrete or continuous distributions, allowing event to occur at any time point. (134)

A combination of Markov model and decision tree was also included for studies that addressed more complex research questions. For instance, in the study by Lachaine *et al.* both antipsychotic switching and long term comorbidity were addressed, hence a joined model of decision tree (for the beginning 52 weeks) and Markov model (for subsequent 4 years) was adopted, which appropriately reflected the decision problem specific to this study. (119)

3.3.4 Outcomes

The majority of the included studies assessed the cost-effectiveness of different pharmacological interventions, and 65% (22/34) of the studies used utility instead of related clinical measurements as outcomes. This is due the property of utility (e.g. generalisability) which allow the comparison of health-related effect across different adverse effects and comorbidities as well as across different disease areas. Other relevant clinical outcomes used included BPRS, Positive and Negative Syndrome Scale (PANSS) or response rate. Outcomes that represent relapse were widely used amongst the studies, including cost per relapse avoided, cost per month with non-relapse or proportion of patients in remission.

Amongst the included studies, 22 studies used utility measure as outcome for effectiveness. More frequently used results were taken from study by Lenert *et al.,* Briggs *et al.* or Revicki *et al.* The utility scores used in the literature for schizophrenia in a remission, or a relatively stable state, range from 0.61 to 0.89. Variation in the utility values largely depends the valuation methodology and in some cases, adherence with the standard antipsychotic treatment. On the other hand, the utility values of symptom exacerbation, in many cases described as relapse, showed more variation in the utility values. Definitions of symptom exacerbation were considered based on clinical measurement of symptoms, so that different utility scores can reflect the nature of different states of health for patients. Both utility values and utility deduction (disutility) were used for describing the deviance in utility from that for remission. Table 3-3 and Table 3-4 list the sources of the utility and disutility values that were used for

different health states amongst the studies included in the discussion in this chapter.

Table 3-3 Utility values used for schizophrenia related health states, adverse effects and comorbidities from the included studies

Study ID	Value (SE)	Reference	Method	Notes
Schizophrenia (remission or stable)				
Einarson 2017	0.70 (0.75)	Osborne 2012	Time trade off (TTO)	
Furiak 2012 **Garcia-Luiz 2012** **Graham 2012** **Furiak 2011** **Heeg 2008**	0.88	Lenert 2004	Standard Gamble(SG)	Adherent
Furiak 2012 **Graham 2012** **Furiak 2011** **Lachaine 2014**	0.75	Lenert 2004	SG	Non-adherent
Davies 2008 **Park 2014**	0.856 (0.021)	Briggs 2008	TTO	
Beard 2006 **Wang 2004**	0.83	Revicki 1996	SG	Maintenance
Laux 2005	0.61 (0.069)	Chouinard & Albright 1997	SG	
Einarson 2013	0.89	Cummins 1998; Lenert 2004; Briggs 2008; Oh 2001 & Revicki 1996	N/A	Average of the values from 5 studies
Dilla 2014	0.77 (0.12)	Haro 2006	Unclear	
Kasten 2011	0.73	Lönestrukturstatistik 2007	Unclear	
McIntyre 2009	0.75	Revicki 1996	SG	
Schizophrenia (relapse)				
Einarson 2017	0.485	Osborne 2012	TTO	
Einarson 2017	0.27	Osborne 2012	TTO	Hospitalisation
Furiak 2012 **Furiak 2011**	0.74	Expert opinion	N/A	Adherent
Furiak 2012 **Furiak 2011**	0.53	Expert opinion	N/A	Adherent, hospitalisation
Furiak 2012 **Furiak 2011**	0.63	Expert opinion	N/A	Non-adherent
Furiak 2012 **Furiak 2011**	0.42	Expert opinion	N/A	Non-adherent, hospitalisation

Garcia-Luiz 2012	0.67	Lenert 2004	SG	
Beard 2006 **Wang 2004**	0.56	Revicki 1996	SG	Hospitalisation
Graham 2012	0.57	Lenert 2004	SG	
Laux 2005	0.36 (0.073)	Chouinard & Albright 1997	SG	Moderate
Laux 2005	0.29 (0.071)	Chouinard & Albright 1997	SG	Severe
Einarson 2013	0.659	Cummins 1998; Lenert 2004 & Briggs 2008	N/A	Average of the 3; exacerbation
Einarson 2013	0.49	Oh 2001 & Revicki 1996	N/A	Average of the 2; hospitalisation
Diabetes				
Garcia-Luiz 2012	0.76	Cases 2003	Unclear	
Furiak 2011	0.7095	Lenert 2004	SG	88.8% of the remission utility from Lenert 2004
Heeg 2008	0.7128	Landy 2002	SG	81% of the remission utility from Lenert 2004
EPS				
Garcia-Luiz 2012	0.7095	Lenert 2004	SG	88.8% of the remission utility from Lenert 2004
Furiak 2012	0.7095	Lenert 2005	SG	88.8% of the remission utility from Lenert 2005
Wang 2004	0.69	Glennie 1997	Unclear	Tardive Dysknesia

Abbreviations: SE, standard error; TTO, time trade off; SG, standard gamble

Table 3-4 Disutility values used for schizophrenia related health states, adverse effects and comorbidities from the included studies

Study ID	Value (SE)	Reference	Method	Notes
Schizophrenia (relapse)				
Davies 2008 **Park 2014**	0.358 (0.025)	Briggs 2008	TTO	
Diabetes				
Dilla 2014	0.18 (0.03)	Haro 2006	Unclear	
McIntyre 2009	0.19	Revicki 1996	SG	
Davies 2008 **Park 2014** **Lachaine 2014**	0.151 (0.019)	Briggs 2008		

McIntyre 2009	0.06/0.05	Schultz 2003	Unclear	Male/Female
Kasteng 2011	0.02	Konig 2009	TTO	
EPS				
Davies 2008 **Park 2014**	0.256 (0.022)	Briggs 2008	TTO	
Lachaine 2014 **McIntyre 2009** **Graham 2012**	0.074 (0.011)	Lenert 2004	SG	
Dilla 2014	0.054 (0.014)	Hao 2006	Unclear	

Abbreviations: SE, standard error; TTO, time trade off; SG, standard gamble

3.4 Discussion

Research has been published on using decision modelling to assess the cost and effectiveness (not necessarily cost-effectiveness) of the pharmacological treatment options for schizophrenia patients. In a recently published systematic review that examined the relationships between modelling techniques and reported outcomes, nearly 50% of the paired antipsychotic treatment comparisons showed contradictory results for different methodological approaches, whereas the other half showed consistent results regardless of the choice of modelling techniques. (97) Discrepancies may be explained not only by the choice of the modelling techniques, but input data source also played an important part in deciding the results of the analysis.

In this chapter, most of the included studies took relapse into account, not only because recurring relapse is a major event in patients with schizophrenia, it also may incur significant extra cost to the health care system. (135) As for the case of RADAR, patients who are allocated to the antipsychotic reduction intervention may be faced with increasing risk of relapse. It is therefore essential to model relapse, as well as the management of relapse in RADAR, which could become a substantial part of the cost-effectiveness of the reduction strategy.

Apart from health care services, the costs of schizophrenia also include a range of societal services, most of which are indirect costs. Loss of productivity due to absenteeism, presentism, unemployment or incarceration are also important components in the costs of schizophrenia. However the based on the Philips checklist, majority of the included studies adopted a payer's perspective (public or third party). Many schizophrenia patients are economically inactive, and unemployment rate is much higher than general population. (96) Therefore, there is also a significant part of indirect cost relating to patients with schizophrenia due to unemployment. Recent evidence showed that people with SMI did not show significant increases in job retention after attending an employment educational programme that aimed to help them return to work. (136)

The rate of criminal activity amongst patients with schizophrenia is higher than in the general population. (137) The costs in prison related to schizophrenia is manifested in the increased psychiatric consulting among patients who are in prison, and the transfer of them into secure hospital setting. The costs to the criminal justice system (CJS) may also reflect a major part of schizophrenia's costs to the society. Lin *et al.* designed a 3-year Markov model to estimate the cost burden of psychiatric relapse and recidivism for schizophrenia patients who were released from prison recently. By increasing 20% of the released patients to take antipsychotics to prevent relapse could save up to 1.8 million USD over the 3 years. (138) Hence, a public sector perspective, particularly considering the costs and burdens to the CJS, would be meaningful for when evaluating the management of schizophrenia. When costs and outcomes fall on more than one sector, there has not been systematic considerations into what it should

include and how it should be included practically. The extended impact inventory framework provided some guidance on this by emphasising the importance of a disaggregated presentation of costs, effects and opportunity costs by different dimensions to be considered in the scope. (139) To define social values under an explicit social welfare function defined across individuals and dimensions necessiates an aggregated methods and defined set of dimentions agreed in advance. Analysis with explicit value judgement can then help to inform deliberations between decisions made for different dimensions. The difficulty lies in achieving consensus in the dimensions that are considered socially valuable and their relative values, which often see conflicts contradition between each other.

Using the Philips checklist to assess the quality of these included decision analytic model may give a general view of the study quality. Although it is worth noting that the appropriate study methodology may evolve over time as the technology develops. Cohort Markov model and decision tree in Excel or TreeAge were relatively prevalent between 2000 and 2010. As computation ability rapidly grew in recent years, more complex models could be designed to model the treatment management and disease progression, which benefit from the use of individual patient level data.

In deciding on the modelling technique for estimating the cost-effectiveness of RADAR's strategy, a few key factors must be taken into account. First of all, relapse is pivotal throughout the trial and patients may have multiple recurring relapse episodes. It is therefore important to model the management of relapse accurately. The nature of recurring relapse suggests a state-transition model

consisting of live states of stable and relapsed, therefore a cohort Markov model might be a sensible choice. Wang *et al.* described a similar design using 4 different states to discuss whether clozapine might be used as first-line treatment for schizophrenia, given the suggested risk of developing adverse effects (agranulocytosis on clozapine, compared with tardive dyskinesia on conventional antipsychotics) and benefit of controlling symptoms and preventing relapse from the literature. (109) The assumption of a cohort Markov model necessitates a linear relationship between patient demographics, such as age, and model outcomes (e.g. QALYs). If non-linear relationship exists, then simply taking average characteristics for a cohort of patients will provide biased estimation of the outcome of interest, in which case, a patient level simulation may be more appropriate for the research question.(140) As majority of the studies included adopted a cohort approach and have either made assumptions on the average age of the cohort or made no mention of the population. It is difficult to investigate whether the linearity assumption holds. The quality of the reporting of these included CEA showed variation and it was not always transparent how some of the fundamental decision and assumptions were made from the text of the published paper. Transparent reporting, particularly in CEA studies can considerably improve research quality and ensure reproducible analysis. Through scrutiny such as the Philip checklist could be helpful in pointing out the caveats that may lie behind a CEA. In the next chapter I will examine the relationships between population and outcomes in order to validate the linear assumption as well as the feasibility of a Markov Model design.

The included studies showed a variety of methodological approaches, of which Markov model was predominantly adopted in decision modelling in

schizophrenia. The cycle lengths were based on the event of interest with the shortest time intervals. Previous reviews also suggested that more complex structures and patient-level models are expected for future modelling exercise.(141)

In this chapter I reviewed the existing published decision models for patients with schizophrenia, as well as their design characteristics. In the next chapter I will describe and discuss the decision model which compares the effectiveness and cost-effectiveness of antipsychotic reduction and discontinuation strategy to the antipsychotic maintenance strategy, based on some of the findings in this chapter.

4 Chapter 4 – A simulated patient-level Markov model to assess the cost effectiveness of antipsychotic reduction strategy

In Chapter 3, I reviewed the previously published decision models for patients with schizophrenia. The design characteristics of the reviewed studies, in terms of the clinical context of the population, could be informative to the decision model I will discuss in this chapter. I describe and discuss a simulated patient-level Markov model to assess the effectiveness and cost-effectiveness of the antipsychotic reduction and discontinuation strategy. I use previously published evidence available to suggest what may occur in the RADAR trial and potentially the advantage and disadvantage of the intervention compared to the maintenance strategy.

The aims of this chapter, are a) to explore what is likely to occur in the RADAR trial, based on previous published evidence, using a patient level simulation approach; b) to investigate whether the RADAR intervention is likely to be cost-effective in an early cost-effectiveness decision model.

4.1 Introduction

Decisions on HTA of therapeutic strategies should be made on the balance between the expected health benefits and the estimated cost of each pharmacological or interventional strategy. (142) Economic evaluation compares the costs and consequences of strategies and ensures the maximum health benefit is gained from finite resources, for which cost-effectiveness is widely used as one of the aspects in NICE for consideration of recommendation. (8) This requires making use of the best available evidence of both clinical measures and costs, as well as incorporating different levels of

uncertainty. For randomised trials where sufficient information for efficacy, effectiveness and costs can be collected (well designed and conducted), trial-based CEA may suffice as a piece of evidence for cost-effectiveness. Where there is insufficient information on either the health benefit, or cost-effectiveness needs to be estimated over a longer time horizon, a decision-analytic modelling approach may be more appropriate to assess the long-term impact of the strategy.

In previous chapters, I introduced a series of frequently used decision analytic approaches for HTA. For the case of schizophrenia, a chronic condition manifested by recurring relapses, the choice of modelling techniques depends on not only the imminent decision question, but also the highly heterogeneous nature of the population, and the availability of data source. Hence, a flexible approach is preferable in modelling schizophrenia, favouring patient level simulation or DES model (143). In Chapter 3, the review suggested that previous studies have extensively focused on comparing cost-effectiveness of different antipsychotics, covering adverse effects such as EPS and cardiovascular events. The effects of reducing antipsychotic dosages, however, have not yet been thoroughly discussed and antipsychotic reduction still remains under debate as to the potential benefits and disadvantages (18, 144).

It has been suggested that repeated encounters with the CJS are more frequent amongst people with serious mental illnesses (138). CJS therefore inevitably bears significant costs of providing mental health services for people who are imprisoned (145). The average annual cost of a prison place in 2016-17 was £38,042 (146). Psychiatric relapse amongst patients with schizophrenia can

lead to substantial cost burden from a societal perspective, which may result in a huge cost contribution to antipsychotic reduction intervention due to the likely increasing relapse risk, driving the societal cost higher (138).

The RADAR study aims to provide information on the safety, efficacy, effectiveness and cost-effectiveness of antipsychotic reduction and discontinuation strategy, by means of a RCT with by now the largest sample size (n=402) of trials which investigated a similar intervention. To support this intervention into NICE recommendation, exploration on potential costs and benefits, from both societal and clinical perspective, will also be important. As also suggested by Németh *et al*. , as well as in the previous chapter, the majority of models adopted a payer's perspective, while the public sector perspective of long-term non-communicable mental health condition is not negligible. (147)

In this chapter I will describe and discuss a decision analytic model for an antipsychotic reduction and discontinuation strategy, and the potential impact of this strategy in terms of both public sector and healthcare costs and benefits. The proposed decision model does not only take into account the previous questions, but also addresses some issues from a public sector perspective, e.g. including costs to the English CJS. Public sector perspective includes both costs for the NHS and CJS.

4.2 Methods

4.2.1 Study design

I developed a simulated patient level Markov state transition model to assess the long term health service and CJS cost-effectiveness of antipsychotic reduction and continuation strategy. The decision model has 3 states – stable,

relapsed and dead as the capture state. (60) Each individual enters the model at the stable state and transfers to the relapsed state if symptom exacerbation requiring hospitalisation occurs. This assumption is mainly for the decision model to capture the primary event of interest, severe relapse, as well as the costs and outcomes associated with it, as severe relapse costs including hospitalisation are the major cost driver in caring for patients with SMI (135). As hospitalisation due to symptom exacerbation is used as a proxy for relapse in RADAR (also as the key factor of the confirmed definition of relapse in schizophrenia), the approximated average duration of hospital length of stay (LoS = 28 days, (135)) for schizophrenia patients determines the cycle length of this model. Conveniently this length is also the cycle which has been used by the antipsychotic reduction schedule in the RADAR trial. (148) Previous economic models also employed this design. (107, 128) The main outcome is incremental cost per QALY. The model has a 10-year horizon to capture the treatment effect and cost-effectiveness of the antipsychotic discontinuation strategy, without being too demanding on memory and time for computation which prohibits a life-time horizon. The model also captures the incidence of a series of adverse effects and comorbidities, including EPS, CVD and diabetes. I also consider public sector costs due to the substantial burden to the CJS. Figure 4-1 schematically describes the model structure. Individuals are assumed at risk of developing a series of complications and adverse effects, including EPS and cardiovascular events upon entering the model and transferring between stable and relapsed states. Moreover, I also consider the potential benefit of antipsychotic discontinuation, as suggested by the primary outcome of RADAR trial. I consider the purported improvement in wellbeing upon discontinuation in a series of scenario analyses. Different probabilities of

developing these complications and adverse effects depend on the relevant risk prediction models or data from the literature, which are described below.

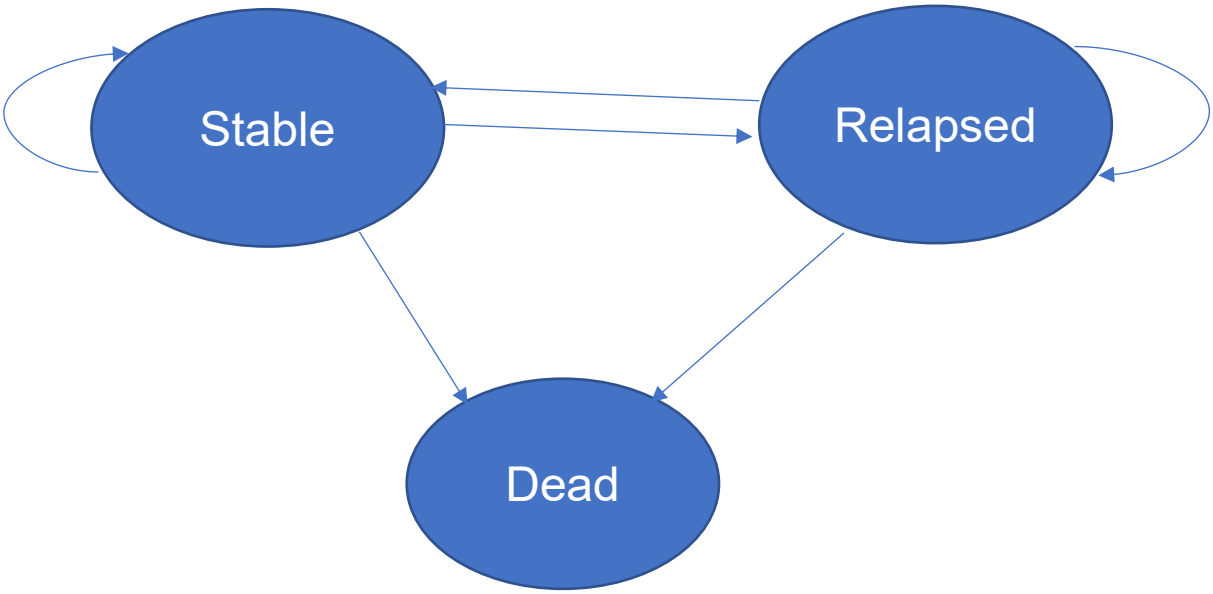

Figure 4-1 A schematic description for the model structure

The model is parameterised with existing evidence from meta-analyses and RCTs where information is relevant. Expert opinions are also referred to if no suitable current published resources are available. Some assumptions are made for the convenience and simplicity of the decision process. The whole decision model is built by R language, in the RStudio environment. I illustrate the structure of the iterative model with R implementation in Figure 4-2. The R code for the CEA is detailed in Appendix F.

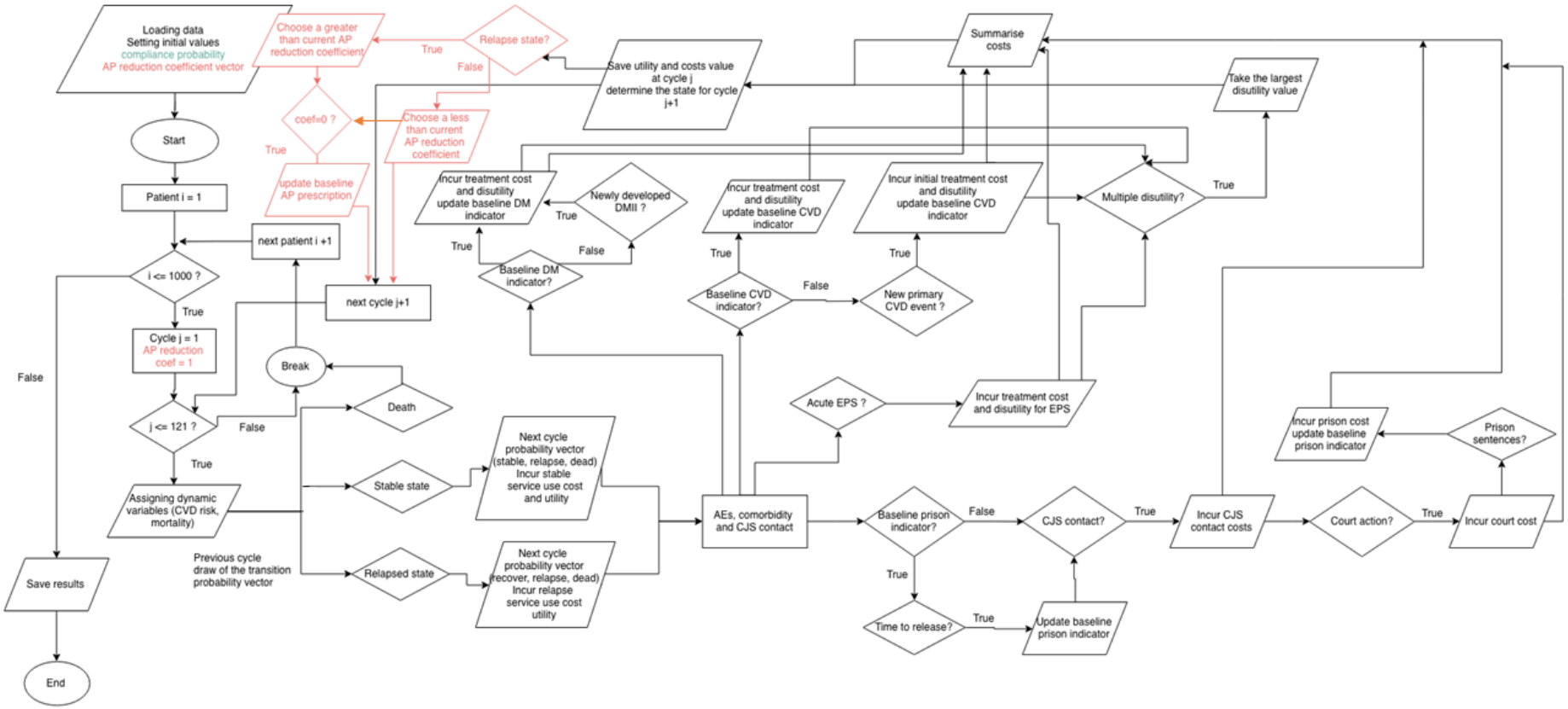

Figure 4-2 The iterative model structure implemented in R

Abbreviations: i, patient identifier; j, cycle; DM, diabetes mellitus; CVD, cardiovascular disease ; EPS, extrapyramidal symptoms; AE, adverse event

The design process involved conceptualising the clinical question and the model structure with the RADAR trial operation team as well as the chief investigator, through a presentation during a weekly trial team meeting. I took some informative feedback from the team, for instance the input parameter value of choice, and the antipsychotic prescription data.

4.2.2 Population

A cohort of schizophrenia patients was simulated based on the population used in the Prediction and Management of Cardiovascular Risk in People With Severe Mental Illnesses (PRIMROSE) study (17, 149). Patients who had a schizophrenia diagnosis and having previously taken antipsychotic treatment were included into the cohort, from which 1000 individuals were simulated using random sampling with replacement. Both PRIMROSE and simulated data include basic social demographics (sex and age) and clinical information (height, weight, systolic blood pressure, history of hazardous drinking, smoking status, antipsychotic prescription status, diagnosis of depression, anti-depressant prescription status, lipid drug status, cholesterol level, Townsend score of deprivation). I simulated a list of antipsychotic prescriptions according to these estimates, with the list of investigational drugs described in RADAR protocol. Due to the lack of prescription details from the PRIMROSE model, all antipsychotics were given equal probabilities of prescription. This is not likely to be the case in real clinical settings. Scenario analysis of all patients being prescribed the same antipsychotics was performed in 4.3.3.4 to see the likely impact of this assumption. As for the case of polypharmacy and clozapine augmentation, olanzapine with amisulpride or risperidone, or quetiapine with

risperidone are the most commonly prescribed antipsychotic combinations. Clozapine is often prescribed with amisulpride, haloperidol or surpiride (150). Other baseline risks, such as having EPS or diabetes, were simulated based on existing literature on the risks of these conditions in similar populations, or using clinical expert opinion. Sensitivity analyses were carried out varying the value of these risks to investigate the likely impact of the baseline risk. Table 4-1 shows a brief summary of the antipsychotic prescription on patients with schizophrenia in the UK, according to the National Audit of Schizophrenia.

Table 4-1 A brief summary of antipsychotic prescriptions in patients with schizophrenia in the UK in 2014

Prescription type	No. Cases	% of total sample
One oral	1900	33.9
One LAI[11]	1336	23.8
Clozapine	1135	20.2
Two orals	241	4.3
One LAI + one oral	428	7.6
Two LAIs	1	<0.1
Three orals	4	0.1
One LAI + two orals	21	0.4
Clozapine + one oral	381	6.8
Clozapine + one LAI	10	0.2
Clozapine + two orals	7	0.1
Clozapine + one LAI + one oral	1	<0.1

Each simulated individual enters the model sequentially and remains at one health state during each monthly cycle. All patients are assumed to be at stable state upon entering the model (cycle 1). Then the risk of relapse will be assigned to each individual, based on antipsychotics prescription and the intervention patients are receiving (maintenance vs. antipsychotic reduction and

[11] LAI = Long-acting injectable

discontinuation). An individual who is currently in the relapsed state will be assigned a probability of symptom remission, returning to the stable state in the next cycle, or can stay in the relapse state in the following cycle with a given probability. Mortality risks are estimated separately. Cardiovascular death are modelled in the PRIMROSE risk prediction model, whereas death of other causes are modelled in the supplementary Weibull model in PRIMROSE study. (151) Patients in the relapse state are assumed to have elevated risk of self-inflicted death, which is adjusted using the mortality gap from an epidemiological study. (152)

Patients are assigned with probabilities of developing adverse effects and comorbidities while they are in the model. According to the economic models discussed in the previous chapter, weight gain and EPS, are most commonly included adverse effects of antipsychotic treatment. Weight gain and diabetes risks were considered significant during first year of any specific antipsychotic treatment, but the long-term risks of developing weight gain or diabetes amongst patients with antipsychotic treatment are difficult to determine. Such events were therefore not modelled in the NICE clinical guideline for schizophrenia. (131) Instead, primary cardiovascular events were modelled, partly as the diabetes complication. Here I used a similar approach by estimating the cardiovascular events in the PRIMROSE model, with the Weibull survival model to estimate primary CVD events. (153-155)

Patients' clinical information, such as age, body mass index (BMI), systolic blood pressure and dosage are updated on a yearly basis, then related risks and transition parameters are re-calculated annually based on these updated

parameters. Sampling from appropriate distributions is applied where data or estimates for updating information are not available. All events or state transitions are assumed to be randomly sampled based on corresponding probabilities, or probability distributions.

4.2.3 Clinical parameters

The decision model captures the clinical effect of the antipsychotic reduction regime, primarily in terms of incidence of recurring symptom exacerbation, and also social functioning and quality of life. Results from Wunderink *et al.* suggested that a 2.1-fold increase in relapse incidence occurred in patients who were reducing antipsychotics during the study period, in which relapse was defined as an exacerbation of symptoms during at least 1 week with at least 1 relevant PANSS item score above 3. However the definition of relapse was different from that used in RADAR, which was defined as serious symptom exacerbation and hospitalisation was used as a proxy. Indeed, the definition of relapse covers a very broad range and usually varies between studies. In RADAR trial, relapse will be assessed as the principal secondary outcome, in ways such as admission to inpatient care or any acute care, in addition to follow-up assessments and clinical notes. Although many antipsychotic RCTs used relapses as a measure of safety profile, one may presume that the nature, and the relapse risks of patients who are on maintenance antipsychotic treatment reducing and discontinuing, are different from that of patients who only start specific antipsychotic treatment or switch to a different antipsychotic in antipsychotic RCTs. Hence the relapse risk data from these antipsychotic RCTs should be used with caution. Evidence from a meta-analysis of studies comparing antipsychotics versus placebo for relapse prevention was used for

the clinical input, in which the frequencies of hospital re-admission were pooled and compared between antipsychotics and placebo. (20) This estimate has also been used to form the treatment effects hypothesis of the RADAR trial, in which the sample size calculation was powered on incidence of serious relapses using a non-inferiority margin of 10%. Scenario analysis is performed using parameters from both studies. The probabilities of relapse are often defined as 1-year time to event outcome. To fit the parameters into the decision model, which requires monthly probabilities, it is assumed that relapse happening in each month is independent. Therefore, a monthly risk was calculating, using:

$p_m = 1 - (1 - p_a)^{\frac{1}{12}}$, where p_a is the annual probability of experiencing relapse. Alternatively, annual probabilities were transferred to rates using formula

$p = 1 - e^{-rt}$, where p is probability of interest and r is the rate. Table 4-2 summarises the transition probabilities between states and the risk of other events in the model.

The cardiovascular event risk is estimated using the PRIMROSE prediction model. (17) The PRIMROSE algorithm is a Weibull model applied to each individual to calculate annual risk of primary cardiovascular events respectively. Relevant patient level data are updated annually and fed into a separate survival model to obtain the annual cardiovascular event risks. Non-fatal primary CVD event included coronary heart disease (CHD) and cerebrovascular accident (CVA). Primary CHD event included stable angina, unstable angina, myocardial infarction (MI), coronary artery surgery unclassified CHD and fatal CHD while a primary CVA event consisted of haemorrhagic stroke, ischaemic stroke, transient ischaemic attack and fatal stroke. Probabilities of primary CVD

events at each year were estimated using Weibull survival models on data from the PRIMROSE economic model. All primary CVD events were considered mutually exclusive and enduring till the end of the model. Secondary CVD events were not considered, Zomer *et al*. discussed the risk of subsequent secondary CVD events in more depth elsewhere. (149) Risk of metabolic comorbidities, such as diabetes are referenced from existing literature. It is assumed that a patient's risks of developing diabetes and EPS decrease to zero if he or she discontinues antipsychotic medication in that cycle.

Table 4-2 A summary of transition probabilities and risks of other events in the model

	Base case	**PSA[12] distribution**	**Reference**
Relapse risk (maintenance)	0.0085	Beta (11.06, 1281.9)	(20)
Relapse risk (antipsychotic reduction)	0.024	Beta (14.1, 730.4)	(20)
Adherence probability	0.6	Beta (3, 2)	(156) Vague distribution for PSA
Diabetes risk	0.01	Beta (10,90)	Assumption
CJS contact risk	0.035	Beta (35,965)	(138)
Court risk given CJS contact	0.203	Beta (203,797)	(138)
Sentence risk (given court)	0.394	Beta (394, 606)	(138)
Monthly discharge probability	0.8	Beta (40, 10)	Assumption
EPS risk			
OLA	0.002	N/A	(131)
AMI	0.0027	N/A	
ZOT	0.0012	N/A	
ARI	0.0019	N/A	
PAL	0.0022	N/A	
RIS	0.003	N/A	
HAL	0.0046	N/A	
LAI	0.0087	N/A	
Mortality	from PRIMROSE model	N/A	(149, 152)

[12] PSA – probabilistic sensitivity analysis

Cardiovascular	from PRIMROSE model	N/A	(149)

Abbreviations: CJS, criminal justice system; EPS, extrapyramidal symptom; PSA, probabilistic sensitivity analysis; OLA, olanzapine; AMI, amisulpride; ZOT, zotepine; ARI, aripiprazole; PAL, paliperidone; RIS, risperidone; HAL, haloperidol; LAI, long-acting injectable;

4.2.4 Cost data

Direct costs include antipsychotic medication costs, hospitalisation costs due to relapse, outpatient primary and psychiatric care costs, costs of care for adverse effects and other co-morbidities. Antipsychotic costs are based on the advised dose and NHS indicative prices in the British National Formulary (BNF) (12), or the dosage and costs from the Cost Comparison Chart developed by Regional Drug and Therapeutics Centre (157). Table 4-3 summarises the average dose, frequency and acquisition costs for the investigational antipsychotics. The Cost Comparison Chart consists of the costs for 21 different second-generation antipsychotics treatment strategies. For antipsychotic treatment options which are not included in the Cost Comparison Chart but in RADAR's list of investigational medication, I estimated the acquisition cost based on the information from BNF using the recommended dosage and medical form.

Table 4-3 A summary of average dose, frequency and annual costs for antipsychotics (2018)

Antipsychotics strategy	Frequency	Dosage (mg)	Annual cost (£)
First generation antipsychotics (FGA)			
Benperidol	OD	1	1407.72
Chlorpromazine	TD	75 - 300	1350.36
Flupentixol	BD	9	167.04
Haloperidol	TD	9	358.8
Levomepromazine	OD	50	162.08
Pericyazine	OD	75 - 300	1760
Perphenazine	OD	10 - 24	600
Pimozide	OD	2 - 20	406.3248
Prochlorperazine	BD	25-75	30
Promazine Hydrochloride	TD	100-200	331.407437
Surpiride	BD	200-400	105.6

Trifluoperazine	BD	10	809.34
Zuclopenthixol	OD	20-50	108.3936
Clozapine	OD	200 -- 450	798.36
Second generation antipsychotics (SGA)			
Paliperidone	OD	12	3161.6
Quetiapine	OD	600	2062.67
Paliperidone	OD	9	1876.96
Paliperidone	OD	3	1264.64
Paliperidone	OD	6	1264.64
Lurasidone (Latuda®)	OD	18.5 - 74	1179.36
Asenapine (Sycrest®)	OD	10	622.44
Asenapine (Sycrest®)	OD	5	622.44
Risperidone (disp)	OD	4	606.97
Olanzapine	OD	20	447.33
Amisulpride	BD	400	427.21
Olanzapine oral dispersible	OD	20	400.27
Aripiprazole	OD	30	136.63
Olanzapine	OD	5	111.54
Olanzapine oral dispersible	OD	5	102.83
Amisulpride	BD	200	54.96
Risperidone	OD	6	39.78
Quetiapine	BD	225	38.58
Quetiapine	BD	150	27.54
Aripiprazole	OD	15	15.6
Risperidone	OD	4	10.07
Depot			
Flupentixol decanote (depot Depixol®, Depixol Conc.®, Depixol Low Volume®)	4 weeks	50-600	234.216
Aripiprazole (depot Abilify Maintena®)	monthly	9.75	41.16
Fluphenazine decanoate (depot Modecate®, Modecate Concentrate®)	monthly	400	3426.24
Haloperidol (depot)	4 weeks	300	181.872
Olanzapine embonate (depot)			
Paliperidone (depot)	3 months	175 - 525	3768.84
Risperidone (depot)	2 weeks	50	3426.24
Zuclopenthixol decanoate (depot)	2 weeks	200	151.248

Abbreviations: OD, every daily; BD, twice daily; TD, three times per day

Average unit costs of hospitalisation, including treatment, inpatient stay costs, are estimated from PSSRU (158) or the NHS reference cost (159). For adverse

effects and management of co-morbidities, e.g. acute EPS, diabetes, costs were obtained from relevant studies.

Indirect costs mainly include CJS costs due to aggression or criminal behaviour of patients. Unemployment and loss of productivity are also often included in the indirect cost. However estimation of indirect costs in the published models considered in the previous chapter shows considerable variability. (160) In the 2 studies of COI in the UK, both have identified that productivity losses are major contributors to the COI of schizophrenia. (161, 162). However, patients with long term schizophrenia often have little employability. Plus, evidence suggested that although intervention such as Individual placement and support model of supportive employment would increase job initiation rates amongst patients with SMIs, there was no differences in the length of employment, hourly wages, or hours worked with the intervention compared to the psychosocial rehabilitation programme, highlighting difficulties in retaining them at their jobs.(136) Hence I did not include employment or losses of productivity in the model for indirect costs. Table 4-4 summarises the unit costs and average units consumed for service costs in patients with schizophrenia. Table 4-5 summarises the costs for treating primary cardiovascular events, including the costs of initial treatment and follow-up treatment throughout the rest of the time horizon.

Table 4-4 A summary of unit cost of service use and criminal justice contacts (2014)

	Unit cost (£)	**PSA distribution**	**Average unit consumed (over 6 months)**		**Reference**
			Stable	Relapse	
Outpatient psychiatric visits	283.97	Gamma	1.4	2.1	(131, 159)
Outpatient other visits	119.84	Gamma	0.1	0.3	

Day hospital visits	119.84	Gamma	2.3	2.1	
Community mental health centre	133	Gamma	2.4	1.4	
Day care centre visits	54	Gamma	5.9	0.9	
Group therapy	54	Gamma	0.4	0.1	
Specialist education	54	Gamma	2.9	0	
Psychiatrist visits	300	Gamma	2.5	2.3	
GP visits	66	Gamma	1.8	1.6	
Cost of CJS contact	540	Gamma	N/A	N/A	(146)
Cost of tribunal	500	Gamma	N/A	N/A	
Prison cost (month)	3518	Gamma			

Abbreviations: GP, general practitioner; CJS, criminal justice system; PSA, probabilistic sensitivity analysis; N/A, not applicable

Table 4-5 A summary of treatment cost for cardiovascular events

CVD co-morbidity	**Initial treatment (£)**	**PSA Distribution**	**Subsequent treatment (£, annual)**	**Reference**
Unstable angina	566	Gamma	220	(149)
Stable angina	220	Gamma	220	
MI	5,720	Gamma	220	
Surgery	6,008	Gamma	N/A	
Unclassified CHD	2,169	Gamma	220	
TIA	1,368	Gamma	340	
Stroke	10,347	Gamma	2,782	
Unspecified CVA	5,858	Gamma	1,561	

Abbreviations: MI, myocardial infarction; CHD, coronary heart disease; TIA, transient ischaemic attack; CVA, cerebrovascular accident; PSA, probabilistic sensitivity analysis; N/A, not applicable

4.2.5 Antipsychotic reduction strategy

To mimic the nature of antipsychotic reduction, the reduction coefficients, in the form of a vector containing percentage of the maintenance dose, e.g. (0, 0.2, 0.3, 0.5, 0.6, 0.8, 0.9), were set *a priori*. The choice of the values in the vector was to reflect dosing changes at different stages of the RADAR intervention. For instance, a patient may take 100% of the prescribed dosage at the beginning of the trial, in month 2 the dosage is reduced to 90%, 80% at month 3, etc., if the patient remains stable. Patients in the RADAR intervention group

start with the maintenance dose, then reduce from the next cycle, using a random draw from the vector. Risk of relapse is assumed also to depend on the reduction coefficients. I then explored a series of possible relationships between the reduction coefficients and the adjusted relapse risk, including linear, exponential, square root, quadratic, or cubic. I introduced a monotonic function which allowed for changes in risk of relapse according to the reduction coefficients.

$$r_{coef} = \frac{r_{discontinu.}}{e^{\alpha k_{radar}}}$$

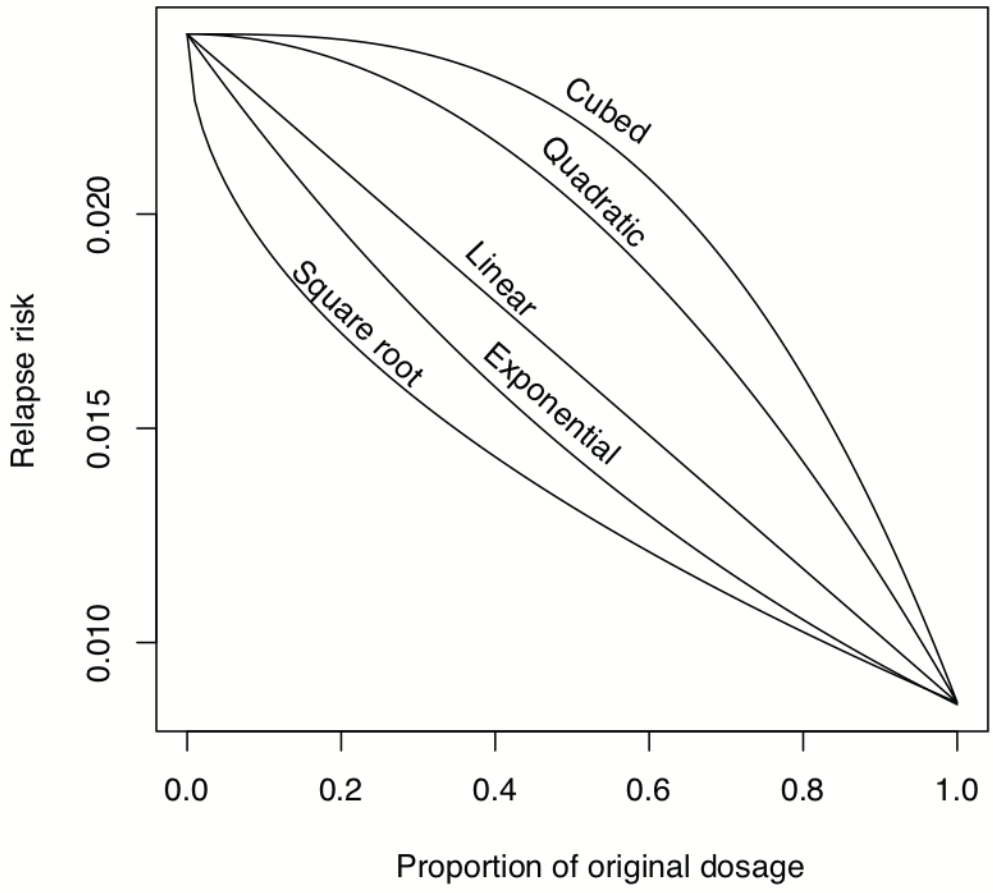

Figure 4-3 Relation between relapse risk and the proportion of original dosage

In this function, $r_{discontinu.}$ represents the risk of relapse when under no antipsychotic treatment (for instance, 0.0242 according to Leucht *et al.*) and

k_{radar} is the proportion of original dosage. Consider r_{coef} should equal the relapse risk for antipsychotic maintenance (that is $r_{coef} = 0.0085\ if\ k_{radar} = 1$), by plugging these values into

$$r_{coef} = \frac{r_{discontinu.}}{e^{\alpha k_{radar}}}$$

$$0.0085 = \frac{0.0242}{e^{\alpha \times 1}}$$

$$\alpha = \log\left(\frac{0.0242}{0.0085}\right) = 1.046$$

the value of α can be solved 1.046. Similarly, for the RADAR trial design of 10% non-inferiority margin, which gives a relapse probability of 0.0094, α is solved 0.041. Other relationships between the two parameters are listed as follows.

Linear $r_{coef} = \alpha \times k_{radar} + r_{discontinu.}$
Square root $r_{coef} = \alpha \times \sqrt{k}_{radar} + r_{discontinu.}$
Quadratic $r_{coef} = \alpha \times k_{radar}^2 + r_{discontinu.}$
Cubic $r_{coef} = \alpha \times k_{radar}^3 + r_{discontinu.}$

Figure 4-3 illustrates the relation between the proportion of original dosage, k_{radar} and r_{coef} ,under different relation functions. The exponential and the square root relations show concave curves, which allows for slower changes of the relapse risk at the beginning of reduction, compared with a linear relationship between the proportion of original dosage and the relapse risk which treats all stages of reduction equally. Quadratic and cubed relationships show more rapid change of relapse risk at the beginning of reduction. I explored how these relations will change the overall cost effectiveness in a scenario analysis.

4.2.6 Adherence

Non-adherence with prescribed dosages is reportedly highly associated with relapse. (135, 163, 164) The process of manualised antipsychotic reduction and discontinuation strategy resembles that of antipsychotic non-adherence, to the extent that patients are no longer taking the original prescribed dosages of medication, despite the fact that antipsychotic reduction may be more carefully managed during the trial. One could speculate that non-adherence may incur similar symptom exacerbation during the study and in real world. Therefore, a component of non-adherence is built in the model, with an indicator showing whether patients are fully adherent with the medication. The risk of relapse is then adjusted according to the adherence indicator c_i for the individual i, which follows binomial distribution with a predefined probability. Then the adjusted relapse risk can be calculated as below:

$$r_{i,comp} = r_i \times e^{(1-c_i)} \quad \begin{cases} c_i = 1, if\ fully\ adherent; \\ c_i = 0, if\ not \end{cases}$$

It only becomes active if the RADAR analysis shows an interaction between non-adherence and treatment allocation in the subgroup analysis model. (42) Sensitivity analysis was also performed regarding different levels of adherence (by varying the probability) for patients who were under maintenance treatment.

4.2.7 Criminal Justice System

Patients' encounters with the CJS are defined as a simple decision tree model of criminal behaviour and justice system. Patients will have probabilities of

committing criminal behaviour, hence incurring costs for the police force, depending on the state of being stable or relapsed. Different probabilities are assigned for the likelihood of entering court for a serious criminal offence, thus incurring the cost to the court. Subsequently, patients will have probabilities of being released after their court hearing or being given a prison sentence. The latter will then incur prison costs. A random sentence period (in monthly term) will be assigned once the decision of imprisonment is made based on Prison population statistics from House of Commons (165). Individuals then remain in prison until the end of the assumed sentence period. Both prison and service costs are assumed to be incurred while an individual is experiencing relapse during a period of incarceration.

4.2.8 Improvement in wellbeing for the reduction intervention

RADAR trial purports that patients who successfully discontinue antipsychotic treatment may achieve improvement in wellbeing, as measured as SFS, which is the primary efficacy outcome. The model also captures the purported improvement in this aspect, despite the fact that there is a lack of evidence on utility value for such improvement in patients with schizophrenia. The model assumes that in the reduction intervention, when a patient is in stable state and the antipsychotic prescription is discontinued, a utility increase is incurred. I then will explore to what amount this increase could help supporting the cost-effectiveness of the reduction strategy via a series of deterministic sensitivity analyses and threshold analysis, using a £30,000/QALY threshold.

4.2.9 Utility data

Lenert *et al.* estimated that a moderate schizophrenia state has a utility of 0.799 using the standard gamble approach, alongside utility values for different states of schizophrenia, as well as disutility of some adverse effects, such as EPS. (166) Utility of other adverse effects and comorbidities are obtained from relevant literature. For patients who are in either relapsed or stable state, a utility reduction will be applied to the utility value of the two states. As for the scenarios where more than one AE or comorbidity are presented concomitantly, the disutility of only the most debilitating AE or comorbidity is considered. The same approach was adopted by Lachaine *et al*. (119) .

Table 4-6 summarises the utility values of different health states and utility decrements for adverse effects and comorbidities. A discount rate of 3.5% annually was applied to both costs and outcomes.

Table 4-6 A summary of utility values for different health states in the model

Utility	**Base case**	**PSA distribution**	**Reference**
Stable	0.799	Beta (799, 211)	(166)
Relapsed	0.27	Beta (54,146)	(167)
Utility reduction			
Unstable angina	0.216	Beta (186.4, 676.6)	(149)
Stable angina	0.216	Beta (186.4, 676.6)	
MI	0.072	Beta (192.3, 2479.3)	
Surgery	0.072	Beta (192.3, 2479.3)	
Unclassified CHD	0.101	Beta (254.64, 2266.55)	
TIA	0.088	Beta (196.09, 2032.24)	
Stroke	0.185	Beta (193.52, 582.53)	
Unspecified CVA	0.153	Beta (198.12, 1096.79)	
Diabetes	0.151	Beta (53.47, 300.65)	
EPS	0.074	Beta (41.83, 523.48)	

Abbreviations: MI, myocardial infarction; CHD, coronary heart disease; TIA, transient ischaemic attack; CVA, cerebrovascular accident; PSA, probabilistic sensitivity analysis

To test the impact of the uncertainty of the input parameters on the overall cost-effectiveness, one-way sensitivity analysis was performed on all variables, including relapse probability in the reduction intervention, probability of being discharged within one cycle, risk of CJS contact and the following incidences (court contacts and incarceration), probability of being fully adherent with

antipsychotic prescriptions in the maintenance intervention, and utility values of both being stable and relapsed. All selected parameters were increased and decreased by 20% of their original value.

4.3 Results

4.3.1 Baseline characteristics

Patients were simulated from the PRIMROSE decision modelling dataset with criteria of antipsychotics medication (typical and atypical) and diagnosis of schizophrenia. The dataset is anonymised and originally from The Health Improvement Network (THIN) database, with the individual identifier removed. I then used random sampling with replacement to expand the sample to 1000 individuals.(168) Table 4-7 describes the characteristics of the original PRIMROSE dataset and simulated datasets.

The simulated data consisted of 587 males and 413 females, and the average age at baseline is 49.27 years. 122 individuals were diagnosed with diabetes at baseline. 35.1% were prescribed with FGAs and 72.0% were prescribed with SGAs.

Table 4-7 A summary of baseline characteristics of the PRIMROSE dataset and the simulated data

	PRIMROSE data (N=185)	**Simulated data (N=1000)**
Sex (%)	Male = 109 (58.92%) Female = 76 (41.08%)	Male = 587 (58.70%) Female = 413 (41.30%)
Age (years, Mean, SD)	49.55 (SD = 12.08)	49.27 (SD = 11.95)
Weight (kg, Mean, SD)	84.69 (SD = 19.65)	85.08 (SD = 20.01)
Height (m, Mean, SD)	1.69 (SD = 0.10)	1.70 (SD = 0.10)
BMI (Mean, SD)	29.29 (SD = 6.22)	29.29 (SD = 6.22)
Systolic Blood Pressure (Mean, SD)	129.38 (SD = 14.81)	129.25 (SD = 15.25)
Diabetes diagnosis (N, %)	23 (12.43%)	122 (12.20%)
Antipsychotic prescription (N, %)	FGA = 70 (37.84%) SGA = 129 (69.73%)	FGA = 351 (35.10%) SGA = 720 (72.00%)
Anti-hypertensive prescription (N, %)	31 (16.76%)	176 (17.60%)

High-density lipoprotein cholesterol (Mean, SD)	1.27 (SD = 0.38)	1.28 (SD = 0.37)
Lipid drug (N, %)	27 (14.60%)	141 (14.10%)
History of heavy drinking (N, %)	20 (10.81%)	108 (10.80%)
Depression diagnosis (N, %)	95 (51.35%)	517 (51.70%)
Antidepressant treatment (N, %)	62 (33.51%)	350 (35.00%)
Lithium treatment (N, %)	8 (4.32%)	36 (3.60%)
Townsend score of deprivation (N, %)		
Score = 1	14 (7.57%)	71 (7.10%)
2	20 (10.81%)	99 (9.90%)
3	36 (19.46%)	209 (20.90%)
4	58 (31.35%)	310 (31.00%)
5	57 (30.81%)	311 (31.10%)
Smoking status (N, %)		
Non-smoker	62 (33.51%)	336 (33.60%)
Ex-smoker	13 (7.03%)	84 (8.40%)
Smoker	110 (59.46%)	580 (58.00%)

Abbreviations: SD, standard deviation; BMI, body mass index;

4.3.2 Base case

4.3.2.1 Cost effectiveness of the antipsychotic reduction strategy

Over the 10-year period, the cost-effectiveness for antipsychotic reduction and discontinuation strategy is dominated by the maintenance strategy. The average incremental cost is around £15034.21 (SD = 122383.2) per person whereas the average incremental QALYs is -0.048 (SD = 1.221) per person, which makes the ICER in the northwest quadrant, meaning that within a 10-year horizon, the antipsychotic reduction and discontinuation strategy could incur more costs, but less effectiveness (although not by much). (Figure 4-4)

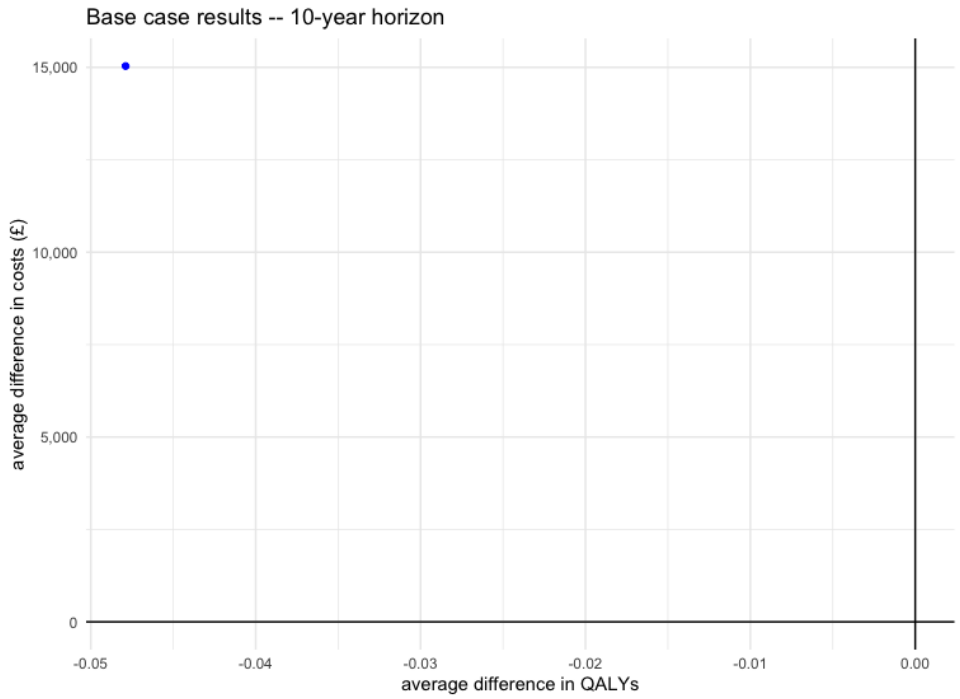

Figure 4-4 10-year average cost-effectiveness of the antipsychotic reduction and discontinuation strategy versus the maintenance strategy

Costs from the NHS perspective over 10 years in the maintenance intervention are £62018, whereas reduction intervention has incurred £75185 per person over 10 years (SD=25102). The overall cycles of serious relapse, the sum of all relapse cycles for the 1000 individuals over the 10-year horizon, in the maintenance intervention is 1145, whereas the reduction intervention has 2131 more cycles in which patients were hospitalised. In terms of CJS costs, over 10 years the maintenance intervention costs £43533 on average (SD = 78633) and the reduction intervention costs £45399 (SD = 83871) on average. 10-year average service costs for relapse, as part of the service costs, is £9427 per individual (SD = 11005.32) for the maintenance intervention and £26912 (SD = 18749.78) for the reduction intervention. Table 4-8 lists the results for the main outcomes.

Table 4-8 A summary of the results for the main outcomes

	Maintenance	**Reduction**
Total cycles of relapse	1145	3276
Mean NHS costs (£, per person)	62018.0 (SD = 20716.97)	75185.6 (SD = 25102.18)

Mean CJS costs (£, per person)	43533.0 (SD = 78633.09)	45399 (SD = 83871.71)
Mean public sector costs (£, per person)	105551.0 (SD = 83614.8)	120585.6 (SD = 88941.9)
Mean relapse costs (£, per person)[13]	9427.0 (SD = 11005.32)	26912.0 (SD = 18749.78)
Mean CVD costs (£, per person)	929.1 (SD = 3862.2)	1517 (SD = 8461)
Median cycles of discontinuation	N/A	101
Overall monthly unit antipsychotic prescription (10-year horizon)	120	9.50 (SD = 5.06)[14]
Number of individuals who remain stable (10-year horizon)	403	92
Total deaths (10-year horizon)	139	155
10-year median survival (month)	120	120
Total primary cardiovascular event rate (per 1000-person years)	13.4	11
Total diabetes diagnosis	722	307

Abbreviations: SD, standard deviation; CVD, cardiovascular disease

In the antipsychotic reduction intervention, all 1000 individuals have had cycles in which they discontinued antipsychotics at least once. The median of total accumulative discontinuation cycle was 101 cycles, suggesting that more than 50% of the patients, either intermittently or continuously, are not taking any antipsychotics for 8.5 years out of 10 years in the reduction intervention. Moreover, the overall amount of antipsychotics prescribed in the reduction intervention, were remarkably lower compared with the maintenance strategy (on average 7.9% of the amount in the maintenance intervention). Overall there are 403 individuals who did not relapse over the 10-year horizon in the maintenance intervention and 92 in the reduction intervention. Figure 4-5 illustrates the antipsychotic reduction pattern for a random selection of 4 individuals in the population. Except for individual 501 (bottom left), who

[13] Relapse costs are a part of the service costs (breakdown).
[14] p-value < 0.00001 based on t-test

seemed completely discontinued antipsychotic and remained off treatment for the 10-year time horizon, the other three all had similar patterns, in which they had intermittent periods of discontinuing antipsychotic treatment, but with different frequencies. Individual 301 (top right) had most episodes recurring relapses, which may require antipsychotic to control symptoms. 201 (top left) and 701 (bottom right) had fewer relapses, indicating some of the more successful reduction and discontinuation.

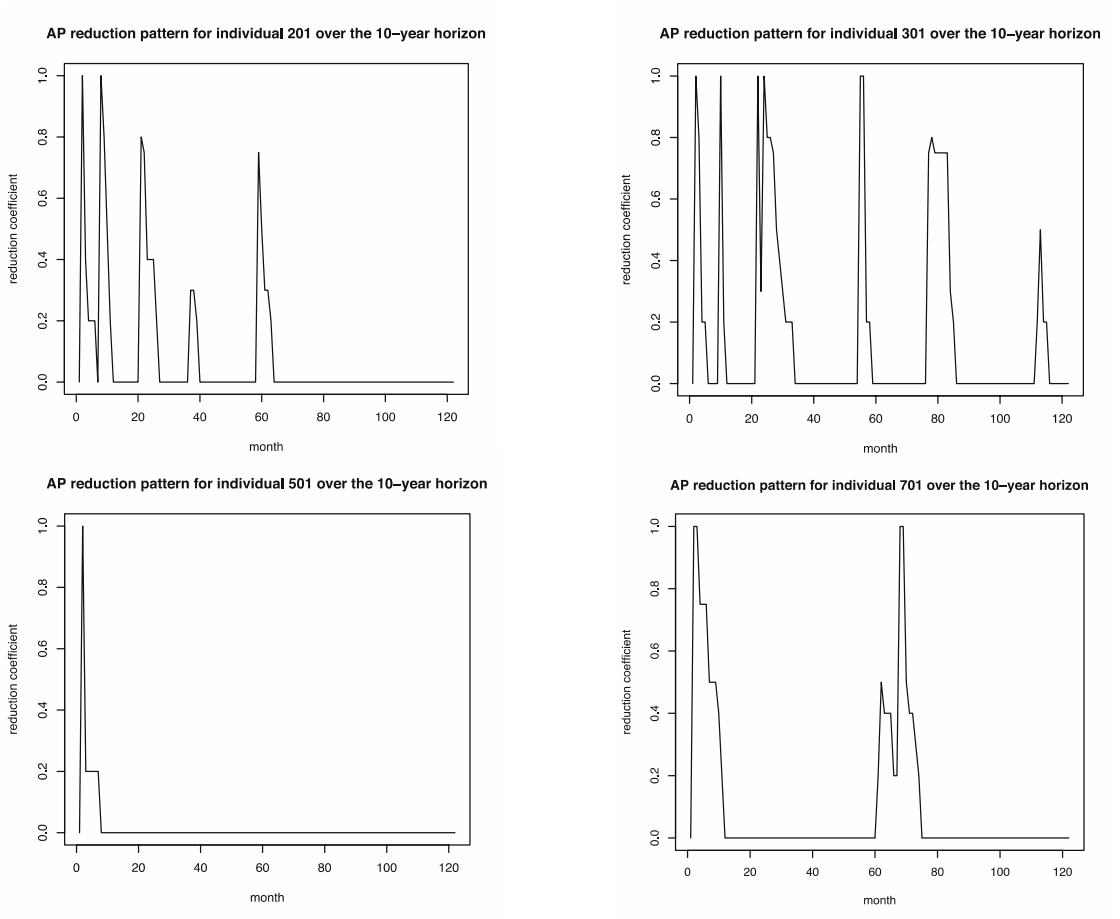

Figure 4-5 An illustration of the simulated antipsychotic reduction and discontinuation pattern of a random selection of 4 individuals in the simulated population

There were 139 cases of total all-cause death in the maintenance arm and 155 cases in the reduction arm, although the median of 10-year survival is 120 cycles, which is the time horizon. Besides, the Kaplan-Meier curve of total

survival suggests no meaningful difference in terms of mortality between the groups. (Figure 4-6)

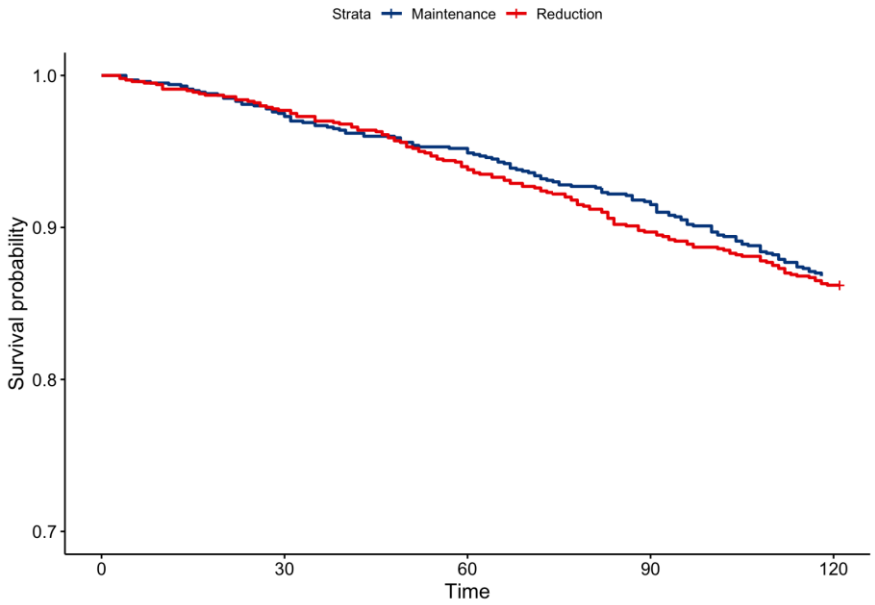

Figure 4-6 Kaplan-Meier curve for total survival on the two intervention

Primary cardiovascular events in the reduction intervention showed moderate reduction compared to the maintenance intervention. The total primary CVD event rate in the maintenance group is 13.4 per 1000-person year whereas that in the reduction group is 11.0 per 1000-person year. However, the Kaplan-Meier of the total CVD event comparison of two interventions almost overlap, suggesting insignificant difference in the total CVD events for the 10-year horizon. (Figure 4-7) Compared to the baseline of 122 diabetes diagnosis, there were 600 more cases of diabetes developed over the 10 year horizon for the maintenance and 185 more cases for the reduction intervention.

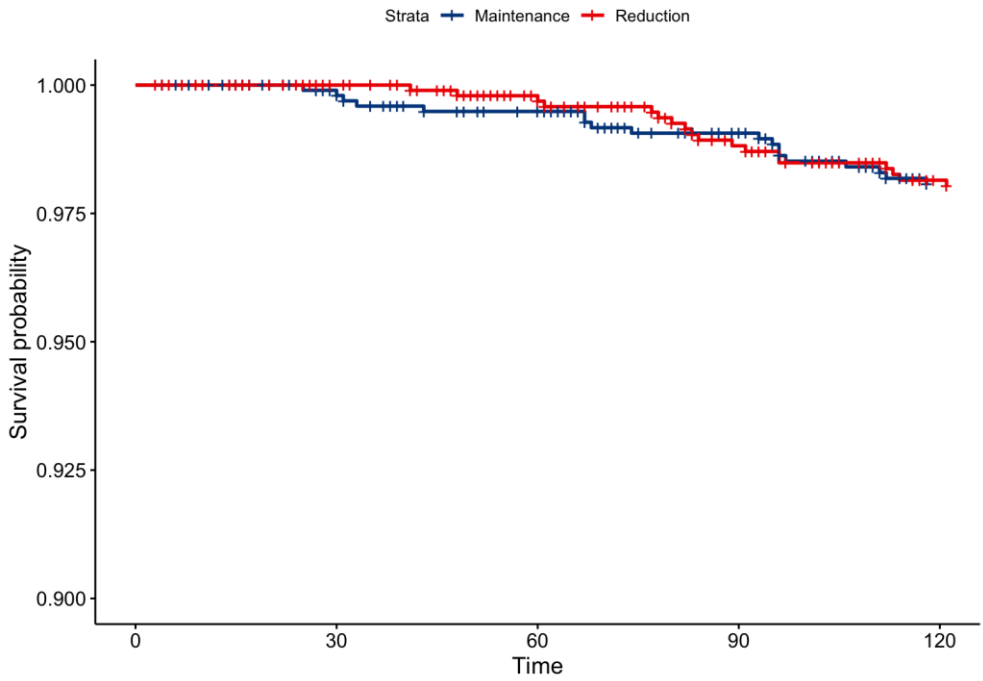

Figure 4-7 Kaplan-Meier curve for total CVD survival on the two intervention

4.3.3 Scenario analyses

4.3.3.1 RADAR's Design

I also performed scenario analysis on the hypothesis used in the RADAR design. Based on the 10% non-inferiority margin of the event rate on serious relapse (hospitalisation), I used 5% and 10% threshold of increase compared to the maintenance intervention for some scenario for the relapse risk in the reduction intervention for the analysis, which translates to monthly risk of 0.0089 (5%) and 0.0094 (10%). The results of a 5% increase of the event rate yields the results that the reduction intervention is dominating, and the ICER is located in the southeast quadrant, meaning less cost and better outcome. On the other hand, a 10% increase on the event rates still gives a dominated result.

4.3.3.2 Different relations between proportion of original dosage and risk of relapse

In 4.2.5 I introduced a series of different relationships between the risk of relapse and the proportion of original dosage. The ICERs remain negative

regardless of the relationship, except the square root relationship that gave a positive but very high ICER. However, overall the changes in incremental QALYs is small-scale where the incremental costs remain positive. This could be due to random sampling simulation error. Table 4-9 shows the result of each relation in one simulation.

Table 4-9 ICER, incremental cost and QALY under different relation assumptions between the risk of relapse and proportion of original dosage

Relation	ICER	ΔQALY	ΔCosts
Exponential	Dominated	-0.048	15034.21
Linear	Dominated	-0.029	11596.1
Square root	293011.6	0.035	10364.26
Quadratic	Dominated	-0.024	9854.34
Cubic	Dominated	-0.054	12597.21

Abbreviation: ICER, incremental cost-effectiveness ratio; QALY, quality adjusted life year

4.3.3.3 Improvement in social functioning for discontinuation

I also performed a scenario analysis to consider the purported improvement in wellbeing, as measured in SFS in RADAR trial. A utility increase of 0.064 to represent the improvement on discontinuation while remaining stable at each cycle, could bring the ICER to £36360/QALY. Similarly, an increase in 0.096 can result in the ICER to £24610/QALY. The threshold of £30000/QALY showed that a utility increase of at least 0.0815 can make the overall cost effectiveness fall between the £20000 – £30000/QALY region. The benefit of the purported wellbeing upon discontinuation, using the 0.0815 threshold value, could lead to an ICER that falls in the £20000 – £30000/QALY region, compared to the base case dominated results.

4.3.3.4 All patients prescribed with one antipsychotics

To test the assumption of equal probability for all antipsychotics to be prescribed, a scenario analysis on all patients receiving the risperidone was performed. The result was still dominated, with the incremental costs of £15931.0 and incremental QALY of -0.089.

4.3.4 One-way sensitivity analysis

Table 4-10 summarises the one-way sensitivity results on parameters that has most impacts on the results. Except a 20% decrease on the monthly relapse risk for the reduction intervention can lead to a positive but high ICER of £200405.9/QALY, all other results were still dominated regardless of the changes on the parameters, suggesting not cost-effective of the intervention by varying key input parameters by 20%. Sensitivity analysis was also run varying discount rate for both costs and outcomes at 3% and 4% annually, but the cost-effectiveness remained dominated.

Table 4-10 One-way sensitivity analysis results on key parameters

Variable	Inc. Costs (£)	Inc. QALY	ICER
-20%			
Relapse (reduction intervention)	8581.004	0.04281813	200405.9
Diabetes risk	9926.552	-0.03641	Maintenance dominates reduction
CJS risk	16161	-0.0669351	Maintenance dominates reduction
Court risk	12117.75	-0.0087155	Maintenance dominates reduction
Incarceration risk	13766.63	-0.0891034	Maintenance dominates reduction
Probability of discharge	18285.42	-0.0627886	Maintenance dominates reduction
Stable utility	15034.21	-0.0226135	Maintenance dominates reduction
Relapse utility	15034.21	-0.0557909	Maintenance dominates reduction
+20%			
Relapse (reduction intervention)	14166.79	-0.0275159	Maintenance dominates reduction
Diabetes risk	13489.4	-0.0305081	Maintenance dominates reduction
CJS risk	19087.05	-0.0498374	Maintenance dominates reduction
Court risk	16252.32	-0.0610878	Maintenance dominates reduction
Incarceration risk	9400.632	-0.066185	Maintenance dominates reduction
Probability of discharge	6116.89	-0.0237837	Maintenance dominates reduction
Stable utility	15034.21	-0.073166	Maintenance dominates reduction

Relapse utility	15034.21	-0.0399886	Maintenance dominates reduction

4.3.5 Probabilistic sensitivity analysis

Figure 4-8 illustrates the results of probabilistic sensitivity analysis (PSA). The ICER estimates were located above the threshold and largely in the north-west quadrant, indicating non cost-effectiveness of the intervention. Figure 4-9 shows the Cost-effectiveness Acceptability Curve (CEAC) of the reduction intervention based on the standard PSA result. The probability of being cost effective increases as the WTP threshold increases, only reaching a mere 10% at around £10,000,000/QALY.

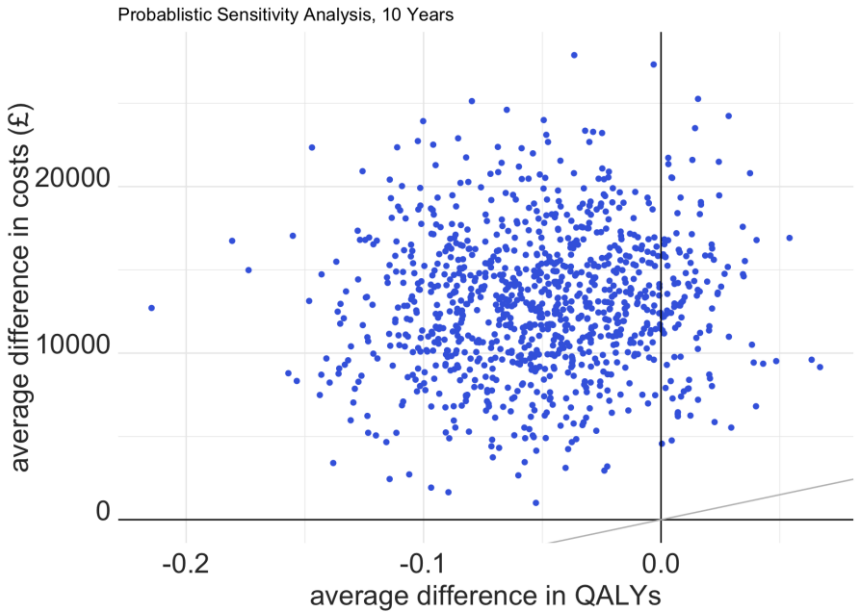

Figure 4-8 Probabilistic sensitivity analysis

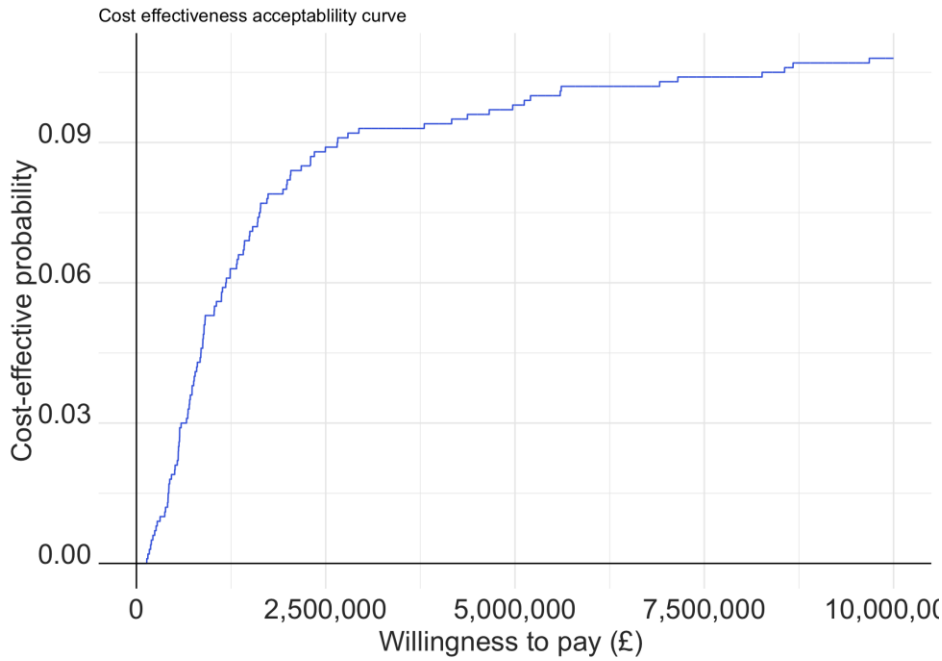

Figure 4-9 Cost effectiveness acceptability curve

4.4 Discussion

The results of the decision modelling suggest that with the current level of evidence available, the reduction and discontinuation strategy leads to a dominated result in cost-effectiveness. For the base case, the overall incremental costs for the antipsychotic reduction intervention to the maintenance intervention remain positive whereas the incremental QALY is negative, hence the negative ICERs. This suggests that the intervention may incur more costs in the long run while the effectiveness is not meaningful. The incremental costs over 10-year horizon may be attributed to the increased episodes of relapse, which costs notably more than the stable state. However, by varying the risk of relapse in the scenario analysis and sensitivity analysis, it showed that the intervention could potentially achieve cost effectiveness if the difference in relapse rates between two arms are smaller. Furthermore, it showed 21.8% reduction in the primary CVD event rate in the reduction intervention compared to the maintenance. The model also showed reduction in

risk of developing diabetes for the reduction intervention. Risk of diabetes amongst patients who are taking antipsychotics is discussed extensively, although most of the studies investigated short-term (maximum 18 months) risk of developing diabetes since treatment initiation, whereas the long-term risk for antipsychotic-induced diabetes or weight gain requires more research. (169 and 170) However, the long-term follow-up nature of the risk may be subject to any other factors such as diet or lifestyle, in addition to antipsychotic medication. Diabetes complications commonly seen as cardiovascular events amongst patients with SMIs, are modelled using the PRIMROSE prediction model. (171)

A scenario analysis considering the purported benefit, the improvement in social functioning upon discontinuation, shows that an average of 0.0815 increase in utility can achieve cost-effectiveness under the £30000/QALY threshold. Despite that evidence on social functioning improvement in terms of utility is still under developed, it can be a good opportunity for RADAR to provide a precise estimate for this. Meanwhile, the CJS costs, are higher in the reduction intervention compared with the maintenance intervention. This is due to the elevated incidence of relapse episodes, which were suggested to be associated with higher risk of criminal offences. (138)

This model takes both public sector perspectives (NHS + CJS) to discuss the potential risks and benefits of the antipsychotic reduction and discontinuation strategy. The results show potential risks of relapse in the antipsychotic reduction intervention may be determinant to the overall cost-effectiveness, as the reduction intervention does not seem to show considerable benefits in terms of reducing the risk of the co-morbidities or adverse effects, which could

potentially compensate for the effect of relapses. The reduction intervention suggests there is potential reduced risk of cardiovascular events upon discontinuation, although the PRIMROSE cost-effectiveness study had a higher prevalence using the same model. This could be due to the difference in data source which was fed into the model and the difference in population characteristics. The PRIMROSE study used the original THIN data from a population with SMIs including bipolar disorder, which I excluded in this chapter. A nearly 20% decrease in cardiovascular risk suggests the benefits of reducing and discontinuing antipsychotics amongst this population.

Also, the utility value used for relapse is significantly lower compared with stable state, with or without comorbidities, which means the effectiveness of one relapse cycle can equalise two or three cycles of stable with comorbidities. This is due to the definition of relapse in the model as well as in the RADAR trial. Serious relapse with hospitalisation is thought to be a more disagreeable state than moderate relapse, for which the utility values are reported to be around 0.6 as discussed in the previous chapter. The one-way sensitivity analysis suggests that increasing the utility for relapse may increase the overall incremental QALYs, hence making the overall ICER more favourable towards the reduction strategy.

The RADAR trial aims to demonstrate that not only is it feasible to reduce and discontinue antipsychotics, the potential improvement in wellbeing is also one of the benefits of RADAR intervention. One of the limitations of the model is that potential improvement in social functioning was only considered in a scenario analysis, as there is a lack of body of evidence of projected improvement of

wellbeing (SFS) in patients with schizophrenia and its mapping into health-related quality-of-life outcome measure, which is RADAR primarily seeks to investigate. Since the RADAR baseline data was not available by the time this thesis is being written, attempt to map the purported improvement in wellbeing to health utility was difficult. It could have supported the intervention from a perspective where more health benefits might be gained from this intervention had it been integrated into the decision model.

There is substantial uncertainty risk of relapse of patients reducing antipsychotics in this model as the body of evidence remains to be explored further. The trial by Wunderink *et al*. compared the safety and effectiveness of a similar antipsychotic reduction strategy, however the definition of relapse was broader compared with the one in RADAR. Hence the relapse risk may not be suitable for this model. I instead designed a monotonous function accounting for relapse risk and proportion of reducing antipsychotic, based on previous conclusion on partial adherence may increase risk of relapse. (172, 173) More research should be performed to focus on this aspect and in RADAR it is the key safety outcome that will determine the safety profile of the intervention.

The decision model also suggested that it may be possible for patients to discontinue antipsychotics for a period of time after gradual reduction. But symptom exacerbation could occur, which may again require antipsychotics medication to control the symptoms. Many individuals relapse more than once in the 10-year horizon. However, for those who did not relapse within 10-year horizon (for instance individual 501 in Figure 4-5), it is still possible that patient may still be susceptible to symptom exacerbation beyond the 10-year horizon.

This could be suggestive to the extent that antipsychotic may be used intermittently as means of controlling the symptoms when serious relapse occurs, in combination with interventions that can help with early signs of moderate relapse symptoms, despite the risk of unpleasant experience of serious relapse. (131)

The debate of reducing antipsychotics on patients with SMI is centred around primarily, the incidence associated with reduction and withdrawal. Although possibly avoidable, rebound psychosis, withdrawal reactions and psychological reactions which are often mistaken for relapse, can result in similar aggressive behaviour of patients, therefore endangering themselves and others. (174) Antipsychotics in this case, perhaps from a broader perspective, can effectively control the risk of aggression behaviour amongst patients. (175) The Mental Health Act allows compulsory treatment for patients, mainly in form of medication under specific circumstances. This in a way deprives patients' right of refusing treatment and treats patients disrespectfully, despite being backed up by law and justified by appeals to the patient's best interest. For patients with schizophrenia, some may find their treatment more harmful than helpful, and compulsory medication could evoke their negative feelings. This could be further supported by measuring health related quality of life using an appropriate questionnaire tool.

While the advocacy for antipsychotic treatment is made amongst carers and their clinicians, it should not be neglected that unless a case of risk of violence or harm, patients have the right of accepting or in this case, refusing the treatment they receive, as indicated in NICE's service user experience in adult

mental health clinical guideline. (176) It is above all, many patients' inclination that at some point they do not have to receive long-term antipsychotic medication. (177) RADAR trial may find itself in a meaningful position that it could potentially provide the solution for both patients and the society. Different from majority of the clinical trials which investigate the efficacy, safety and effectiveness of a novel treatment, RADAR is one of the few which investigates an intervention that "removes" the treatment in patients. The appropriate comparison of such a supervised reduction intervention would be to leave the patients as they behave in the real world, where more "unsupervised" non-adherence may occur. However, it would be unethical to randomise patients into such an unsupervised intervention. If RADAR can demonstrate the benefits of reducing antipsychotics, which outweighs the risk of relapse and the associated incidence, the antipsychotic reduction and discontinuation strategy may deal with the dilemma. Expectation that patients could safely reduce and discontinue antipsychotic is particularly meaningful in the way that antipsychotic may no longer be used regularly and long-term. Instead, safely discontinuing antipsychotic medication, or the intervention of doing so, could become the recommended care after symptom is controlled by antipsychotics. For many years antipsychotics work as the invisible strait-jacket and RADAR provides an opportunity for patients to eventually remove it safely, to both themselves and the rest.

4.5 Conclusion

This chapter described a decision analytic model to estimate the cost-effectiveness of antipsychotic reduction and discontinuation intervention. The intervention did not demonstrate cost-effectiveness compared to the

maintenance strategy in the results. The incidence of relapse is determinant in the overall cost-effectiveness and more cases of serious relapse appeared in the antipsychotic reduction and discontinuation intervention over the 10-year horizon. More empirical research could benefit the evidence base in this area, as the evidence base for relapse risk when reducing antipsychotics, long term effect of weight gain or diabetes risks, and the potential health benefits of reducing antipsychotics are still hard to come by.

Although the results suggested the antipsychotic reduction may not be cost-effective compared to the maintenance strategy, without the support of RADAR trial, there can still be uncertainty on the overall cost-effectiveness. It is therefore potentially informative of reducing the decision uncertainty by refining the estimate of the model parameters. In the next chapter, I will discuss the potential benefit and value of conducting future research to reduce the decision uncertainty, under the VoI framework.

5 Chapter 5 – Value of Information (VoI) analysis to determine the extent to which future research is worthwhile

In the last chapter, I introduced a decision model to compare the proposed antipsychotic reduction and discontinuation strategy to the maintenance strategy. The PSA suggested that the RADAR intervention was not cost – effective. The uncertainty of the decision for reimbursement recommendation might be reduced by investing in future research. In this chapter I will use the VoI approach to discuss the potential meaning of conducting future research on the key parameters in the decision model in Chapter 4 to reduce the risk of making the wrong decision for adoption, and further, discuss how the VoI approach may also be informative to research design in terms of the net benefit. The aim of this chapter is to use VoI approach to discuss what future research may be contributing to reducing the decision uncertainty based on the economic analysis for two scenarios a) the PSA results; b) based on a hypothetical PSA results to demonstrate how VoI may inform future research planning in a hypothetical scenario.

5.1 Introduction

In HTA for single payer healthcare systems, answers to whether an innovative therapeutic intervention should be adopted do not only depend on the efficacy and effectiveness, for which randomised control trials can provide unbiased and precise evidence. Moreover, the decision also depends upon the relative value for money, which is the incremental cost-effectiveness of the technologies. Trial based economic evaluation estimates the ICER for the new intervention versus the comparator. In the UK, the conventional threshold to adopting a new technology is between £20,000 and £30,000 per QALY. (178) Meanwhile, the

ICER may also be interpreted as NMB, in which a WTP, λ is taken into account, together with the incremental effectiveness ΔE and costs ΔC.

$$NMB = \lambda \times \Delta E - \Delta C$$

When making decisions using the existing evidence, the goal is to maximise the health gain for the population as a whole. In this case interventions with highest NMB are favourable. When performing decision analytic models or trial-based economic evaluations, parameter uncertainties need to be taken into account to reflect the decision uncertainties from the analysis. This is often achieved by performing PSA, which has become standard practice when performing economic analysis. (178, 179) Adoption decision is therefore taken under a certain level of uncertainty for the NMB. The risk of making a wrong decision on funding a new programme due to the uncertainty, and consequently the opportunity cost of making the wrong decision, could be reduced with more research carried out on to reduce the uncertainties on the parameters. On the other hand, extra research will incur extra cost, so the trade-off between investing extra into reducing the decision uncertainty and making the decision with current level of certainties but potential higher risk of making a “wrong” decision, is what VoI analysis seeks to answer. If the expected value of the further research exceeds the cost, then the research should be undertaken. The key statistics in the VoI analysis include EVPI, EVSI, expected net ENGS and EVPPI. The use of VoI alongside CEA in HTA has been increasing in recent years. EVPI is the expected value of learning perfect information about all input parameters, ideally by eliminating all parameter uncertainties. This can be directly derived from the PSA without any further analytical models.

However, it is often unrealistic to gain estimates of all parameters in one study. Therefore, interest leans towards obtaining perfect information on a subset of parameters. The EVPPI is calculated as the difference in the monetary value of health gain when making a decision to fund therapeutic alternatives (and represented in an economic model), between when a decision is made on the basis of information that is currently available (i.e. uncertainty in parameters of interest) and when the decision is made based on perfect information (no uncertainty in parameters) on key parameters. Challenges are associated with the computational complexity when calculating EVPPI, which involves integrating out the rest of the uninteresting parameters. Analytical solutions to integrating out random variables that follow complex probability distributions is often unachievable. Approximation using 2-level Monte Carlo simulation based on the PSA sample is widely used, which often involves intensive computational tasks. Methods that seek to simplify the computational burden have been proposed in recent years and considerable strides were made into making the computation of EVPPI more efficient, although the additional assumptions required may not always be appropriate. (180, 181) On the other hand, it is worth bearing in mind that uncertainty, whether in parameters or in the overall decision, can only be reduced but not eliminated, hence both EVPI and EVPPI always provide the solution for an idealistic situation (that is the “best” case scenario, but often unrealistic). EVSI estimates the value of proposed future research in order to reduce some of the parameter uncertainty. In this situation, EVSI can help determine the optimal research design that can maximise both the reduction in uncertainty and the value to the society of conducting the study. (182) Similar to EVPPI, numerical solution of EVSI commonly requires a 2-level Monte Carlo approach, where an outer loop generates plausible data sets,

conditional on the selected parameters of the economic decision model which get updated using Bayes theorem and are sampled at the inner loop, where the decision model is evaluated and the NMB results are calculated at each iteration. (65) This is particularly computationally demanding and may take days or even weeks to run on a computer. Generalised additive models (GAM) have been introduced in a non-parametric approach in recent years in order to simplify the computation process by only using the PSA sample. This takes a shorter time to compute. (183)

Recent strides in simplifying the computational intensity of calculating EVPPI adopt numerical approximation approaches with statistical tools such as Gaussian process (GP) and GAM. (184) GAM is efficient for single-parameter and low-dimensional EVPPI but can become time consuming and unstable for larger subsets of parameters ($n \geqslant 5$), in which case GP may be more appropriate. (185) However, GP calculations can also be time consuming as the number of parameters increases. Heath *et al.* introduced a fast GP calculation using the Stochastic Partial Differential Equations - Integrated Nested Laplace Approximations (SPDE-INLA) algorithm, which provided an efficient and stable way of estimating EVPPI.

The RADAR trial was designed on a 10% non-inferiority margin for serious relapse rates between treatment groups. In the previous chapter I analysed the cost-effectiveness of the antipsychotic reduction and discontinuation strategy, under different assumptions of the serious relapse risks (best and worst cases scenario). Decision of adopting this strategy may benefit from reducing the uncertainty on relapse event rate by further research. In this chapter I introduce

a series of VoI analyses, to demonstrate how decision risk on adopting antipsychotic reduction and discontinuation strategy may be reduced by future research into this topic.

5.2 Methods

The calculation of EVPI often involves taking the expectation over all values of parameters θ of interest with regard to the optimal decision in terms of net benefit.

$$EVPI = E_\theta \left[\max_t NB_t(\theta)\right] - \max_t E_\theta[NB_t(\theta)] \qquad (60)$$

Wilson described in fine details as to how EVPI can be calculated using the numerical solution with PSA sample. (9)

EVPPI of a group of parameters ϕ, within all parameters $\theta = (\phi, \psi)$, where φ are parameters we are not interested (also sometimes called “nuisance” parameters), is calculated by taking the expectation of the value of the optimal decision under the condition where perfect information of ϕ, across the support of ϕ and less the value of the current optimal decision. ϕ can be a single parameter, or a group of different parameters.

$$EVPPI = E_\phi \left[\max_t E_{\psi|\phi}[NB_t(\phi, \psi)]\right] - \max_t E_{\phi,\psi}[NB_t(\phi, \psi)] \qquad (181)$$

EVSI for a group of parameters, in other word, the value of a study which aims to reduce the uncertainty for these parameters, is calculated assuming that study may give rise to data X, which update the information on model parameters by Bayes theorem.

$$EVSI = E_{\mathrm{X}} \left[\max_t E_{(\theta|\mathrm{X})}[NB_t(\theta)]\right] - \max_t E[NB_t(\theta)] \qquad (183)$$

EVPI is the difference between the maximum net benefit with the perfect information and that with current information, which is the PSA results. Numerical approach was calculated according to explanations in the PhD thesis by Wilson. (186) The method was incorporated in both Sheffield Accelerated Value of Information (SAVI) and Bayesian Cost-Effectiveness Analysis web (BCEAweb) applications. I choose £30,000 willingness to pay threshold for EVPI. I consider methods developed both Strong *et al*. and Heath *et al.* for the EVPPI calculation, in their user friendly interface BCEAweb (https://egon.stats.ucl.ac.uk/projects/BCEAweb/). Both single parameter EVPPI and group EVPPI were calculated on parameters of interest. For EVSI, I use the methods proposed by Strong *et al.*, which estimates the EVSI using GAM regression. Then

$$\widehat{EVSI} = \frac{1}{N}\sum_{k=1}^{N} \max_{d} \hat{g}_d^{(k)} - \max_{d} \frac{1}{K}\sum_{k=1}^{K} \hat{g}_d^{(k)} \text{,(183)}$$

where $\hat{g}_d^{(k)}$ are GAM fitted values for decision option d. The first part of the equation involves taking the mean of all of the maximum between 0 and the fitted regression values, the second item on the equation involves taking the mean of the regression fitted values and finding the maximum between 0 and the mean.

Estimates for EVSI are calculated by running 500 iteration of the GAM model and taking the mean as the point estimate of EVSI with regard to sample size N, with corresponding 2.5% and 97.5% quantiles calculated in the same fashion. Population-level EVPI and EVPPI can be obtained by multiplying the estimate by the total number of affected population. Population-level EVSI are obtained by multiplying the estimate by the number of population who will in the future benefit from the research, which is the total affected population less the number

of participants in the proposed study. For ENGS with the proposed study design, $ENGS_n = EVSI_n - TC_n$, where TC_n is the cost of sampling. For single arm studies, $TC_n = C_f + n_s \times C_v$, where C_f is the fixed cost of study and C_v is the variable cost (recruitment) per patient. For parallel group RCTs, $TC_n = C_f + 2 \times n_s \times C_v + n_s \times |b_0|$, where b_0 is the mean net benefit calculated from PSA results. (65) The values of C_f and C_v for RCTs were estimated using the NIHR Open Data Platform. I extracted the observations on mental health RCTs with the monetary value of grant awarded and the planned sample size and fitted into a simple linear regression $Grant\ amount = \widehat{C_f} + \widehat{C_v} \times sample\ size$, to get the estimate of C_f and C_v. The ENGS-maximising n is then the optimal sample size for the proposed study.

In this chapter I conducted the VoI analysis for 2 scenarios: the base-case scenario and a hypothetical scenario. For the base-case scenario I used the PSA results from 4.3.5, along with its 60 input parameters for the VoI analysis. I also used hypothetical PSA results presented below alongside the corresponding parameter inputs as a demonstrative case for the hypothetical scenario. (Figure 5-1) As shown in the Chapter 4, the PSA results showed that it is highly unlikely that the RADAR intervention is going to be cost-effective at £30,000 threshold, it is anticipated that the EVPI would be close to zero and preclude the need of calculation other quantities (EVPPI, EVSI or ENGS). In this Chapter, I used the PSA result in Chapter 4, and the PSA for a hypothetical scenario to calculate the VoI for these two different scenarios to demonstrate the possible usage of the VoI analysis. In addition to the PSA parameter distribution for relapse in the discontinuation group that was derived from the meta-analysis, a different PSA was run using a vague uniform distribution Unif(0, 0.03) to account for the scenario that the probability distribution for the

relapse risk in the discontinuation group is still with greater uncertainty. This leads to 34.2% of the strategy being cost-effective under a £30,000/QALY WTP threshold. Figure 5-1 illustrates the results for this PSA.

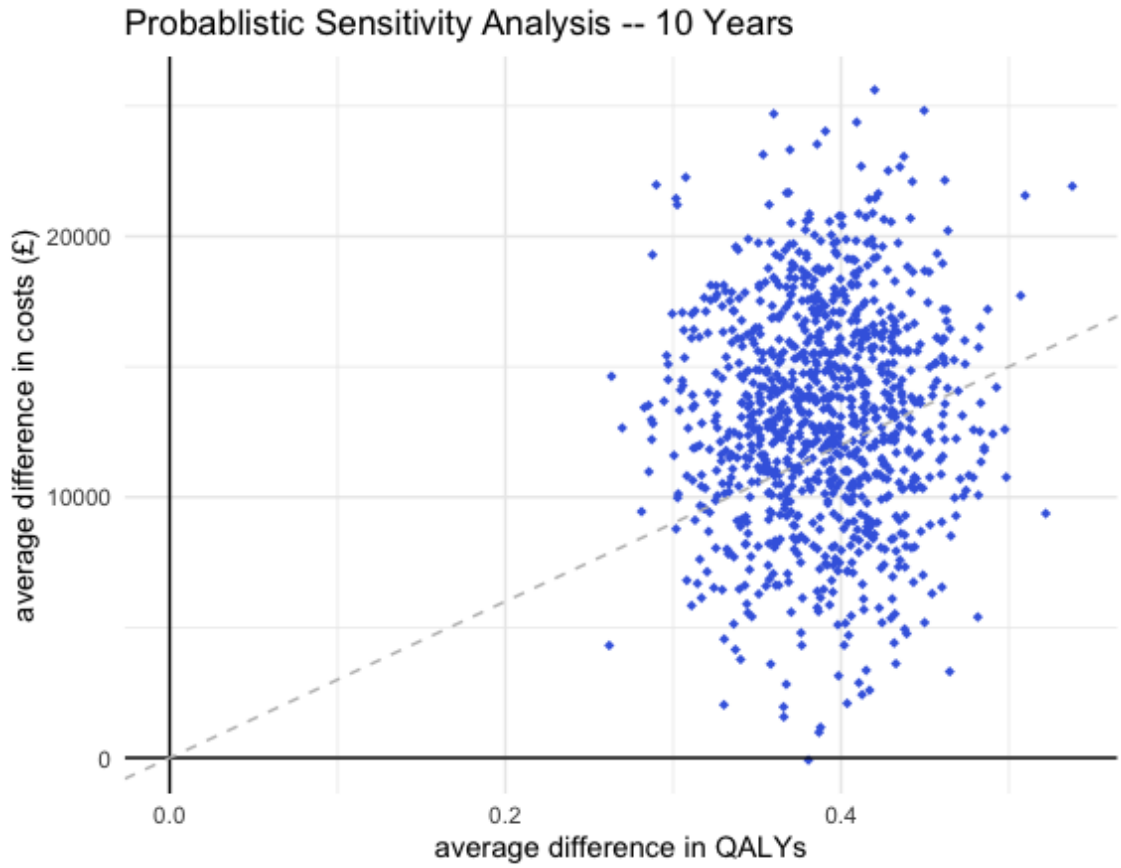

Figure 5-1 PSA results for a hypothetical scenario using Unif(0,0.03) for relapse probability

5.3 Results

5.3.1 EVPI

Figure 5-2 illustrates the EVPI based on the base case PSA results. As shown in Chapter 4, the cost-effectiveness of the RADAR intervention was dominated with little uncertainty at £30,000/QALY WTP threshold using 1000 PSA iterations. Therefore, based on the 1000 PSA sample, the estimated EVPI under this threshold is approximately zero under WTP up to £50,000/QALY (BCAEWeb evaluates EVPI up to thresholds of £50,000/QALY). Further analyses on EVPPI and EVSI in this case are not deemed necessary.

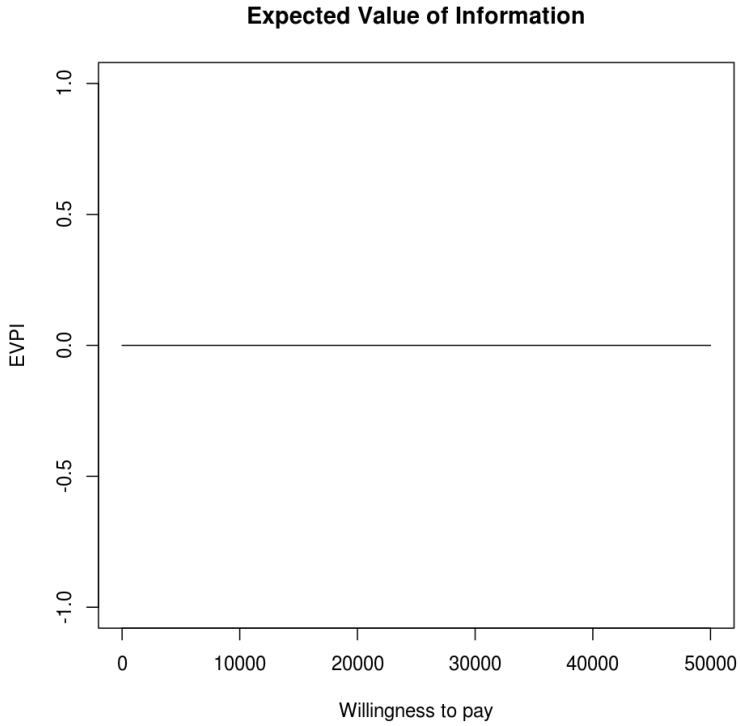

Figure 5-2 EVPI in relation with WTP based on the PSA results

On the other hand, using the hypothetical PSA result, at a £30,000/QALY gained WTP threshold, the overall EVPI per person affected by the decision is estimated at £1084.70 per person. This is equivalent to 0.03616 QALYs per person in decision uncertainty when valuing uncertainty on the QALY scale. With current annual numbers of patients with schizophrenia, which are to be affected by this decision, a total of 210,450 according to McCrone *et al.*, the overall EVPI per year is therefore £228.28 million. (96) Figure 5-3 illustrates the relationship between the EVPI per patient and WTP threshold. The EVPI per patient increases as the WTP threshold increases, and reaches maximum of £1,714.19 when the WTP threshold is at £34,000. Then it gradually declines as the WTP threshold continues to increase because the decision uncertainty diminishes as WTP threshold increases. From this point on, results of EVPPI and EVSI in the rest of Section 5.3 are based on the hypothetical PSA results in order to show how VoI may be done.

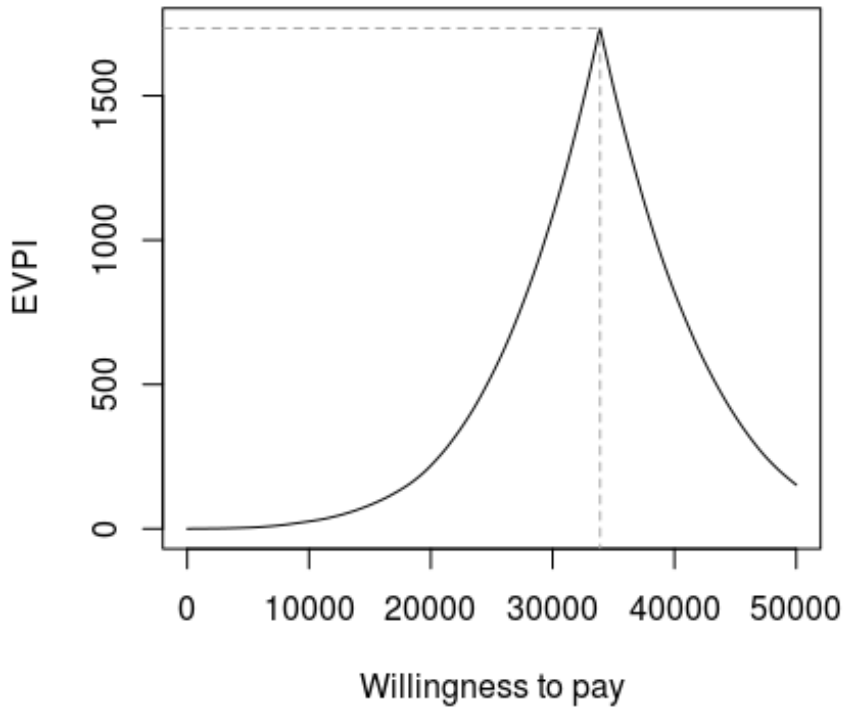

Figure 5-3 EVPI relations with changes to WTP in a hypothetical scenario

5.3.2 Single parameter EVPPI

Single parameter EVPPI was estimated using the BCEAweb algorithms which include GAM, GP regression and SPDE-INLA. The resulting EVPPI describes the value of learning about a particular parameter in the model described in Chapter 4, while all the other parameters remain uncertain at current level of knowledge.

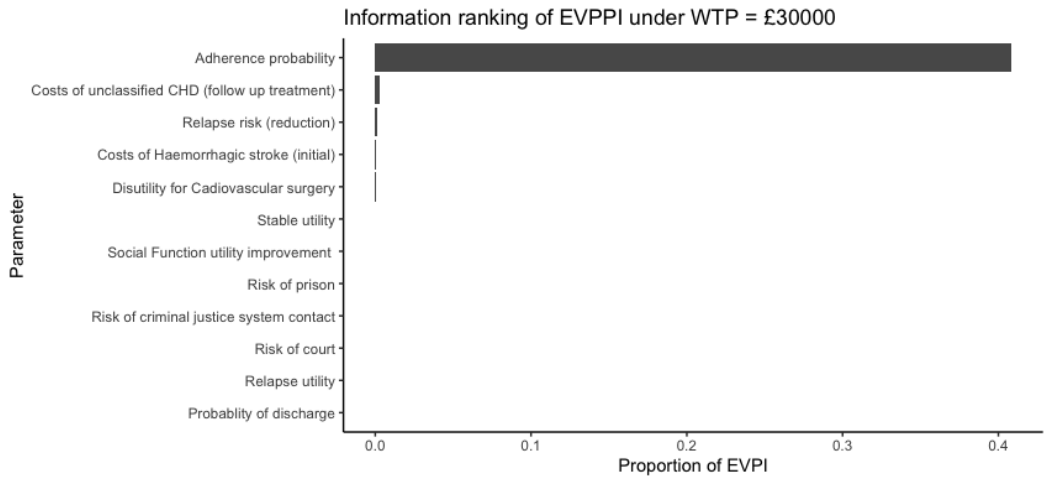

Figure 5-4 Information ranking of single parameter EVPPI

Figure 5-4 shows the ranking of EVPPI of each parameter in terms of its proportion of total EVPI. Under the WTP threshold of £30,000/QALY gained, probability of full adherence takes up 40.8% of the total EVPI, yielding EVPPI of £93.2 million, followed by the cost of unclassified of CHD, which take up 0.24% of the EVPI (EVPPI = £566,000) and risk of relapse in the reduction intervention, 0.1% of the total EVPI (EVPPI = £215,000).

5.3.3 Group parameter EVPPI

EVPPI for groups of all transition probabilities, utilities, and costs parameters were calculated using GP regression and SPDE-INLA algorithm. All transition probabilities and risk parameters make up 48.00% of the overall EVPI, resulting in £109 million. EVPPI for cost parameters are 27.5% of the overall EVPI, around £62 million. EVPPI for the utility group only takes 0.4% of the overall EVPI, translating into £8 million. Figure 5-5 illustrates the EVPPI information ranking of 3 different parameter groups.

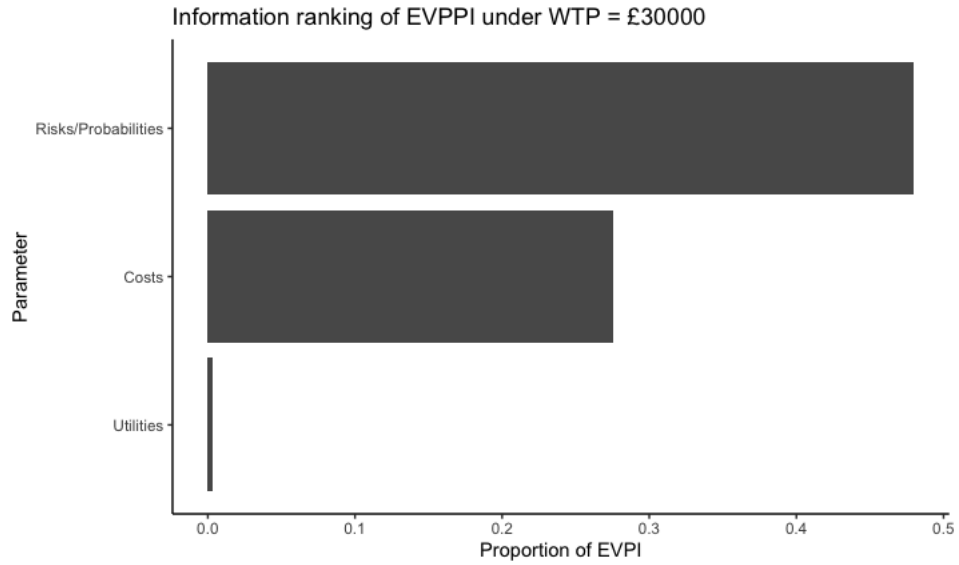

Figure 5-5 Group EVPPI proportion and rankings

5.3.4 EVSI and proposed study designs

5.3.4.1 Estimation of C_f and C_v

A total number of 42 trials in mental health were identified on the NIHR Open Data Platform, using the filters of funding stream "HTA Clinical Trials and Evaluation" and UKCRC health categories 1 & 2 "Mental Health" (https://nihr.opendatasoft.com/). Six were excluded because they were either meta-analyses, or cohort studies. The planned sample size of each trial was obtained in the ISRCTN Registry with the corresponding registration number. A total of £48,229,750 was awarded and a total of 12,590 participants were included in 36 trials. The point estimate for C_f is £1,115,841.3 and £640.1 forC_v. Table 5-1 estimates for fixed trial cost and variable trial cost summarises the estimation results.

Table 5-1 Estimates for fixed trial cost and variable trial cost

Variable	Mean	Standard Error
C_f	1115841.3	226295.2[15]
C_v	640.1	600.0

5.3.4.2 A single arm trial to investigate the probability of adherence with antipsychotics maintenance treatment

Amongst all the input parameters for the decision model, full-adherence probability shows the most value that could be gained via future research. Here we propose an observational study to investigate the probability of being fully adherent with antipsychotic medication. A binary outcome of fully adherent with the prescribed medication could be collected monthly, with 12 months follow up. A sample of parameter X is generated from binomial distributions, with probabilities of adherence generated in the PSA input and different sample

[15] P<0.0001

sizes N, ranging from 10 to 1000 in steps of 10. Then the NMB is regressed on X, which is the summary statistic for X. Figure 5-6 shows the EVSI (red dashed line) and ENGS (black solid line) estimations. Using the previous results on estimation of fixed cost of trial, C_v £1,115,841 and recruitment cost C_v of £604 per patient, the maximum ENGS yielded a total of £91,372,999, where the sample size is 510. As a single arm observational study can also (perhaps better) be in a form of a survey, the cost of setting up such study could be significant lower compared to setting up an RCT. Assuming £50,000 for the cost of setting-up gave the same ENGS-maximising sample size of 510, whereas the ENGS got maximum of £92,438,840. Further reducing the setting-up cost showed that the optimum sample size is insensitive to the setting up cost, giving the same result. When recruitment costs increase into £2,000 per patient, the optimal sample size reduced to 380, giving a maximum ENGS of £90,770,940. When recruitment costs reach as low as £300 per patient, the optimal sample size raised to 670, which obtained from a maximum ENGS of £91,557,166.

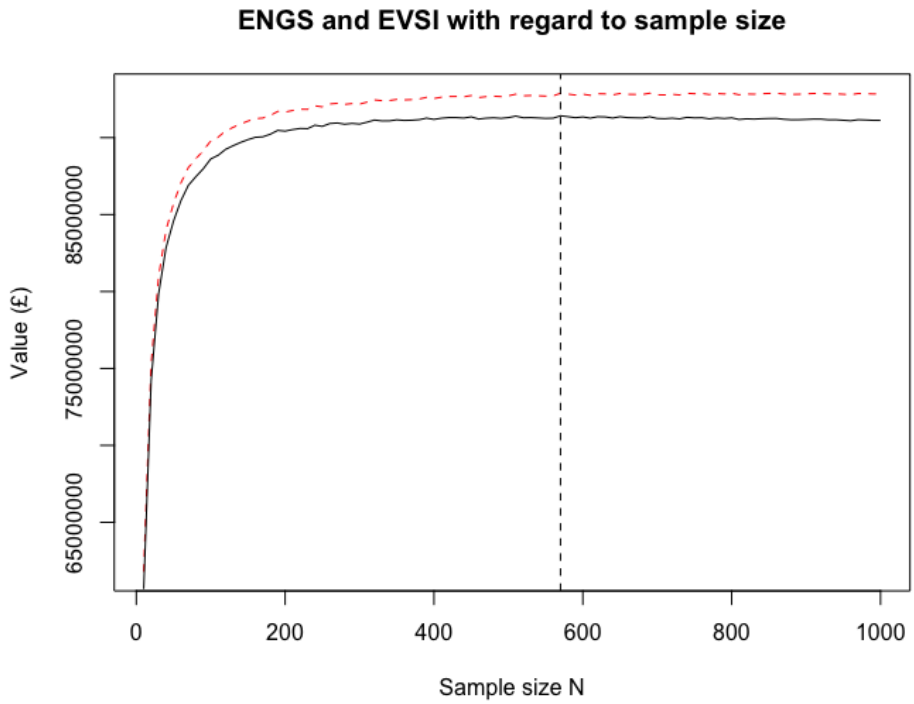

Figure 5-6 EVSI (red dashed line) and ENGS (black solid line) estimation for a proposed observational study on adherence with regard to sample size N and optimal sample size (vertical dashed line)

5.3.4.3 A parallel group randomised trial for serious relapse rate between the two interventions

As the results in 5.3.2 indicate, uncertainty in the relapse risk in the reduction intervention accounts for 0.1% of the overall EVPI. A study to compare rates of serious relapse event between antipsychotic maintenance and the reduction strategy within a 2-year timeframe could be proposed, although the overall EVPPI of the research will not be high. A transformation from monthly to 2-year probability was performed using the function $1-(1-p)^{24}$. Samples of the parameters $\mathrm{X}_M, \mathrm{X}_R$ are generated from binomial distributions, with probabilities of relapse for each of the two different interventions generated (M for maintenance and R for reduction) in the PSA input, and different sample sizes N, ranging from 10 to 3000 in steps of 10. The NMB then is regressed on tensor product smooth (GAM function in the mgcv R package) of the two summary statistics, $^{X_M}/_N$ and $^{X_R}/_N$. Figure 5-7 shows the EVSI of the proposed parallel

group RCT with regard to sample size of each arm N. As the EVPPI for the relapse probability parameter is only a fraction of the overall EVPI, EVSI of the proposed parallel group study did not increase as the sample size did. When taking into account the cost of sampling, with variable recruitment costs using previous results of £640 per patient, and £1,115,841 per trial for the fixed costs of running a trial in the UK, the ENGS of the study became negative regardless of the sample size. The maximum value of ENGS for the affected population (schizophrenia) is -£890,742, at a sample size of 10 for each arm, indicating that the research is pointless from a cost-effectiveness point of view.

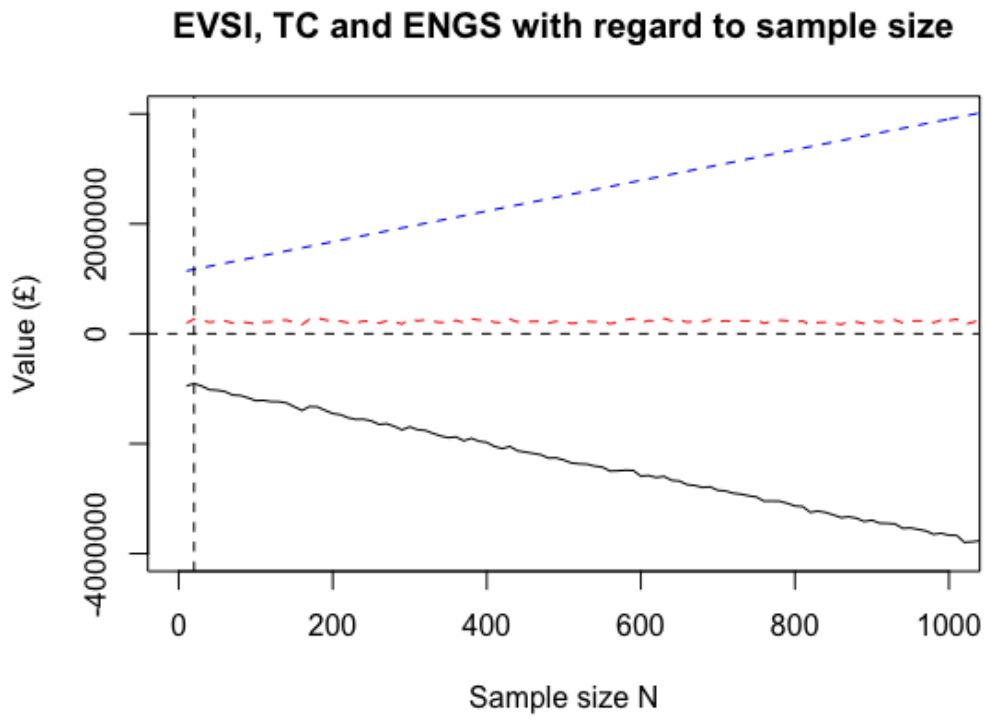

Figure 5-7 Sampling costs (TC, blue dashed) EVSI (red dashed) and ENGS (black straight) estimations for a proposed parallel group RCT on relapse event rates between maintenance and reduction treatments with regard to sample size N per arm

5.4 Discussion

5.4.1 Scenario in Chapter 4

For the PSA results from Chapter 4, the overall EVPI is £0 up to £50,000/QALY WTP. This implies that no further research is needed to reduce the uncertainty

in the adoption decision for the antipsychotic reduction and discontinuation strategy. In this scenario the EVPI is zero within an appropriate WTP range (up to £50,000/QALY). As the WTP threshold increases to very high values, the curve which looks flat at zero will rise. This however will not change the conclusion that is reached here as calculating the EVPI at a very large WTP threshold is not useful. Hence the main discussion point from here on is based on the hypothetical scenario. As the hypothetical scenario is not from the base case PSA, the paragraph below only shows a demonstrative case of how the results from Section 5.3.3. and Section 5.3.4 may be interpreted hypothetically and does not serve as a piece of evidence to inform the future research into RADAR programme.

5.4.2 Hypothetical scenario

For the hypothetical scenario, the overall population EVPI was £228 million, indicating a substantial value of future research regarding the antipsychotic reduction strategy. The probability of real-world full adherence to the medication was a key parameter, whose EVPPI took up more than 40% of the overall EVPI. Future research on this could be particularly important to the RADAR trial to the extent that non-adherence to antipsychotic treatment in the real world was likely to be more prevalent than that in the trial. The point of the RADAR trial itself was to provide a “reduced adherence” environment, in which the intervention offered active monitoring and support when patients reduce their antipsychotics, as compared to the real world, where patients might stop on their own without supervision, which might result in a series of withdrawal effects without being carefully and actively monitored and could incur higher cost to the NHS and society as well as greater distress to the patient. Indeed, previous evidence had

suggested that taking antipsychotics is often associated with unpleasant side effects and the non-adherent (discontinuation) rate is high. (15)

The computational burden of EVPPI and EVSI has long been a major barrier to using these quantities to inform decision making. Strides to improve the computational efficiency of EVPPI and EVSI by avoiding running the economic model for multiple times rely on using regression techniques including GP regression, GAM and SPDE-INLA. (183-185, 187) These methods enable relatively fast calculation and accurate estimation. In this chapter, the methods we used to compute EVPPI, SPDE-INLA on the BCEAweb took only seconds to generate the EVPPI results for each single parameter. As for EVSI, 500 iterations by the GAM regression methods took 1 hour on an Intel i7 laptop with 16G RAM. These methods ensure the wider applicability of these methods for future research project planning. However, there have not been a gold standard to compare the results of using these fast calculation methods. The conventional calculation could take up to months, even years, making it almost impossible to make comparisons. It is also worth bearing in mind that any of the fast calculating methods used in this analysis (BCEA, SAVI) have not been thoroughly stress tested, hence using the results require some caution. Cross validation using different algorithm can be helpful to ensure the accuracy of the results.

The EVSI and ENGS results for the single arm study (the hypothetical scenario) to investigate antipsychotics adherence rate amongst patients with schizophrenia showed an optimal sample size of 510 as it is the N that gives the maximum value of ENGS. However, Figure 5-4 showed that the ENGS value

starts to plateau from sample size of around 250, after which the magnitude of increase is only moderate. This is due to some of the simulation error which could give different results at each time of the simulation. It is therefore appropriate to ascertain a sample size at which the ENGS curve starts to plateau. In this case a sample size of 250 should achieve a very similar ENGS to the sample size that gives the actual maximum ENGS values in the simulation.

The value of investigating further into the antipsychotic adherence in real world, is to show what the RADAR intervention will be comparing to in reality. Different from randomised into the maintenance arm in the RCT, patients in real world may receive far less support and supervision when they wish to reduce and discontinue the medication. The impact of unsupervised withdrawal, which is regarded as non-adherence, may induce serious symptom exacerbation. In this case, a proposed study to reduce the uncertainty around real world adherence rate can be informative to how the potentially more appropriate comparator to RADAR appears, hence there will be less uncertainty of the cost-effectiveness under the current WTP threshold.

When calculating the EVSI for the proposed RCT with series relapse rate outcome, according to the methods proposed by Strong *et al.* for a parallel arm RCT with binary outcomes, the summary statistics showed were the log odds ratio of the two parameters sampled from binomial distributions. However, in this case, the risks of relapse with each of the two treatments are small and led to many zeroes appearing in the simulation if the proposed sample sizes were insufficient, which means that calculating the regression fitted value is

impossible as infinite values are generated after taking the log odds ratio. I tried using a different value, the proportion of patients who relapsed in each group, as the summary statistics, and regressed on the tensor product smooth function of the two. As the GAM regression was only used for its fitted regression values, the appropriateness of this approach should be appropriate.(183) The result of 10 participants per arm (total 20 participants) were not comparable to the standard sample size calculation with 0.05 type-I error rate and 0.1 type-II rate described in the trial protocol, which gives a sample size of 372. In the hypothetical scenario, the EVSI analysis showed no values of investing further into reducing uncertainty in relapse rate in the intervention arm, which is a key question to answer in the RADAR trial, whereas investing in real world antipsychotic adherence has substantial value in the possibility of altering decision on adopting this intervention. Because the weight of EVPI largely concentrates on the non-adherence parameter of the long-term antipsychotic maintenance strategy and the differences of relapse rates between arms only accounts for a small fraction of the EVPI. In the EVPPI analysis, adherence also takes up the majority of EVPI, resulting in the largest potential for uncertainty reduction by future research. This mirrors the real-world scenario, where patients with antipsychotic medication often fail to comply with their prescribed medication, which has been suggested associated with increased risk of symptom exacerbation, incurring considerable costs to the NHS as well as the society. (173) As for the case of RADAR, in which patients consent to adhere to the protocol as well as the intervention, the chance of non-adherence in the trial should be lower than in the real world. Patients often find long term antipsychotic treatment unpleasant, and discontinue without any supervision or supported care environment. Future research on this could be done via survey

to collect monthly adherence data, the cost of which should be taken into account to calculate the expected net gain of sampling. This study may also have been able to imply the evidence base for patients with mental health conditions who are on long-term antipsychotic maintenance treatment as adherence has been rarely studied in a relatively large cohort to obtain robust estimates for the adherence outcome. This ideally may provide more precise estimates so as to reduce the uncertainty of decision under £30,000/QALY threshold. Meanwhile, as the decision model in Chapter 4 was a pre-trial early model, hence the VoI may also be informative as to what the RADAR trial programme may need collecting to reduce this uncertainty. This VoI analysis for the hypothetical scenario meant that in addition the trial, a single-arm observational study in a real-world setting may be worthwhile to reduce the uncertainty for the decision. The adherence monitoring between RCT and real-world settings are very distinctive, hence proposing a separate observational study is more appropriate than informing the design perspective of the RADAR trial itself. On the other hand, if ENGS for a different, trial-related parameter had been meaningful to calculate, it could be seen the optimal sample size for the RADAR trial from a VoI point of view. This would work well to the power calculation as discussed previously, the procedure of which often follows a rule of thumb of the average cost of recruiting one patient and the overall budget of the research project. If the ENGS sample size required were smaller than the one using power calculation, potentially there may be resource saved for not having to over-recruit.

The single parameter EVPPI calculations via SAVI and BCEAweb tools showed good agreement between each other. However, when grouping the parameters

for group-EVPPI calculation, the level of accuracy still requires improvement. There is a trade-off between computation time and accuracy of the estimation when performing EVPPI calculation in using these user-friendly web tools. The results to an extent could be informative to the decision making. The overall EVSI and ENGS framework may also be useful alongside the interim analysis as an economic consideration to terminate the trial early for safety, efficacy or futility. For instance the results in 5.3.4.3 could be used during the interim analysis of RADAR as an argument for futility in addition to the regular interim analysis, as recruiting more patients would not lead to an increase on the probability of altering the decision of adopting this intervention under the could be used under the £30,000/QALY threshold.

This chapter showed that at £30,000/QALY threshold, there is no need for further study to investigate how uncertainty of decision may be reduced for the CEA in Chapter 4. Based on hypothetical demonstrative VoI analysis, a future single arm study to investigate the adherence to antipsychotics on patient with schizophrenia may benefit the antipsychotic reduction and discontinuation strategy for reducing decision uncertainty on cost-effectiveness. The ENGS showed that an optimal sample size of 510 patients may be recruited, when considering the balance between the benefit of reduction of uncertainty and the costs of setting up a new study, under this hypothetical scenario.

In this chapter I used the VoI approach to discuss how decision uncertainties may be reduced by conducting more research and more importantly, what research may result in better efficiency in terms of the benefit gained. The EVSI quantity could be informative to trialists who also consider the economic benefit

of conducting a trial. This could potentially be used in line with the standard sample size calculation to both ensure generalisability and economic benefits of the trial. In the next chapter I will summarise and discuss the findings from previous 4 chapters, and highlight some of the area that could be explored in the future.

6 Chapter 6 – Discussion and conclusion

In previous 4 chapters I introduced a number of studies which could aim to improve the efficiency of RCTs. In this chapter I will summarise and discuss some of the main findings of the topic based on previous chapters, and highlight future research area which may be worth exploring.

6.1 Summary of main findings

Evidence on the efficacy and safety of reducing and discontinuing antipsychotics in patients with psychosis has not been broadly developed. However, as antipsychotics are, in the beginning, not designed for the purposes of treating people with psychosis, the evidence of its 'efficacy', therefore, only stands in terms of controlling the symptoms. (190) Hence the aim of treating patients with psychosis should not only be limited in controlling the symptoms, let alone achieving this by altering the brain chemistry, but also more long-term health benefits, the RADAR trial was set up to provide good estimates for the risks and benefits.

In terms of the hierarchy of medical evidence, RCTs are often considered the highest level of proof experimental design for the reason that randomisation deals directly with confounding, which is ubiquitous in other non-randomised research designs. (23) Despite the merits of providing unbiased statistical estimates for the treatment effects (if properly conducted), RCTs often get criticisms for their lack of applicability to the real world clinical practice. In order to maintain a high standard of scientific validity, many RCTs require narrow eligibility criteria, specialised providers and settings and high treatment adherence and exclude patients with major co-morbidities, all of which deviate

from the real world scenario, bringing the generalisability of the result into question. (23)

Meanwhile, the cost of setting up an RCT is also significant, which can also explain the use of different stages for trial designs (Phase I, II, III and IV, in which the sample size gradually increases in most cases) in a progressive way, in addition to the paramount safety question of the intervention. The next phase can only be recommended if the current phase shows convincing results in terms of efficacy, safety, and feasibility. Certainly, the argument above manifests itself in an industrial environment, for instance, an innovative therapeutic molecule or biologic product, as industrial development programmes should always consider and weigh up the investment and return on clinical development. In the case of publicly funded research programmes, where complex interventions often cover a significant part, the evidence base may not be as progressive as it is in the industry. Evidence may be sought in various places and there may not be as formal development “stages” to inform what the most appropriate experimental design would be, although there has been a proposal of a similar framework. (23) In the case of RADAR, the design process largely depended on evidence from previous case series and RCTs on similar interventions from outside the UK. (67)

Recruitment often plays a major part in determining the feasibility and reflecting whether the appropriate trial design is applied. Only 56% of the publicly funded RCTs between 2004 and 2016 reached their targeted sample size, highlighting the challenge in recruitment. (35) Methods and strategies that hope to improve recruitment were discussed in Chapter 2. The current exercise of assessing the

effectiveness of recruitment strategy contains observational and interventional. The former records any specific strategies which were used to recruit whereas the latter involves a more pre-specified design process. Planning an interventional randomised comparison of different recruitment strategies requires previous evidence or insights (which are often observational) on what might work for a particular trial based on the population, intervention and design. It is common practice to apply different strategies in an *ad hoc* manner and the choice of strategies may hugely depend on the trial's characteristics. Although randomised SWATs provide unbiased estimate for the effectiveness of the strategy, one superior strategy night not apply to the current trial simply due to minor difference in the trial characteristics. Conclusions on the superiority of certain recruitment strategy ought to be made with caution, even with the evidence from a randomised SWAT.

SWAT can provide unbiased estimate of interventional recruitment or retention strategy. On the other hand, however, this design sometimes may inevitably be fraught with some pitfalls. For instance, insufficient power can easily become an issue in SWATs for retention strategies, as the sample size have already been powered to the primary outcomes in the host trial, but may not be enough to test a meaningful difference in retention rates. Particularly, by randomising participants into 2 different strategies during the follow up period would inevitably lead to "more" loss to follow up in the inferior strategy. Therefore it may be advisable to use complex strategies such as adding new additional strategies to the standard practice, or using allocation ratio different from 1:1 in favour of the hypothesised superior strategy to ensure loss to follow up is within

acceptable amount, although the latter would inevitably lead to a reduced efficiency.

Recruitment SWATs often need a sample size larger than that of their host trial. (52) But it is often unknown what the success rates for the comparing recruitment strategies are in a SWAT, making the approximate sample size of the recruitment SWAT difficult. In order to recruit successfully to the planned samples size for the host trial, this may take longer compared with recruiting using a superior strategy alone. However on the current level of evidence base, it is not yet clear what strategies work for what particular group of patients or what kind of interventions, it can be very helpful to lay the fundamentals knowledge about the recruitment methodologies in RCTs.

Also, from the results in Chapter 2, it appears that majority of the mental health conditions amongst the included studies were more common mental health conditions (depression or anxiety related). Different strategies such as recruiting via different advertisements, or via different health practitioners were used. Interestingly, in my observation, none of these strategies were used during RADAR recruitment. It may be explaining that patients with psychosis are frailer and have less access to media such as internet. Due to their severe condition, more intensive and specific recruitment methods are needed. Examining psychiatrists' case load and outreaching to psychiatrists and patients are used intensively during RADAR's recruitment. It reflects the design and nature of recruitment strategy and understanding the motivation for patients to participate. Relations between research team and clinicians, as well as patients are essential to recruitment's success in this case and qualitative work in

advance may help to understand the cultural cause of patient's willingness to participate into clinical research. This is partly due to the characteristics of mental health patients, particularly those with psychosis, who may be challenged by the cognitive demands of the informed consent process when enrolling in trials. (191) Although it is suggested that patients with psychosis still retain a level of understanding of risk of benefits, efforts to improve recruitment by facilitating the understanding of consent do not show significant effect on recruitment efficiency. (80, 192) While Robert *et al.* suggested that patients with schizophrenia and psychiatrists could make discerning risk assessments and participating decision with hypothetical protocols in which variable design elements were introduced, this could imply that patients' perceptions ability may be undermined. (193) Literature also suggested that during consent process, many participants have reported unwillingness to be randomised. (194) Summer *et al.* discussed the reasons for participation or non-participation for a trial in psychological therapy in amongst patients with SMI, suggesting that a number of interviewees have expressed preferences on treatment allocation or disliked the nature of randomisation. Concerns of being randomised into the unfavourable arm (often placebo), or worries of symptoms worsening during washout phase are also important reasons for non-participation. (50) Hemminki *et al.* compared recruitment with an open design with a blinded placebo-controlled design and found open design helped increase recruitment by avoiding resentful demoralization. However, the pitfall of an open design is that it would inevitably introduce a certain degree of bias as patients know beforehand which treatment they will be receiving. (51) Indeed, many patients who are willing to participate in a trial because they are offered a chance of receiving the "new" treatment. Consequently, patients may inevitably favour the

new, as they perceive that the intervention to be in a way superior to the control, which is not always the case. Sometimes patients will even take the risk of accepting randomisation, despite the fact that there is an (most usually) equal chance of being allocated into a treatment or controlled arm, the latter of which they may perceive as a rejection. Investigators try to minimize selection bias using inclusion and exclusion criteria so as to ensure the generalisability. Judicious and reasonable criteria are essential when balancing recruitment challenge and generalizability. Khan *et al.* explained under an intensive recruitment how the usual inclusion and exclusion criteria often used in antipsychotic trials narrowed down the eligible population from 3.6 million to a mere 655, by applying each item on the selection list. (195) Preskorn *et al.* discussed a similar situation in antidepressant trials to see how each criterion selected a population, with a broadly inclusive criteria. (196)

On the other hand, recruitment difficulties also exist from a recruiter or trialist's perspective. Recruitment via primary care faced significant difficulties as some GPs worry that referring their patients to mental health trials will affect doctor-patient relationships. A great number of GPs consider patients with mental health illnesses are vulnerable, which leads to their protectiveness towards their patients.(70) Some GPs believe that introducing their patients to clinical research may give patients more stress and could deteriorate their health. (70) But do their GPs get to decide how much psychological stress is too much for anyone to bear? Or indeed what level of capacity would one need to have to demonstrate that one knew that this is what one was being asked to do? (197) For GPs who possess an interest in clinical research, the case is also difficult when it comes to the practicality of selection criteria. Jenkinson *et al.* found

noticeable variations amongst GPs in the way they put into practice a mental health trial's selection criteria particularly with regards to comorbidities. (198)

Patients participate in research are based on autonomy and altruistic, understanding that research is to benefit the larger group of population and society. Everyone has the right to participate and no one should be excluded due to a disability. This is particularly an issue in patients with SMI, whom are often considered have impaired capacity to consent. (197) But it should not be the reason to exclude them in research, or to include them in absence of consent. RADAR trial has made substantial effort in order to support, augment and respect an individual's decision-making autonomy during the informed consent process, in order to improve recruitment outcome.

Improving recruitment in mental health trials requires significant efforts in various respects and extensive collaborations. Helping potential participants to correct their misunderstandings of RCTs in a more encouraging manner would make the more discerning in participation. Having more comprehensive criteria would help GPs screen patients more confidently for inclusion. (198) Recruiting patients with mental health conditions may be more likely to be successful from many sites in a short period than at a few sites over a long period. (199) Certainly, researchers and trialists would use various strategies to promote available trials for a wide range of potential participants.

There is abundant literature discussing factors that affect attrition in clinical trials. Participants' age, gender, level of education and ethnicity are suggested to be predictors of dropouts. Condition-specific factors appear to have strong relation with co-morbidity and health states. (200) However, a decision on

altering inclusion criteria so as to achieve higher retention needs to be made with caution. One of the statistical rationales underlying experimental design is that the sampling should be an appropriate representation of the study population in order to draw a conclusion based on the findings. Not only can altering inclusion criteria affect a trial's external validity and generalisability, it may also inevitably increase recruitment difficulty by introducing extra restrictions on the eligible population. It seems, looking at the question as a whole, more reasonable to encourage participants' retention to clinical trials rather than changing study design, which ought to be determined only by the clinical questions in hand.

Similar as strategies to improve recruitment into RCTs, most of the retention strategies discussed in the literature aim to counteract the operation and logistic complexity that modern trials engender. Brueton *et al*. summarised retention strategies into 6 categories in a review on retention strategies which focus on all healthcare, including incentive strategies, communication strategies, new questionnaire strategies, behavioural or motivational strategies, case management and methodologies strategies. Postal or electronic questionnaire response during follow up period were the main measure of retention, however strategies to encourage participants' return to trial sites were lacking. Often in a form of voucher, incentive was widely used to encourage questionnaire response and suggested to be effective and cost-effective. Communication strategies in different forms were also discussed. However, the effects varied between each specific strategy, suggesting these methods should be used with more care. Strategies such as postal questionnaires may no longer be suitable

for certain populations while other strategies may be laborious or require substantial monetary input. (78)

Another interesting finding in Chapter 2 is the obscure reporting of costing on recruitment. Alongside advocates for transparent and comprehensive reporting on all trial costs, recruitment costs sometimes may be challenging to pin down as co-costing or shared cost can happen within a research project. Even though, it should be considered good practice to report detailed methods of collecting empirical data. For instance, if recruitment involving a research assistant, then the detail of dedicating into recruitment tasks should be specified (full time or part time, approximate number of hours spending on recruiting, etc.). The included studies in Chapter 2 could date prior to 2000, the difficulty of ascertaining the costing becomes more challenging.

In Chapter 3, I reviewed existing decision analytic models on antipsychotics for patients with schizophrenia and schizoaffective disorders, the review mainly focused on the study design characteristics of the decision modelling and the essential clinical events considered in this population. The choice of modelling designs varies and depends on clinical questions and data available. In Chapter 3 a collection of modelling technique was identified in the included studies, however cohort Markov models and decision trees still make up the majority of the design. With the advancement of computational capacity and data storage, more complex and stochastic models have become feasible. DES and patient-level simulation are more computationally costly methods however more suitable to handle populations with considerable heterogeneity. Pharmacological intervention on patients with schizophrenia comprise the

majority of included studies, which have the main aim of the safety and efficacy profile of specific antipsychotics. Parameter inputs in these models often come from corresponding RCTs of antipsychotics, such as the CATIE trial, in which patients were either antipsychotic-naïve or switching to a different one being investigated. These studies were helpful in identifying the nature of antipsychotic intervention on patient with schizophrenia in terms of their benefits and risks. Patients with schizophrenia who are under long-term antipsychotic treatment may already have developed some of the co-morbidities (such as weight gain, or diabetes) hence the clinical context is thought to be different. Besides summarising the relevant clinical events and potential social costs driver for the population and utility values for each health state, the model input had little contribution to the design in Chapter 4.

Decision analytic models essentially are trial simulation that provides the most possible ‘guess’ without the actual data collection practice (observational studies or randomised trials). Information which parameterises the decision model is essentially from evidence in the existing literature or sought from expert opinion. As for the decision analytic model described in Chapter 4, one of the pivotal parameters, relapse risk, had little previous evidence. However, the definition of relapse varied in previous RCTs, therefore cannot be used directly for the model. The relapse risk (hospitalisation) in the meta-analysis of antipsychotics in preventing relapse in a way, resembles the nature of the antipsychotic reduction and discontinuation strategy. The assumption made with relapse risk monotonously increasing as the antipsychotic dosage decreasing originated from previous evidence that suggested that non-adherence essentially has similar characteristics to antipsychotic reduction to

the extent that patients are not taking the originally prescribed antipsychotic dosage. (164) The decision model did not show cost-effectiveness at 10-year. The time horizon was longer than that was seen on average in the published decision models discussed in Chapter 3. This is due to the inclusion of CJS in the model, which could take relatively longer than 5 years to model the average costs. Besides, different from the antipsychotic cost-effectiveness model described in Chapter 3, the population being modelled in Chapter 4 were on long-term maintenance treatment, hence in addition to short term effectiveness, long term safety and outcome improvement was also considered key to the intervention. Meanwhile, one of the main limitation of the CEA in Chapter 4 was that the model was based upon relapse, which is not the primary outcome for the RADAR trial, improvement in SFS. This was only conducted as a scenario analysis as there was not sufficient data to translate SFS into health related quality of life measures, particularly for this population. The RADAR trial has a planned trial-based economic evaluation at the end of the trial. This could provide a more holistic view on how the intervention compares with the maintenance strategy, making use of the main trial results and centring around the primary outcome, rather than previous related evidence which is subjected to considerable uncertainty. In addition, as RADAR is an RCT investigating a complex intervention in nature, the question of whether evidence of a purely pharmacological perspective is appropriate to use as input for the key parameter for effectiveness may be asked as a result of the fundamental differences between the interventions. In complex interventions a whole array of therapeutically active elements may be operating simultaneously and synergistically, and it could be far more complex than trials on pharmacological interventions. It is often standard practice that early phase trials may provide

preliminary evidence for certain parameters to support the design of a pivotal trial. On the other hand the evidence development for complex intervention has not been as straight forward. Although NIHR had allocation for resource to support feasibility trials of complex interventions, the nature of the intervention still pose challenge on informing the design of larger trials. Reduction and discontinuation process could be relatively smooth and fewer cases of serious relapses may occur than one would fear, if the intervention is being delivered by clinicians who have undergone considerable training in psychiatry.(201) Expert opinion was also sought, with the overall relapse risk no more than 10% compared with maintenance treatment. However, in the sensitivity analysis, the relapse risk appears a major drive of the overall cost-effectiveness, along with parameters such as adherence probability, utility of stable state or relapse. The decision to adopt the antipsychotic reduction intervention, had it become more likely to be cost-effective, could be largely affected by the uncertainty of these parameters, which could potentially be reduced by conducting future research. The impact of future research could be measured in the economic term as to how much is worth spending on this population in order to reduce the decision risk. The 3 statistics of VoI described in Chapter 5 could provide such estimates, as seen in the demonstrative hypothetical scenario. EVSI could be used to inform on the value of reducing these parameter uncertainties. By using the EVSI together with recruitment cost, the ENGS can provide an estimate on the economically optimal sample size. This method may be compared in line with the orthodox frequentist sample size calculation approach based on the sampling theory. As it is often very easy to back calculate the power of certain clinical effect size based on the desired sample size, in practice it is not often very meaningful in determining the optimal sample size as it is always the larger

the better statistically, despite that there is often budget restrictions to recruit indefinitely. Then clinical research, especially RCTs, should set off more than just demonstrate comparative in characteristics, but also take into account how economically beneficial it might be, within a single payer system in which public research funding largely comes from taxpayers.

The RADAR trial investigates such an intervention that aims to remove a class of pharmacological intervention from a patient population. The NICE TA programme has mainly focused on the opportunity costs of investing in new interventions and acknowledged this in the cost-effectiveness threshold that NICE makes its recommendations based on. On the other hand, there has not been a clearly defined programme on when an intervention needs to be removed from the current health care budget.(202) The strategy that universal health care plan adopts is to maximise the health of the entire population under a constant, or constantly growing budget, there should be elimination process for the least effective, or least cost-effective technology.(203) NICE has started making disinvestment recommendations, however unlike adopting technologies, these recommendations is currently only advisory and not based on cost-effectiveness. A similarly structured disinvestment programme should be in place to help make decisions to support interventions such as RADAR. Given the limited published evidence for how disinvestment is carried out, let alone those were based on cost-effectiveness, if any, more evidence is imminent in this space to stimulate the discussion on what the ideal disinvestment threshold NICE should take, in comparison to the £20,000 – £30,000 investment threshold.

In Chapter 5 I took a series of VoI analyses to demonstrate how the future research areas which may reduce the decision risk of the antipsychotic reduction and discontinuation strategy in a hypothetical scenario. The VoI approach provides a way of foreseeing future research to reduce uncertainty of the parameter estimates. The value of future research, from a single payer's point of view, relies on the WTP for the proposed therapeutic intervention. EVPPI may be helpful to point out which parameters or group of meaningful parameters are worth more future research. EVSI takes this idea further by estimating the value of designing particular types of data collection practice (research project). This could potentially be a powerful feature to be incorporated in the planning stage of the research project. The EVSI less sampling cost (recruitment), ENGS, could be informative to the sample size determination in line with standard power calculation. Indeed, the sample size calculation often involves consideration for all different outcomes included in the trial, on which the trialist should make sure to demonstrate sufficient power. RADAR's power calculation considered not only the primary outcome, SFS, but also the safety outcome, serious relapse rate during power calculation. As recruitment cost (sampling costs) increases, the sample size for maximum ENGS decreases, resulting smaller sample to reduce uncertainty for decision making.

Furthermore, EVSI may provide a useful information in the DMSC. As data from clinical trial accumulate, not only can interim analysis inform trial results, emerged data may also carry value of reducing uncertainties from the EVSI's point of view. As more data being included in the analysis, the costs of recruiting more patients to reduce uncertainty of outcomes declines, meaning

more research participants will not necessarily change the decision of reimbursement under the current WTP threshold. This could be used in interim analysis in addition to the futility analysis, as an economic argument for terminating the trial early. It could be potentially meaningful to the value-based reimbursement decision bodies such as NICE, while reviewing clinical guidelines and HTA applications.

Costs for recruitment on the other hand, plays another very important role in ascertaining ENGS, which is the EVSI less the recruitment (sampling) cost. The optimal sample size maximising ENGS may not be sufficient from EVSI's point of view if expensive recruitment strategies are used. In Chapter 2 I discussed the cost-effectiveness of the recruitment strategy, which is an area of research expecting more work. Using cost-effective recruitment strategies would help with higher ENGS, which results in better approximation of the optimal sample size of reducing uncertainty of outcomes.

6.2 Qualitative work

The thesis mainly takes an analytical approach towards the RCTs and intervention of antipsychotic reduction and discontinuation. While this approach has many merits in providing reasonably accurate information to the question, its limitations also cannot be overlooked. Clinical trials, and more broadly clinical research, demand collaboration between researchers, clinicians and patients. Qualitative approach could be helpful to investigate the underlying reasons for some of the findings, and often beneficial and informative to carry out beforehand in finding the "why". (204) For instance, in determining the best optimal strategies for recruitment or retention in a certain population, qualitative studies have the advantage to find out the underlying behavioural cause and

opinions about patients participating in randomised trials. Previous works have investigated reasons for non-participation in clinical trials, and some strategies such as improving the clarity and friendliness of trial material, or using multimedia approach to introduce RCT, may have emerged as a solution to some of the issues. Similar approach could be applicable to retention in trials and strategies that attempt to tackle attrition problems may hugely contribute to the validity issue and ease the burden of developing missing data analytic models using extra assumption.

Also, the RADAR research programme may find it helpful to conduct qualitative studies as to the motivations and concerns of clinicians' and patients' participating in the RADAR trial, as well as the purported quality of life improvement by once reducing even coming off antipsychotic treatment in the long run. Qualitative evidence-based research may also shed light on the patients' views and experiences for the antipsychotic reduction and discontinuation strategy in terms of how the purported improvement in social functioning and wellbeing could be manifest.

6.3 Future works

Based on the current finding in the thesis, there are, in the meantime, a number of areas worth investigating in the future. First of all, the evidence for SWATs, especially recruitment and retention SWATs, have been burgeoning. Efforts to integrate the evidence for recruitment strategies have been initiated and progressed with more SWATs being now conducted and published. Outreaching psychiatrists and nurturing supporting relationships between research team and trial sites might be promising to improve recruitment with

appropriate SWATs design to demonstrate. Secondly, attention is needed in costing RCTs, especially recruitment as well as other conduct. Introducing clear financial statement on research projects may be suitable in recording costs and such information should be make accessible upon request. It could be beneficial with research where economic analysis is needed, for instance, cost-effectiveness of different strategies to improve recruitment and retention, or cost-effectiveness into other strategies which may optimise RCT's procedure and improve efficiency. Furthermore, a guideline on cost reporting of RCTs is expected in the future. This can help trialists and methodologists to identify recruitment and retention strategies that may be appropriate for a trial.

Furthermore, despite the need of reducing antipsychotics, nearly 25% of the patients with schizophrenia take more than one antipsychotic, and more so with patients who are under the condition for long time. Common reasons for polypharmacy prescribing include *pro re nata* (prn) use which often act as a temporary strategy and controlling aggressive behaviour using antipsychotic augmentation. However, the evidence has not yet allowed for any consensus or guideline endorsement for antipsychotic polypharmacy in routine practice, nor repudiated its possibility of being a favourable intervention for some clinical cases. Well-designed studies are called for to investigate the risk and benefit for combining antipsychotics and electronic health record should keep update with the prescribing so as to improve the quality of evidence base research in this area.

In Chapter 4, one of the limitations with the decision modelling is that the model does not reflect the purported improvement in wellbeing, as measured by SFS

in the RADAR trial. Mapping from SFS to QALY in SMIs using the appropriate mapping methodology is absent from the literature. Hence a mapping study may contribute to refining the decision model in terms of arguing for the benefit of reducing and discontinuing antipsychotics. Moreover, as the EQ-5D is criticised with lack of consistency amongst patients with SMIs, ICEpop CAPability measure for Adults (ICECAP) was used as measure of quality-of-life and mapped into QALY in the RADAR trial. The extent to which EQ-5D and ICECAP may be helpful in reflecting the utility and quality of life changes amongst this population could be studied in the future.

In Chapter 5, I used some most recent developed methods and web applications to calculate the VoI quantities. Despite the remit of enabling fast and easily accessible VoI calculation, there is still a lack of a gold standard to compare the accuracy of these methods referencing the traditional calculation methods. More methodological work is required on the estimation of these quantities. Meanwhile, similar to the EVPI and EVPPI quantities, a user friendly interface for calculating EVSI and ENGS could make the VoI anlaysis more accessible to trialists and health economists, promoting its wider application. Based on the hypothetical PSA, there may be a further study to improve the precision of estimation for compliance of antipsychotics in routine practice for patients with severe mental health illnesses. A prospective cohort study on the same population may be carried out with monthly follow-up on antipsychotic adherent status of the patients. The outcome is the average probability of adhering the antipsychotic maintenance treatment over time. The outcome may also be conditional on patient characteristics, antipsychotic description for more precise estimates for each subpopulation of interest. This study could potentially

reduce the uncertainty of decision making based on a £30,000/QALY WTP threshold, given that the relapse incidence in the RADAR intervention arm are well managed.

6.4 The ethics behind RADAR trial

The Mental Health Act demands that patients with SMIs can be detained and treated without their consent, on the basis that they may be at risk of harm to themselves or others. While this considers the greater good of society, it does raise the question of whether it is ethical to remove someone's right in receiving treatments, or in this case rejecting them. Moreover, as they start antipsychotic treatment, they are usually advised to keep taking them in order to control the symptoms. Setting aside the argument of the effectiveness of antipsychotic in recovery, the longer patients take antipsychotics, the more they may want to discontinue due to the side effects. While it seems that society insists on them taking the treatment simply because it could make the others feel safer, this may just be an ideal case, as non-adherence rate is high amongst schizophrenia patients who are taking antipsychotic medication. (173) The off-putting experience of long-term antipsychotic treatment may force some patients to discontinue, but without notice to their psychiatrists. Withdrawal without professional supervision may easily induce an acute episode, rebound or withdrawal effects, which are often mistaken for relapse (174). Aggressive and violent behaviours may ensue, and CJS involved to section them, which they are forced back onto antipsychotic treatment again. The cycle continues. The efforts involved into implementing the Mental Health Act in this situation could be considerable and a simplistic decision modelling, no matter how complex, cannot possibly capture all of it.

There is the debate whether antipsychotics are helpful from service users' perspective. Apart from controlling abnormal behaviours, or symptoms, it hardly benefits the patients in terms of trying to alter the biological pathology of the brain, the idea on which many other pharmacological treatments for different conditions are based. Moncrieff summarised drug action into two models, the disease centred model and the drug centred model. The former one refers to drug actions that exert on specific disease process. (10) For instance, insulin reverse the supply deficiency of hormone insulin on patients with diabetes, by providing replacement from the medication. On the other hand, the drug centred model suggests that psychiatric drugs induce an abnormal state, modify the way brain functions and create alterations in sensations. This is particularly common in mental health medication. The pathology behind some of mental health conditions still requires clarification. The majority of the psychoactive drugs focus on creating an abnormal brain state which superimposes on the current diseased nervous system, instead of helping to correct the current abnormal state. (188) When antipsychotics were first introduced in 1950s, they were known as 'major tranquilisers' primarily for their sedative effects. This drug-centred model for antipsychotic was then replaced by the disease centred model by 1970s although in fact there is seldom convincing evidence for disease-centred model. There should be more research and debate into alternative theories of drug action for antipsychotics, and more other psychiatric drugs.

Since the current antipsychotic maintenance treatment for patients with schizophrenia does not provide a scientifically sound and effective therapeutic solution, let alone that some psychiatric drugs may lead to brain damage, or

CVD, why should patients still keep taking them? (189) It is the “abnormality” in patients’ behaviour and emotion that firstly made sedation a choice when it comes to keep patients with schizophrenia under control. However, instead of stopping taking the medication after the symptom is under control, patients are still advised to take them due to the concern that abnormal behaviours may occur again, thus putting the safety of themselves and others at stake. Although studies did find association between the long-term antipsychotic treatment and reduced crime risk, this does not explain the absence of arguments, or questions, at least, on the possibility of safely reducing it. There have not been enough effort into at least trying to make patients come off antipsychotics safely, which is where RADAR trial stands firm. It is after all, ethical to help patients come off antipsychotic treatment if they wish to do so, because they have every right to decide what treatment they would like or not, just as people with any other conditions. However due to the potential negative impact of coming off antipsychotic treatment, some believe that it would be safer to keep patients on antipsychotics. The RADAR trial provides a great opportunity to demonstrate that patients can safely and gradually reduce and discontinue the treatment, without causing considerable negative impact to the society. Moreover, the intervention is compared with the maintenance treatment, but in an RCT environment patients received more monitoring and support even when they are in the control group. This makes a difference when in fact RADAR should be comparing itself to the real-world scenario, where patients are reducing and discontinuing antipsychotics without supervision, close monitoring or support. However, it is ethically wrong to randomise patients where they are a vulnerable group with potentially high needs and would receive a reduction in monitoring and support compared to standard of care, such that it could put

them at risk of harm. Hence the real-world scenario cannot be used for comparison in an RCT. The economic analysis showed that cost effectiveness could be a disadvantage compared to the closely monitored maintenance treatment. However this does lead to further questions – when evaluating an intervention that tries to remove the existing treatment, in the absence of an appropriate comparator (at least in an RCT setting), to what degree can standard cost-effectiveness thresholds still be applicable?

6.5 Concluding remark

My thesis looked at several ways mental health RCTs could conduct with more efficiency, using RADAR trial as a case. Using different strategies to improve recruitment and retention in RCTs, either via interventional SWATs, or via *ad hoc* observational ways, can both help to maintain the validity of the trial. Although the evidence of using SWATs to compare the effectiveness of different strategies has not been completely established, the methodology can provide a sound basis for recruitment strategies for specific interventions on specific cohorts.

Using decision modelling technique to simulate the cost-effectiveness of the RADAR intervention, I found that the intervention is not cost effective from base case analysis, and has very little chance of being cost effective under the £30,000/QALY WTP threshold. However as this was an early model, many input parameters were subject to uncertainty. Sensitivity analyses were performed on key parameters. Antipsychotic reduction strategy could result in some reduced risk of cardiovascular events, despite the risk of relapse and the

purported improvement in wellbeing remain unknown until the trial finishes. The cost-effectiveness of the RADAR strategy may be decided based upon whether patients are adherent to the antipsychotic treatment in real world, which is difficult to measure in the RADAR trial due to the difference in patients' behaviour between RCTs and real world.

The efficiency of mental health RCTs was discussed from various aspects in previous chapters. Methods to improve patients recruitment, specifically those may reduce average costs per patient recruited, may help with better efficiency in terms of having resources saved potentially for more research in the future to reduce uncertainty of the decision making. Meanwhile, using the decision model approach may help with research programme planning by predicting the likelihood cost-effectiveness of the intervention as well as some of the related clinical effectiveness. VoI analyses based on cost-effectiveness can help with future research planning to conduct research project with sufficient and economically optimal sample size to reduce the uncertainty of decision under certain WTP threshold.

There are a few weaknesses in this study. The evidence base of recruitment strategy comparison in mental health trials were found largely observational hence it is difficult to make robust conclusion on the comparative effectiveness of each strategy, even so on cost-effectiveness which always adopts an incremental approach. Average cost per patients recruited were used as outcome but other factors also may need consideration when it comes to recruiting patients into mental health RCTs. Secondly, there were not enough evidence to support the design of the cost-effectiveness model for RADAR trial due to the nature of the intervention, the results presented in Chapter 4 showed

a dominated cost-effectiveness based on existing evidence. The trial-based CEA for RADAR trial can certainly provide more information on the comparative safety and efficacy of the intervention for the model to make a more robust conclusion on the cost-effectiveness case.

Different from RCTs on drugs, RCTs for complex intervention often needs less of standard. This also applies to RADAR, which does not only investigate a non-pharmacological intervention, but also takes away the medication which has been prescribed based on a yet debatable scientific hypothesis. Similarly, the cost effectiveness framework has been widely used with HTA submission on funding new medicine from manufactures. However, when it comes to a treatment management programme like RADAR, which in fact removes medications from patients, what is the evidence from cost-effectiveness suggesting? Furthermore, would the same cost-effectiveness threshold for decision making still apply?

Reference

1. Who. WHO Mental disorders - Fact sheet. 2016.
2. Jorm AF. Mental health literacy: Public knowledge and beliefs about. British Journal of Psychiatry. 2000;177(396):317-27.
3. Sellers R, Collishaw S, Rice F, Thapar AK, Potter R, Mars B, et al. Risk of psychopathology in adolescent offspring of mothers with psychopathology and recurrent depression. Br J Psychiatry. 2013;202(2):108-14.
4. Fann JR, Burington B, Leonetti A, Jaffe K, Katon WJ, Thompson RS. Psychiatric illness following traumatic brain injury in an adult healthMaintenance organization population. Archives of general psychiatry. 2004;61(1):53-61.
5. World Health Organization. Schizophrenia: Fact sheet 2019 [Available from: https://www.who.int/en/news-room/fact-sheets/detail/schizophrenia.
6. Sims A. Symptoms in the mind: An introduction to descriptive psychopathology: Bailliere Tindall Publishers; 1988.
7. Shah JN, Qureshi SU, Jawaid A, Schulz PE. Is there evidence for late cognitive decline in chronic schizophrenia? Psychiatric Quarterly. 2012;83(2):127-2720.
8. Schultze-Lutter F. Subjective symptoms of schizophrenia in research and the clinic: the basic symptom concept. Schizophr Bull. 2009;35(1):5-8.
9. Weinstein JJ, Chohan MO, Slifstein M, Kegeles LS, Moore H, Abi-Dargham A. Pathway-Specific Dopamine Abnormalities in Schizophrenia. Biol Psychiatry. 2017;81(1):31-42.
10. Creese I, Burt DR, Snyder SH. Dopamine receptor binding predicts clinical and pharmacological potencies of antischizophrenic drugs. Science. 1976;192(4238):481-3.
11. Behrens MM, Ali SS, Dao DN, Lucero J, Shekhtman G, Quick KL, et al. Ketamine-induced loss of phenotype of fast-spiking interneurons is mediated by NADPH-oxidase. Science. 2007;318(5856):1645-7.
12. B. M. J. Publishing Group Ltd, Society RP. British National Formulary (BNF) 73. 2017.
13. Consensus Development Conference on Antipsychotic Drugs and Obesity and Diabetes. Diabetes Care. 2004;27(2):596-.
14. Osborn DP, Levy G, Nazareth I, Petersen I, Islam A, King MB. Relative risk of cardiovascular and cancer mortality in people with severe mental illness from the United Kingdom's General Practice Rsearch Database. Arch Gen Psychiatry. 2007;64(2):242-9.
15. Lieberman JA, Stroup TS, McEvoy JP, Swartz MS, Rosenheck RA, Perkins DO, et al. Effectiveness of antipsychotic drugs in patients with chronic schizophrenia. N Engl J Med. 2005;353(12):1209-23.
16. Harrow M, Jobe TH, Faull RN. Do all schizophrenia patients need antipsychotic treatment continuously throughout their lifetime? A 20-year longitudinal study. Psychol Med. 2012;42(10):2145-55.
17. Osborn DP, Hardoon S, Omar RZ, Holt RI, King M, Larsen J, et al. Cardiovascular risk prediction models for people with severe mental illness: results from the prediction and management of cardiovascular risk in people with severe mental illnesses (PRIMROSE) research program. JAMA Psychiatry. 2015;72(2):143-51.
18. Moncrieff J. Antipsychotic Maintenance Treatment: Time to Rethink? PLoS Med. 2015;12(8):e1001861.

19. Viguera AC, Baldessarini RJ, Hegarty JD, van Kammen DP, Tohen M. Clinical risk following abrupt and gradual withdrawal of maintenance neuroleptic treatment. Arch Gen Psychiatry. 1997;54(1):49-55.
20. Leucht S, Tardy M, Komossa K, Heres S, Kissling W, Salanti G, et al. Antipsychotic drugs versus placebo for relapse prevention in schizophrenia: A systematic review and meta-analysis. 2012. p. 2063-71.
21. Piantadosi S. Clinical Trials: Wiley; 2005.
22. O'Quigley J, Pepe M, Fisher L. Continual reassessment method: a practical design for phase 1 clinical trials in cancer. Biometrics. 1990;46(1):33-48.
23. Everitt BS, Wessely S. Clinical trials in psychiatry. 2008.
24. Roberts C, Torgerson D. Randomisation methods in controlled trials. BMJ. 1998;317(7168):1301.
25. Lachin JM, Matts JP, Wei LJ. Randomization in clinical trials: conclusions and recommendations. Control Clin Trials. 1988;9(4):365-74.
26. Matts JP, Lachin JM. Properties of permuted-block randomization in clinical trials. Control Clin Trials. 1988;9(4):327-44.
27. Piantadosi S. Clinical trials: a methodologic perspective. 2013.
28. Friedman LM, Furberg CD, DeMets DL. Fundamentals of Clinical Trials: Springer; 2010.
29. The Food and Drug Administration. Guidance for Industry Diabetes Mellitus — Evaluating Cardiovascular Risk in New Antidiabetic Therapies to Treat Type 2 Diabetes. Clinical/Medical. 2008.
30. U.S. Food and Drug Administration. Non-Inferiority Clinical Trials to Establish Effectiveness. Guidance. 2016.
31. Julious SA. Sample sizes for clinical trials with normal data. Stat Med. 2004;23(12):1921-86.
32. Chang M. Modern issues and methods in biostatistics: Springer Science & Business Media; 2011.
33. Hemming K, Eldridge S, Forbes G, Weijer C, Taljaard M. How to design efficient cluster randomised trials. BMJ. 2017;358:j3064.
34. Donner A, Brown KS, Brasher P. A methodological review of non-therapeutic intervention trials employing cluster randomization, 1979-1989. Int J Epidemiol. 1990;19(4):795-800.
35. Walters SJ, Bonacho Dos Anjos Henriques-Cadby I, Bortolami O, Flight L, Hind D, Jacques RM, et al. Recruitment and retention of participants in randomised controlled trials: a review of trials funded and published by the United Kingdom Health Technology Assessment Programme. BMJ Open. 2017;7(3):e015276.
36. Treweek S, Pitkethly M, Cook J, Fraser C, Mitchell E, Sullivan F, et al. Strategies to improve recruitment to randomised trials. 2018.
37. Schulz KF, Grimes DA. Sample size slippages in randomised trials: exclusions and the lost and wayward. Lancet. 2002;359(9308):781-5.
38. Carpenter, Jr., Kenward M. Missing data in randomised controlled trials–a practical guide. Birmingham: National Coordinating Centre …. 2007:29-36.
39. Prakash A, Risser RC, Mallinckrodt CH. The impact of analytic method on interpretation of outcomes in longitudinal clinical trials. Int J Clin Pract. 2008;62(8):1147-58.
40. Whitehead J, editor On being the statistician on a data and safety monitoring board1999.
41. Bross I. Sequential Medical Plans. Biometrics. 1952.

42. Freemantle N. Interpreting the results of secondary end points and subgroup analyses in clinical trials: should we lock the crazy aunt in the attic? BMJ. 2001;322(7292):989-91.
43. Pocock SJ. Group sequential methods in the design and analysis of clinical trials. Biometrika. 1977;64(2):191-9.
44. DeMets DL, Pocock SJ, Julian DG. The agonising negative trend in monitoring of clinical trials. Lancet. 1999;354(9194):1983-8.
45. Peto R, Pike MC, Armitage P, Breslow NE, Cox DR, Howard SV, et al. Design and analysis of randomized clinical trials requiring prolonged observation of each patient. I. Introduction and design. Br J Cancer. 1976;34(6):585-612.
46. Pocock SJ. When (not) to stop a clinical trial for benefit. JAMA. 2005;294(17):2228-30.
47. DeMets DL, Lan KK. Interim analysis: the alpha spending function approach. Stat Med. 1994;13(13-14):1341-52; discussion 53-6.
48. Armitage P. The evolution of ways of deciding when clinical trials should stop recruiting. Interview by Iain Chalmers. J R Soc Med. 2014;107(1):34-9.
49. Speich B, von Niederhäusern B, Schur N, Hemkens LG, Fürst T, Bhatnagar N, et al. Systematic review on costs and resource use of randomized clinical trials shows a lack of transparent and comprehensive data. 2018. p. 1-11.
50. Brintnall-Karabelas J, Sung S, Cadman ME, Squires C, Whorton K, Pao M. Improving recruitment in clinical trials: why eligible participants decline. J Empir Res Hum Res Ethics. 2011;6(1):69-74.
51. Hemminki E, Hovi SL, Veerus P, Sevon T, Tuimala R, Rahu M, et al. Blinding decreased recruitment in a prevention trial of postmenopausal hormone therapy. J Clin Epidemiol. 2004;57(12):1237-43.
52. Treweek S, Bevan S, Bower P, Campbell M, Christie J, Clarke M, et al. Trial Forge Guidance 1: What is a Study Within A Trial (SWAT)? 2018.
53. National Institute of Health Research. Annual Efficient Studies funding calls for CTU projects 2020 [Available from: https://www.nihr.ac.uk/documents/ad-hoc-funding-calls-for-ctu-projects/20141.
54. Hunter RM, Baio G, Butt T, Morris S, Round J, Freemantle N. An educational review of the statistical issues in analysing utility data for cost-utility analysis. Pharmacoeconomics. 2015;33(4):355-66.
55. Boyd KA, Briggs AH, Fenwick E, Norrie J, Stock S. Power and sample size for cost-effectiveness analysis: fFN neonatal screening. Contemporary clinical trials. 2011;32(6):893-901.
56. Faria R, Gomes M, Epstein D, White IR. A guide to handling missing data in cost-effectiveness analysis conducted within randomised controlled trials. Pharmacoeconomics. 2014;32(12):1157-70.
57. Gabrio A, Mason AJ, Baio G. Handling missing data in within-trial cost-effectiveness analysis: a review with future recommendations. PharmacoEconomics-open. 2017;1(2):79-97.
58. Leurent B, Gomes M, Carpenter JR. Missing data in trial‐based cost‐effectiveness analysis: An incomplete journey. Health economics. 2018;27(6):1024-40.
59. Earnshaw J, Lewis G. NICE Guide to the Methods of Technology Appraisal: pharmaceutical industry perspective. Pharmacoeconomics. 2008;26(9):725-7.
60. Briggs AH, Claxton K, Sculpher MJ. Decision modelling for health economic evaluation: Oxford University Press, USA; 2006.

61. Claxton K, Martin S, Soares M, Rice N, Spackman E, Hinde S, et al. Methods for the estimation of the National Institute for Health and Care Excellence cost-effectiveness threshold. Health Technol Assess. 2015;19(14):1-503, v-vi.
62. Raftery JP. NICE's cost-effectiveness range: should it be lowered? : Springer; 2014.
63. Barnsley P, Towse A, Britton S, Sussex J. Critique of CHE Research Paper 81: Methods for the Estimation of the NICE Cost Effectiveness Threshold. 2015.
64. Dillon A. Carrying NICE over the threshold. 2015.
65. Wilson EC. A practical guide to value of information analysis. Pharmacoeconomics. 2015;33(2):105-21.
66. Eckermann S, Willan AR. Expected value of information and decision making in HTA. Health Econ. 2007;16(2):195-209.
67. Wunderink L, Nieboer RM, Wiersma D, Sytema S, Nienhuis FJ. Recovery in remitted first-episode psychosis at 7 years of follow-up of an early dose reduction/discontinuation or maintenance treatment strategy: long-term follow-up of a 2-year randomized clinical trial. JAMA Psychiatry. 2013;70(9):913-20.
68. Liu Y, Pencheon E, Hunter RM, Moncrieff J, Freemantle N. Recruitment and retention strategies in mental health trials - A systematic review. PLoS One. 2018;13(8):e0203127.
69. Weisfeld V, English RA, Claiborne AB. Recruitment Challenges in Clinical Trials for Different Diseases and Conditions. 2012. p. 17-30.
70. Mason V, Shaw A, Wiles N, Mulligan J, Peters T, Sharp D, et al. GPs' experiences of primary care mental health research: a qualitative study of the barriers to recruitment. Fam Pract. 2007;24(5):518-25.
71. Walsh BT, Seidman SN, Sysko R, Gould M. Placebo response in studies of major depression: variable, substantial, and growing. JAMA. 2002;287(14):1840-7.
72. Kinon BJ, Potts AJ, Watson SB. Placebo response in clinical trials with schizophrenia patients. Curr Opin Psychiatry. 2011;24(2):107-13.
73. Rutherford BR, Pott E, Tandler JM, Wall MM, Roose SP, Lieberman JA. Placebo response in antipsychotic clinical trials: a meta-analysis. JAMA Psychiatry. 2014;71(12):1409-21.
74. Palmer BA, Pankratz VS, Bostwick JM. The lifetime risk of suicide in schizophrenia: a reexamination. Arch Gen Psychiatry. 2005;62(3):247-53.
75. Spineli LM, Leucht S, Cipriani A, Higgins JP, Salanti G. The impact of trial characteristics on premature discontinuation of antipsychotics in schizophrenia. Eur Neuropsychopharmacol. 2013;23(9):1010-6.
76. Park JJ, Siden E, Zoratti MJ, Dron L, Harari O, Singer J, et al. Systematic review of basket trials, umbrella trials, and platform trials: a landscape analysis of master protocols. Trials. 2019;20(1):1-10.
77. Treweek S, Lockhart P, Pitkethly M, Cook JA, Kjeldstrom M, Johansen M, et al. Methods to improve recruitment to randomised controlled trials: Cochrane systematic review and meta-analysis. BMJ Open. 2013;3(2):e002360--e.
78. Brueton VC, Tierney J, Stenning S, Harding S, Meredith S, Nazareth I, et al. Strategies to improve retention in randomised trials. Cochrane Database Syst Rev. 2013(12):MR000032.
79. Man MS, Healthlines Study G, Rick J, Bower P, Group M-S. Improving recruitment to a study of telehealth management for long-term conditions in

primary care: two embedded, randomised controlled trials of optimised patient information materials. Trials. 2015;16(1):309.
80. Jeste DV, Palmer BW, Golshan S, Eyler LT, Dunn LB, Meeks T, et al. Multimedia consent for research in people with schizophrenia and normal subjects: a randomized controlled trial. Schizophr Bull. 2009;35(4):719-29.
81. Krusche A, Rudolf von Rohr I, Muse K, Duggan D, Crane C, Williams JM. An evaluation of the effectiveness of recruitment methods: the staying well after depression randomized controlled trial. Clin Trials. 2014;11(2):141-9.
82. Morgan AJ, Jorm AF, Mackinnon AJ. Internet-based recruitment to a depression prevention intervention: lessons from the Mood Memos study. J Med Internet Res. 2013;15(2):e31.
83. Rollman BL, Fischer GS, Zhu F, Belnap BH. Comparison of electronic physician prompts versus waitroom case-finding on clinical trial enrollment. J Gen Intern Med. 2008;23(4):447-50.
84. Daley A, Winter H, Grimmett C, McGuinness M, McManus R, MacArthur C. Feasibility of an exercise intervention for women with postnatal depression: a pilot randomised controlled trial. Br J Gen Pract. 2008;58(548):178-83.
85. Woolhouse H, Mercuri K, Judd F, Brown SJ. Antenatal mindfulness intervention to reduce depression, anxiety and stress: a pilot randomised controlled trial of the MindBabyBody program in an Australian tertiary maternity hospital. BMC Pregnancy Childbirth. 2014;14:369.
86. DeBar LL, Yarborough BJ, Striegel-Moore RH, Rosselli F, Perrin N, Wilson GT, et al. Recruitment for a guided self-help binge eating trial: potential lessons for implementing programs in everyday practice settings. Contemp Clin Trials. 2009;30(4):326-33.
87. Le HN, Lara MA, Perry DF. Recruiting Latino women in the U.S. and women in Mexico in postpartum depression prevention research. Arch Womens Ment Health. 2008;11(2):159-69.
88. Schlernitzauer M, Bierhals AJ, Geary MD, Prigerson HG, Stack JA, Miller MD, et al. Recruitment Methods for Intervention Research in Bereavement-Related Depression: Five Years' Experience. The American Journal of Geriatric Psychiatry. 1998;6(1):67-74.
89. Hughes-Morley A, Hann M, Fraser C, Meade O, Lovell K, Young B, et al. The impact of advertising patient and public involvement on trial recruitment: embedded cluster randomised recruitment trial. Trials. 2016;17(1):586.
90. Brown JSL, Murphy C, Kelly J, Goldsmith K. How can we successfully recruit depressed people? Lessons learned in recruiting depressed participants to a multi-site trial of a brief depression intervention (the 'CLASSIC' trial). Trials. 2019;20(1):131.
91. Dirmaier J, Harfst T, Koch U, Schulz H. Incentives increased return rates but did not influence partial nonresponse or treatment outcome in a randomized trial. J Clin Epidemiol. 2007;60(12):1263-70.
92. McLean SA, Paxton SJ, Massey R, Mond JM, Rodgers B, Hay PJ. Prenotification but not envelope teaser increased response rates in a bulimia nervosa mental health literacy survey: A randomized controlled trial. J Clin Epidemiol. 2014;67(8):870-6.
93. Kearney A, Harman NL, Bacon N, Daykin A, Heawood AJ, Lane A, et al. Online resource for recruitment research in clinical trials research (ORRCA) 2017 [Available from: https://www.orrca.org.uk/.
94. Madurasinghe VW, Sandra Eldridge on behalf of MRCSG, Gordon Forbes on behalf of the SECG. Guidelines for reporting embedded recruitment trials. Trials. 2016;17(1):27.

95. Knapp M, Mangalore R, Simon J. The global costs of schizophrenia. Schizophr Bull. 2004;30(2):279-93.
96. McCrone PR, Dhanasiri S, Patel A, Knapp M, Lawton-Smith S. Paying the price: the cost of mental health care in England to 2026. 2008.
97. von Scheele B, Mauskopf J, Brodtkorb TH, Ainsworth C, Berardo CG, Patel A. Relationship between modeling technique and reported outcomes: case studies in models for the treatment of schizophrenia. Expert Rev Pharmacoecon Outcomes Res. 2014;14(2):235-57.
98. Philips Z, Ginnelly L, Sculpher M, Claxton K, Golder S, Riemsma R, et al. Review of guidelines for good practice in decision-analytic modelling in health technology assessment. Health technology assessment (Winchester, England). 2004;8(36):iii-iv, ix.
99. Einarson TR, Bereza BG, Garcia Llinares I, Gonzalez Martin Moro B, Tedouri F, Van Impe K. Cost-effectiveness of 3-month paliperidone treatment for chronic schizophrenia in Spain. J Med Econ. 2017;20(10):1039-47.
100. Phanthunane P, Vos T, Whiteford H, Bertram M. Cost-effectiveness of pharmacological and psychosocial interventions for schizophrenia. Cost Eff Resour Alloc. 2011;9(1):6.
101. Furiak NM, Gahn JC, Klein RW, Camper SB, Summers KH. Estimated economic benefits from low-frequency administration of atypical antipsychotics in treatment of schizophrenia: a decision model. Ann Gen Psychiatry. 2012;11(1):29.
102. Rajagopalan K, Hassan M, O'Day K, Meyer K, Grossman F. Cost-effectiveness of lurasidone vs aripiprazole among patients with schizophrenia who have previously failed on an atypical antipsychotic: an indirect comparison of outcomes from clinical trial data. J Med Econ. 2013;16(7):951-61.
103. Davies A, Vardeva K, Loze JY, L'Italien G J, Sennfalt K, Baardewijk M. Cost-effectiveness of atypical antipsychotics for the management of schizophrenia in the UK. Curr Med Res Opin. 2008;24(11):3275-85.
104. Bounthavong M, Okamoto MP. Decision analysis model evaluating the cost-effectiveness of risperidone, olanzapine and haloperidol in the treatment of schizophrenia. J Eval Clin Pract. 2007;13(3):453-60.
105. Bobes J, Canas F, Rejas J, Mackell J. Economic consequences of the adverse reactions related with antipsychotics: an economic model comparing tolerability of ziprasidone, olanzapine, risperidone, and haloperidol in Spain. Prog Neuropsychopharmacol Biol Psychiatry. 2004;28(8):1287-97.
106. Garcia-Ruiz AJ, Perez-Costillas L, Montesinos AC, Alcalde J, Oyaguez I, Casado MA. Cost-effectiveness analysis of antipsychotics in reducing schizophrenia relapses. Health Econ Rev. 2012;2(1):8.
107. Frey S, Linder R, Juckel G, Stargardt T. Cost-effectiveness of long-acting injectable risperidone versus flupentixol decanoate in the treatment of schizophrenia: a Markov model parameterized using administrative data. Eur J Health Econ. 2014;15(2):133-42.
108. Beard SM, Maciver F, Clouth J, Ruther E. A decision model to compare health care costs of olanzapine and risperidone treatment for schizophrenia in Germany. Eur J Health Econ. 2006;7(3):165-72.
109. Wang PS, Ganz DA, Benner JS, Glynn RJ, Avorn J. Should clozapine continue to be restricted to third-line status for schizophrenia?: A decision-analytic model. Journal of Mental Health Policy and Economics. 2004;7(2):77-85.
110. Graham CN, Mauskopf JA, Lawson AH, Ascher-Svanum H, Bruhn D. Updating and confirming an industry-sponsored pharmacoeconomic model:

comparing two antipsychotics in the treatment of schizophrenia. Value Health. 2012;15(1):55-64.
111. Vera-Llonch M, Delea TE, Richardson E, Rupnow M, Grogg A, Oster G. Outcomes and costs of risperidone versus olanzapine in patients with chronic schizophrenia or schizoaffective disorders: a Markov model. Value Health. 2004;7(5):569-84.
112. Palmer CS, Revicki DA, Genduso LA, Hamilton SH, Brown RE. A cost-effectiveness clinical decision analysis model for schizophrenia. Am J Manag Care. 1998;4(3):345-55.
113. Almond S, O'Donnell O. Cost analysis of the treatment of schizophrenia in the UK: a comparison of olanzapine and haloperidol. Pharmacoeconomics. 1998;13(5 Pt 2):575-88.
114. Heeg B, Buskens E, Botteman M, Caleo S, Ingham M, Damen J, et al. The cost-effectiveness of atypicals in the UK. 2008. p. 1007-21.
115. Park T, Kuntz KM. Cost-effectiveness of second-generation antipsychotics for the treatment of schizophrenia. Value Health. 2014;17(4):310-9.
116. O'Day K, Rajagopalan K, Meyer K, Pikalov A, Loebel A. Long-term cost-effectiveness of atypical antipsychotics in the treatment of adults with schizophrenia in the US. Clinicoecon Outcomes Res. 2013;5:459-70.
117. Laux G, Heeg B, van Hout BA, Mehnert A. Costs and effects of long-acting risperidone compared with oral atypical and conventional depot formulations in Germany. Pharmacoeconomics. 2005;23 Suppl 1:49-61.
118. Einarson TR, Zilbershtein R, Skoupa J, Vesela S, Garg M, Hemels ME. Economic and clinical comparison of atypical depot antipsychotic drugs for treatment of chronic schizophrenia in the Czech Republic. J Med Econ. 2013;16(9):1089-95.
119. Lachaine J, Beauchemin C, Mathurin K, Gilbert D, Beillat M. Cost-effectiveness of asenapine in the treatment of schizophrenia in Canada. J Med Econ. 2014;17(4):296-304.
120. Dilla T, Moller J, O'Donohoe P, Alvarez M, Sacristan JA, Happich M, et al. Long-acting olanzapine versus long-acting risperidone for schizophrenia in Spain - a cost-effectiveness comparison. BMC Psychiatry. 2014;14(1):298.
121. Kasteng F, Eriksson J, Sennfalt K, Lindgren P. Metabolic effects and cost-effectiveness of aripiprazole versus olanzapine in schizophrenia and bipolar disorder. Acta Psychiatr Scand. 2011;124(3):214-25.
122. McIntyre RS, Cragin L, Sorensen S, Naci H, Baker T, Roussy JP. Comparison of the metabolic and economic consequences of long-term treatment of schizophrenia using ziprasidone, olanzapine, quetiapine and risperidone in Canada: a cost-effectiveness analysis. J Eval Clin Pract. 2010;16(4):744-55.
123. Lindner LM, Marasciulo AC, Farias MR, Grohs GEM. Economic evaluation of antipsychotic drugs for schizophrenia treatment within the Brazilian Healthcare System. Revista de saúde pública. 2009;43 Suppl 1:62-9.
124. Geitona M, Kousoulakou H, Ollandezos M, Athanasakis K, Papanicolaou S, Kyriopoulos I. Costs and effects of paliperidone extended release compared with alternative oral antipsychotic agents in patients with schizophrenia in Greece: a cost effectiveness study. Ann Gen Psychiatry. 2008;7:16.
125. Obradovic M, Mrhar A, Kos M. Cost-effectiveness of antipsychotics for outpatients with chronic schizophrenia. Int J Clin Pract. 2007;61(12):1979-88.
126. Yang YK, Tarn YH, Wang TY, Liu CY, Laio YC, Chou YH, et al. Pharmacoeconomic evaluation of schizophrenia in Taiwan: model comparison

of long-acting risperidone versus olanzapine versus depot haloperidol based on estimated costs. Psychiatry Clin Neurosci. 2005;59(4):385-94.
127. Tilden D, Aristides M, Meddis D, Burns T. An economic assessment of quetiapine and haloperidol in patients with schizophrenia only partially responsive to conventional antipsychotics. Clin Ther. 2002;24(10):1648-67.
128. Lecomte P, De Hert M, van Dijk M, Nuijten M, Nuyts G, Persson U. A 1-year cost-effectiveness model for the treatment of chronic schizophrenia with acute exacerbations in Belgium. Value Health. 2000;3(1):1-11.
129. Colombo GL, Caruggi M, Di Matteo S, Rossi A. An economic evaluation of aripiprazole vs olanzapine adapted to the Italian setting using outcomes of metabolic syndrome and risk for diabetes in patients with schizophrenia. Neuropsychiatr Dis Treat. 2008;4(5):967-76.
130. Mehnert A, Nicholl D, Pudas H, Martin M, McGuire A. Cost effectiveness of paliperidone palmitate versus risperidone long-acting injectable and olanzapine pamoate for the treatment of patients with schizophrenia in Sweden. Journal of medical economics. 2012;15(5):844-61.
131. National Institute of Health and Care Excellence. Psychosis and schizophrenia in adults. The NICE guidelines on treament and management. 2014(February):74-80.
132. Davis S, Stevenson M, Tappenden P, Wailoo AJ. Cost-effectiveness modelling using patient-level simulation. NICE DSU Technical Support Document. 2014;15(April):1-62.
133. Chhatwal J, Jayasuriya S, Elbasha EH. Changing Cycle Lengths in State-Transition Models: Challenges and Solutions. Med Decis Making. 2016;36(8):952-64.
134. Heeg BM, Damen J, Buskens E, Caleo S, de Charro F, van Hout BA. Modelling approaches: the case of schizophrenia. Pharmacoeconomics. 2008;26(8):633-48.
135. Pennington M, McCrone P. The Cost of Relapse in Schizophrenia. 2017. p. 921-36.
136. Lehman AF, Goldberg R, Dixon LB, McNary S, Postrado L, Hackman A, et al. Improving employment outcomes for persons with severe mental illnesses. Arch Gen Psychiatry. 2002;59(2):165-72.
137. Monahan J, Arnold J. Violence by People with Mental Illness: A Consensus Statement by Advocates and Researchers. Psychiatric Rehabilitation Journal. 1996;19(4):67-70.
138. Lin I, Muser E, Munsell M, Benson C, Menzin J. Economic impact of psychiatric relapse and recidivism among adults with schizophrenia recently released from incarceration: a Markov model analysis. J Med Econ. 2015;18(3):219-29.
139. Walker S, Griffin S, Asaria M, Tsuchiya A, Sculpher M. Striving for a societal perspective: a framework for economic evaluations when costs and effects fall on multiple sectors and decision makers. Applied health economics and health policy. 2019;17(5):577-90.
140. Latimer N. NICE DSU technical support document 14: survival analysis for economic evaluations alongside clinical trials-extrapolation with patient-level data. Sheffield: Report by the Decision Support Unit. 2011(0).
141. Zhou J, Millier A, François C, Aballéa S, Toumi M. Systematic review of utility values used in the pharmacoeconomic evaluations for schizophrenia: implications on cost-effectiveness results. Journal of market access & health policy. 2019;7(1):1648973.

142. National Institute for Health Care and Excellence. Methods for development of NICE public health guidance. 2006. p. 175-.
143. Brennan A, Chick SE, Davies R. A taxonomy of model structures for economic evaluation of health technologies. Health Econ. 2006;15(12):1295-310.
144. Leucht S, Heres S, Hamann J, Kane JM. Methodological issues in current antipsychotic drug trials. 2008. p. 275-85.
145. Hawthorne WB, Folsom DP, Sommerfeld DH, Lanouette NM, Lewis M, Aarons GA, et al. Incarceration among adults who are in the public mental health system: rates, risk factors, and short-term outcomes. Psychiatr Serv. 2012;63(1):26-32.
146. Ministry of Justice Information. Costs per place and costs per prisoner by individual prison, HM Prison and Probation Service Annual Report and Accounts 2016-17 Management Information Addendum. 2017.
147. Nemeth B, Fasseeh A, Molnar A, Bitter I, Horvath M, Koczian K, et al. A systematic review of health economic models and utility estimation methods in schizophrenia. Expert Rev Pharmacoecon Outcomes Res. 2018;18(3):267-75.
148. Knapp M, Locklear J, Jarbrink K. Impact of psychotic relapse definitions in assessing drug efficacy and costs: comparison of quetiapine XR, olanzapine and paliperidone ER. Curr Med Res Opin. 2009;25(7):1593-603.
149. Zomer E, Osborn D, Nazareth I, Blackburn R, Burton A, Hardoon S, et al. Effectiveness and cost-effectiveness of a cardiovascular risk prediction algorithm for people with severe mental illness (PRIMROSE). BMJ Open. 2017;7(9):e018181.
150. Barnes TR, Paton C. Antipsychotic polypharmacy in schizophrenia: benefits and risks. CNS Drugs. 2011;25(5):383-99.
151. Office for National Statistics. National life tables , UK : 2015 to 2017. Statistical Bulletin. 2018.
152. Hayes JF, Marston L, Walters K, King MB, Osborn DPJ. Mortality gap for people with bipolar disorder and schizophrenia: UK-based cohort study 2000-2014. 2017. p. 175-81.
153. Day JC, Wood G, Dewey M, Bentall RP. A self-rating scale for measuring neuroleptic side-effects. Validation in a group of schizophrenic patients. Br J Psychiatry. 1995;166(5):650-3.
154. Waddell L, Taylor M. A new self-rating scale for detecting atypical or second-generation antipsychotic side effects. J Psychopharmacol. 2008;22(3):238-43.
155. Ohlsen RI, Williamson R, Yusufi B, Mullan J, Irving D, Mukherjee S, et al. Interrater reliability of the Antipsychotic Non-Neurological Side-Effects Rating Scale measured in patients treated with clozapine. J Psychopharmacol. 2008;22(3):323-9.
156. Lacro JP, Dunn LB, Dolder CR, Leckband SG, Jeste DV. Prevalence of and risk factors for medication nonadherence in patients with schizophrenia: A comprehensive review of recent literature. 2002.
157. Palmer M, Williams J, McPake B. Cost Comparison Charts. 2017.
158. Curtis L, Burns A. Unit Costs of Health & Social Care 2016. PSSRU. 2016.
159. Department of Health and Social Care. NHS reference costs - GOV.UK. 2016.
160. Jin H, Mosweu I. The Societal Cost of Schizophrenia: A Systematic Review. 2017. p. 25-42.

161. Guest JF, Cookson RF. Cost of schizophrenia to UK Society. An incidence-based cost-of-illness model for the first 5 years following diagnosis. Pharmacoeconomics. 1999;15(6):597-610.
162. Mangalore R, Knapp M. Cost of schizophrenia in England. J Ment Health Policy Econ. 2007;10(1):23-41.
163. Doering S, Muller E, Kopcke W, Pietzcker A, Gaebel W, Linden M, et al. Predictors of relapse and rehospitalization in schizophrenia and schizoaffective disorder. Schizophr Bull. 1998;24(1):87-98.
164. Robinson D, Woerner MG, Alvir JM, Bilder R, Goldman R, Geisler S, et al. Predictors of relapse following response from a first episode of schizophrenia or schizoaffective disorder. Arch Gen Psychiatry. 1999;56(3):241-7.
165. Berman G. Prison population statistics. 2011.
166. Lenert LA, Sturley AP, Rapaport MH, Chavez S, Mohr PE, Rupnow M. Public preferences for health states with schizophrenia and a mapping function to estimate utilities from positive and negative symptom scale scores. Schizophr Res. 2004;71(1):155-65.
167. Osborne RH, Dalton A, Hertel J, Schrover R, Smith DK. Health-related quality of life advantage of long-acting injectable antipsychotic treatment for schizophrenia: a time trade-off study. Health Qual Life Outcomes. 2012;10:35.
168. Blak BT, Thompson M, Dattani H, Bourke A. Generalisability of The Health Improvement Network (THIN) database: demographics, chronic disease prevalence and mortality rates. Inform Prim Care. 2011;19(4):251-5.
169. Gianfrancesco FD, Grogg AL, Mahmoud RA, Wang RH, Nasrallah HA. Differential effects of risperidone, olanzapine, clozapine, and conventional antipsychotics on type 2 diabetes: findings from a large health plan database. J Clin Psychiatry. 2002;63(10):920-30.
170. Leslie DL, Rosenheck RA. Incidence of newly diagnosed diabetes attributable to atypical antipsychotic medications. Am J Psychiatry. 2004;161(9):1709-11.
171. Lean MEJ, Pajonk FG. Patients on atypical antipsychotic drugs: Another high-risk group for type 2 diabetes. 2003.
172. Csernansky JG, Schuchart EK. Relapse and rehospitalisation rates in patients with schizophrenia: effects of second generation antipsychotics. CNS Drugs. 2002;16(7):473-84.
173. Marder SR, editor Overview of Partial Compliance2003.
174. Moncrieff J. Why is it so difficult to stop psychiatric drug treatment? It may be nothing to do with the original problem. Med Hypotheses. 2006;67(3):517-23.
175. Fazel S, Zetterqvist J, Larsson H, Langstrom N, Lichtenstein P. Antipsychotics, mood stabilisers, and risk of violent crime. Lancet. 2014;384(9949):1206-14.
176. National Institute for Health Care and Excellence. Service user experience in adult mental health Improving the experience of care for people using adult NHS. Practice. 2011.
177. Rofail D, Heelis R, Gournay K. Results of a thematic analysis to explore the experiences of patients with schizophrenia taking antipsychotic medication. Clin Ther. 2009;31 Pt 1:1488-96.
178. National Institute for Health Care Excellence. Guide to the processes of technology appraisal: National Institute for Health and Care Excellence; 2014.
179. Claxton K, Sculpher M, McCabe C, Briggs A, Akehurst R, Buxton M, et al. Probabilistic sensitivity analysis for NICE technology assessment: Not an optional extra. 2005.

180. Strong M, Brennan A, Oakley J. How to calculate value of information in seconds using ‘SAVI’, the Sheffield Accelerated Value of Information web app. Value in Health. 2015;18(7):A725-A6.
181. Heath A, Manolopoulou I, Baio G. A Review of Methods for Analysis of the Expected Value of Information. Med Decis Making. 2017;37(7):747-58.
182. York Health Economics Consortium. Expected Value of Sample Information (EVSI) [online]. 2016.
183. Strong M, Oakley JE, Brennan A, Breeze P. Estimating the Expected Value of Sample Information Using the Probabilistic Sensitivity Analysis Sample: A Fast, Nonparametric Regression-Based Method. Med Decis Making. 2015;35(5):570-83.
184. Heath A. Bayesian computations for Value of Information measures using Gaussian processes, INLA and Moment Matching: UCL (University College London); 2018.
185. Strong M, Oakley JE, Brennan A. Estimating multiparameter partial expected value of perfect information from a probabilistic sensitivity analysis sample: a nonparametric regression approach. Med Decis Making. 2014;34(3):311-26.
186. Wilson E. How much detail is needed in cost estimation in an economic evaluation alongside a clinical trial to optimise evidence for decisions? : University of East Anglia; 2014.
187. Jalal H, Alarid-Escudero F. A Gaussian Approximation Approach for Value of Information Analysis. Med Decis Making. 2018;38(2):174-88.
188. Moncrieff J. Models of drug action. 2013.
189. Ho BC, Andreasen NC, Ziebell S, Pierson R, Magnotta V. Long-term antipsychotic treatment and brain volumes: a longitudinal study of first-episode schizophrenia. Arch Gen Psychiatry. 2011;68(2):128-37.
190. Shen WW. A history of antipsychotic drug development. 1999.
191. Carpenter WT, Gold JM, Lahti AC, Queern CA, Conley RR, Bartko JJ, et al. Decisional capacity for informed consent in schizophrenia research. 2000.
192. Dunn LB, Kim DS, Fellows IE, Palmer BW. Worth the risk? Relationship of incentives to risk and benefit perceptions and willingness to participate in schizophrenia research. Schizophr Bull. 2009;35(4):730-7.
193. Roberts LW, Warner TD, Brody JL, Roberts B, Lauriello J, Lyketsos C. Patient and psychiatrist ratings of hypothetical schizophrenia research protocols: assessment of harm potential and factors influencing participation decisions. 2002.
194. King M, Nazareth I, Lampe F, Bower P, Chandler M, Morou M, et al. Conceptual framework and systematic review of the effects of participants' and professionals' preferences in randomised controlled trials. Health Technol Assess. 2005;9(35):1-186, iii-iv.
195. Khan AY, Preskorn SH, Baker B. Effect of study criteria on recruitment and generalizability of the results. 2005.
196. Preskorn SH, Macaluso M, Trivedi M. How commonly used inclusion and exclusion criteria in antidepressant registration trials affect study enrollment. Journal of Psychiatric Practice. 2015;21(4):267-74.
197. Adshead G. Studying the mind: ethical issues and guidance in mental health research. Clinical Ethics. 2008;3(3):141-4.
198. Jenkinson CE, Winder RE, Sugg HVR, Roberts MJ, Ridgway N, Kuyken W, et al. Why do GPs exclude patients from participating in research? An exploration of adherence to and divergence from trial criteria. Family Practice. 2014;31(3):364-70.

199. McDermott C, Agras WS, Crow SJ, Halmi K, Mitchell JE, Bryson S. Participant recruitment for an anorexia nervosa treatment study. Int J Eat Disord. 2004;35(1):33-41.
200. Arnow BA, Blasey C, Manber R, Constantino MJ, Markowitz JC, Klein DN, et al. Dropouts versus completers among chronically depressed outpatients. J Affect Disord. 2007;97(1-3):197-202.
201. Walach H, Falkenberg T, Fønnebø V, Lewith G, Jonas WB. Circular instead of hierarchical: methodological principles for the evaluation of complex interventions. BMC medical research methodology. 2006;6(1):1-9.
202. Wise J. Leading academics condemn government undermining of NICE. British Medical Journal Publishing Group; 2015.
203. Culyer AJ. Why do/should we do economic evaluation? Value in Health Spotlight. 2015;1:8-10.
204. Freemantle N, Wood J, Crawford F. Evidence into practice, experimentation and quasi experimentation: are the methods up to the task? J Epidemiol Community Health. 1998;52(2):75-81.

Appendices

Appendix A Protocol for the systematic review in Chapter 2

Objectives

To investigate and evaluate different approaches to recruiting and retaining patients into mental health randomised clinical trials (RCT) and what may be done to improve recruitment and retention rate in RCT of such kinds. In addition, to discuss how different these strategies used in RCTs for mental illness enhance recruitment and retention rate.

Background

Recruitment in randomised clinical trials is challenging and delayed recruitment can give rise to a series of issues, such as reducing the power of the study and additional costs and extension of study period. Inadequate or ineffective recruitment may often result in early-termination of a trial and the failure in answering an important clinical question. As the gold standard of clinical research, randomised clinical trials often fail in recruiting enough subjects, particularly in mental health area. Lost to follow-up and patients dropouts can also cause a reduced study power in discovering the difference between trial arms. Although an increasing amount of research has contributed to dealing with missing data in clinical trials using various statistical techniques, the risk of bias due to missing data still cannot be avoided by such imputations. It is suggested that a less than 5% loss to follow-up may lead to minimum bias, while 20% lost to follow-up can pose a threat to a trial's validity. Some modern trials aim to reduce this risk by increasing the sample size by 20%, which in return increases the recruitment difficulties.

This review hopes to contribute to improving both the recruitment and retention rate for mental illness clinical trials.

Searching

MEDLINE, EMBASE, PsycINFO and Cochrane Methodology Register (CMR) are used for search of articles published in English language since inception of each database. The search strategy used in MEDLINE includes MeSH terms such as *exp "bipolar and related disorders"/ or "schizophrenia spectrum and other psychotic disorders"/ or exp schizophrenia/,* other search keywords include (randomi* adj2 clinical trial*).ti

Study Selection

Randomised controlled trials of interventions to improve recruitment and retention for mental illness randomised controlled trials.

Inclusion

- Comparative studies in which at least two different recruitment or retention strategies were compared in an RCT are eligible for inclusion.
- RCTs with adults are eligible for inclusion.
- Mental health illness is defined as conditions that are catogrised in DSM-5 or ICD-10 criteria.
- Studies are funded by public healthcare sector as well as ones by industry are eligible for inclusion.
- Studies to improve recruitment to cluster randomised trials, pilot stage of trials and feasibility studies need further screening for inclusion.
- Strategies both in real settings and in hypothetical settings (studies that ask potential patients' consent to participate in a trial if it was run but without the trial actually running) are eligible.
- As for specific recruitment strategies, we include all strategies that aim to improve recruitment of participants, mainly below but not limited to:

- Newsletters (including mailshots)
- Regular visits/telephone calls
- Posters in clinics/wards
- Amendments on protocol
- Presentations to appropriate groups
- Presentations at conferences
- Extra staffing
- Investigators' meeting/recruiting staff meetings
- Trial material customized to specific sites
- Investigators' visits to centers
- Letter to individuals
- Advertisement in newspapers
- Resource manual for recruiters
- Training
- Monetary incentives

Exclusion

- Studies on non-adults (i.e. children) are excluded.
- Studies that did not randomise patients into different groups are excluded.
- Studies in which mental illness as co-occurrence with primary conditions (i.e. CVD) are excluded.
- Studies which did not report comparison on interventions to improve recruitment nor retention for RCTs are excluded.

- Studies that focused on people with substance misuse problems where comorbid mental health conditions had not been assessed using DSM or ICD-10 criteria are excluded.

Outcome measures

Primary outcomes

For recruitment trials: Difference in numbers of patients recruited using each strategy

For retention trials: numbers of patients who were included in the final analysis over the numbers of patients who were randomised to each retention arm.

Secondary outcomes

- Cost effectiveness of different strategies
- Difficulties during recruitment
- Benefits using different strategies

Analysis will be described in a separate statistical analysis plan.

Assessment of study quality

For each included study, their adequacy of allocation concealment will be assessed (adequate, unclear and inadequate). Completeness of result reporting, loss to follow-up, risk of bias and other aspects will also be assessed. Completeness of reporting will be assessed by the quality of information on participants, intervention, comparator and outcome, as well as whether allocation was concealed. A Consolidated Standards of Reporting Tials (CONSORT) diagram will be used to record the flow of participants through the trial.

Data extraction

A data extraction form will be created to collect the information for the outcomes in each eligible study. Data included in the “outcome measures” section will be

collected. Authors of the studies will also be contacted for necessary but unpublished information.

Appendix B Search strategies for the systematic review in Chapter 2

MEDLINE (Ovid)

1 Patient Selection/ (55467)
2 ((participat$ or recruit$ or enrol$) adj4 trial?).tw. (20036)
3 Informed Consent/ (33169)
4 informed consent.tw. (28318)
5 recruitment.ab. /freq=2 (21355)
6 participation.ab. /freq=2 (20980)
7 (minimi$ adj2 attrition).ab,ti. (85)
8 (prevent$ adj2 attrition).ab,ti. (89)
9 (lessen$ adj2 attrition).ab,ti. (2)
10 (decreas$ adj2 attrition).ab,ti. (97)
11 (reduc$ adj2 attrition).ab,ti. (387)
12 (minimi$ adj2 drop-out).ab,ti. (7)
13 (prevent$ adj2 drop-out).ab,ti. (24)
14 (lessen$ adj2 drop-out).ab,ti. (1)
15 (decreas$ adj2 drop-out).ab,ti. (17)
16 (reduc$ adj2 drop-out).ab,ti. (66)
17 (minimi$ adj2 drop-out$).ab,ti. (9)
18 (prevent$ adj2 drop-out$).ab,ti. (26)
19 (lessen$ adj2 drop-out$).ab,ti. (1)
20 (decreas$ adj2 drop-out$).ab,ti. (22)
21 (reduc$ adj2 drop-out$).ab,ti. (82)
22 (minimi$ adj2 drop$-out).ab,ti. (9)
23 (prevent$ adj2 drop$-out).ab,ti. (28)
24 (lessen$ adj2 drop$-out).ab,ti. (1)
25 (decreas$ adj2 drop$-out).ab,ti. (20)
26 (reduc$ adj2 drop$-out).ab,ti. (72)
27 (minimi$ adj2 dropout$).ab,ti. (33)
28 (prevent$ adj2 dropout$).ab,ti. (111)
29 (lessen$ adj2 dropout$).ab,ti. (0)
30 (decreas$ adj2 dropout$).ab,ti. (63)
31 (reduc$ adj2 dropout$).ab,ti. (203)
32 (strateg$ adj2 drop$-out).ab,ti. (2)
33 (strateg$ adj2 dropout$).ab,ti. (8)
34 (loss adj2 follow-up).ab,ti. (3183)
35 (lost adj2 follow-up).ab,ti. (14548)
36 (loss adj2 followup).ab,ti. (48)
37 (lost adj2 followup).ab,ti. (652)
38 (minimi$ adj2 withdrawal).ab,ti. (93)
39 (prevent$ adj2 withdrawal).ab,ti. (404)
40 (lessen$ adj2 withdrawal).ab,ti. (9)
41 (decreas$ adj2 withdrawal).ab,ti. (687)
42 (reduc$ adj2 withdrawal).ab,ti. (1092)
43 (minimi$ adj2 withdrawal$).ab,ti. (97)

44 (prevent$ adj2 withdrawal$).ab,ti. (412)
45 (lessen$ adj2 withdrawal$).ab,ti. (9)
46 (decreas$ adj2 withdrawal$).ab,ti. (693)
47 (reduc$ adj2 withdrawal$).ab,ti. (1114)
48 (strateg$ adj2 attrition).ab,ti. (17)
49 (strateg$ adj2 drop-out).ab,ti. (1)
50 (strateg$ adj2 dropout).ab,ti. (5)
51 (strateg$ adj2 follow-up).ab,ti. (1092)
52 (strateg$ adj2 followup).ab,ti. (22)
53 (increas$ adj2 retention).ab,ti. (4080)
54 (encourag$ adj2 retention).ab,ti. (77)
55 (maximi$ adj2 retention).ab,ti. (152)
56 (promot$ adj2 retention).ab,ti. (727)
57 (improv$ adj2 retention).ab,ti. (2637)
58 (strateg$ adj2 response$).ab,ti. (2016)
59 (strateg$ adj2 (questionnaire$ adj3 response$)).ab,ti. (9)
60 (increas$ adj2 (questionnaire$ adj3 response$)).ab,ti. (40)
61 (encourag$ adj2 (questionnaire$ adj3 response$)).ab,ti. (1)
62 (maximi$ adj2 (questionnaire$ adj3 response$)).ab,ti. (3)
63 (promot$ adj2 (questionnaire$ adj3 response$)).ab,ti. (1)
64 (improv$ adj2 (questionnaire$ adj3 response$)).ab,ti. (23)
65 (increas$ adj2 response$).ab,ti. (36535)
66 (encourag$ adj2 response$).ab,ti. (601)
67 (maximi$ adj2 response$).ab,ti. (530)
68 (promot$ adj2 response$).ab,ti. (3924)
69 (improv$ adj2 response$).ab,ti. (11236)
70 (retention adj2 strateg$).ab,ti. (692)
71 retention rate$.ab,ti. (3246)
72 (retention adj2 method$).ab,ti. (412)
73 (retention adj2 technique$).ab,ti. (177)
74 attrition rate$.ab,ti. (1748)
75 (questionnaire$ adj3 (response$ adj2 method$)).ab,ti. (29)
76 (questionnaire$ adj3 (response adj2 technique$)).ab,ti. (3)
77 (questionnaire adj response rate$).ab,ti. (689)
78 (difficult$ adj2 (retain$ or retention)).ab,ti. (245)
79 Patient Dropouts/ (7244)
80 1 or 2 or 3 or 4 or 5 or 6 or 7 or 8 or 9 or 10 or 11 or 12 or 13 or 14 or 15 or 16 or 17 or 18 or 19 or 20 or 21 or 22 or 23 or 24 or 25 or 26 or 27 or 28 or 29 or 30 or 31 or 32 or 33 or 34 or 35 or 36 or 37 or 38 or 39 or 40 or 41 or 42 or 43 or 44 or 45 or 46 or 47 or 48 or 49 or 50 or 51 or 52 or 53 or 54 or 55 or 56 or 57 or 58 or 59 or 60 or 61 or 62 or 63 or 64 or 65 or 66 or 67 or 68 or 69 or 70 or 71 or 72 or 73 or 74 or 75 or 76 or 77 or 78 or 79 (256317)
81 randomized controlled trial.pt. (425254)
82 controlled clinical trial.pt. (91288)
83 randomized.ab. (364059)
84 placebo.ab. (176659)
85 clinical trials as topic.sh. (178334)
86 randomly.ab. (260090)
87 trial.ti. (158896)
88 81 or 82 or 83 or 84 or 85 or 86 or 87 (1054276)
89 exp animals/ not humans.sh. (4280763)
90 88 not 89 (972053)

91 80 and 90 (43387)
92 exp "Feeding and Eating Disorders"/ (25424)
93 exp Anorexia nervosa/ (11498)
94 exp Bulimia Nervosa/ (1865)
95 exp Suicide, attempted/ (16837)
96 exp Self mutilation/ (3121)
97 exp Self-injurious behavior/ (59536)
98 exp Mood disorders/ (103490)
99 exp Bipolar disorder/ (34994)
100 exp Neurotic disorders/ (17974)
101 exp Depressive disorder/ (92517)
102 exp Dysthymic disorder/ (1062)
103 exp depression/ (90326)
104 exp Seasonal affective disorder/ (1135)
105 exp anxiety/ or exp anxiety disorders/ or exp anxiety, separation/ or exp dental anxiety/ (128930)
106 exp panic/ or exp panic disorder/ (8717)
107 exp phobic disorders/ (10008)
108 exp combat disorders/ or exp stress disorders, post-traumatic/ (26209)
109 exp Somatoform disorders/ (17354)
110 exp Hypochondriasis/ (2197)
111 exp Hysteria/ (3526)
112 exp Conversion disorder/ (2612)
113 exp munchausen syndrome/ or munchausen syndrome by proxy/ (1739)
114 exp Neurasthenia/ (1346)
115 exp Fatigue syndrome, chronic/ (4796)
116 exp Obsessive-compulsive disorder/ (12603)
117 exp Obsessive behavior/ (1195)
118 exp Compulsive behavior/ (9361)
119 exp Stress, psychological/ (104897)
120 *Mental Disorders/ (111864)
121 exp schizophrenia/ (98181)
122 exp paranoid disorders/ (3903)
123 schizo$.mp. (153123)
124 hebephreni$.mp. (274)
125 oligophreni$.mp. (1102)
126 psychotic$.mp. (60934)
127 psychos#s.mp. (46522)
128 (chronic$ adj mental$).ti,ab. (1841)
129 (sever$ adj mental).ti,ab. (7050)
130 (mental$ adj disorder$).ti,ab. (29226)
131 (mental$ adj ill$).ti,ab. (28537)
132 (emotion$ adj disorder$).ti,ab. (2168)
133 exp "schizophrenia spectrum and other psychotic disorders"/ (137268)
134 or/92-133 (759610)
135 91 and 134 (3448)
136 exp dissociative disorders/ (3797)
137 exp personality disorders/ (37882)
138 134 or 136 or 137 (777550)
139 91 and 138 (3490)

EMBASE(Ovid)

1 ((participat$ or recruit$ or enrol$ or enter$ or entry) and (trial? or study)).ti. (13443)
2 (select$ adj3 (participants or patients or controls)).tw. (142299)
3 recruit$.ab. /freq=2 (60587)
4 participat$.ab. /freq=2 (73083)
5 research.tw. (1342128)
6 2 and (3 or 4 or 5) (9664)
7 (informed consent or consent process$ or consent procedure?).tw. (53284)
8 1 or 6 or 7 (75736)
9 (minimi$ adj2 attrition).ab,ti. (91)
10 (prevent$ adj2 attrition).ab,ti. (101)
11 (lessen$ adj2 attrition).ab,ti. (2)
12 (decreas$ adj2 attrition).ab,ti. (101)
13 (reduc$ adj2 attrition).ab,ti. (480)
14 (minimi$ adj2 drop-out).ab,ti. (9)
15 (prevent$ adj2 drop-out).ab,ti. (39)
16 (lessen$ adj2 drop-out).ab,ti. (1)
17 (decreas$ adj2 drop-out).ab,ti. (31)
18 (reduc$ adj2 drop-out).ab,ti. (101)
19 (minimi$ adj2 drop-out$).ab,ti. (13)
20 (prevent$ adj2 drop-out$).ab,ti. (47)
21 (lessen$ adj2 drop-out$).ab,ti. (1)
22 (decreas$ adj2 drop-out$).ab,ti. (40)
23 (reduc$ adj2 drop-out$).ab,ti. (122)
24 (minimi$ adj2 drop$-out).ab,ti. (10)
25 (prevent$ adj2 drop$-out).ab,ti. (46)
26 (lessen$ adj2 drop$-out).ab,ti. (1)
27 (decreas$ adj2 drop$-out).ab,ti. (36)
28 (reduc$ adj2 drop$-out).ab,ti. (112)
29 (minimi$ adj2 dropout$).ab,ti. (40)
30 (prevent$ adj2 dropout$).ab,ti. (128)
31 (lessen$ adj2 dropout$).ab,ti. (1)
32 (decreas$ adj2 dropout$).ab,ti. (77)
33 (reduc$ adj2 dropout$).ab,ti. (237)
34 (strateg$ adj2 drop$-out).ab,ti. (4)
35 (strateg$ adj2 dropout$).ab,ti. (11)
36 (loss adj2 follow-up).ab,ti. (4737)
37 (lost adj2 follow-up).ab,ti. (22872)
38 (loss adj2 followup).ab,ti. (98)
39 (lost adj2 followup).ab,ti. (891)
40 (minimi$ adj2 withdrawal).ab,ti. (144)
41 (prevent$ adj2 withdrawal).ab,ti. (504)
42 (lessen$ adj2 withdrawal).ab,ti. (9)
43 (decreas$ adj2 withdrawal).ab,ti. (868)
44 (reduc$ adj2 withdrawal).ab,ti. (1416)
45 (minimi$ adj2 withdrawal$).ab,ti. (151)
46 (prevent$ adj2 withdrawal$).ab,ti. (514)
47 (lessen$ adj2 withdrawal$).ab,ti. (9)
48 (decreas$ adj2 withdrawal$).ab,ti. (878)
49 (reduc$ adj2 withdrawal$).ab,ti. (1443)
50 (strateg$ adj2 attrition).ab,ti. (17)

51 (strateg$ adj2 drop-out).ab,ti. (3)
52 (strateg$ adj2 dropout).ab,ti. (9)
53 (strateg$ adj2 follow-up).ab,ti. (1539)
54 (strateg$ adj2 followup).ab,ti. (31)
55 (increas$ adj2 retention).ab,ti. (4810)
56 (encourag$ adj2 retention).ab,ti. (84)
57 (maximi$ adj2 retention).ab,ti. (169)
58 (promot$ adj2 retention).ab,ti. (854)
59 (improv$ adj2 retention).ab,ti. (2932)
60 (strateg$ adj2 response$).ab,ti. (2259)
61 (strateg$ adj2 (questionnaire$ adj3 response$)).ab,ti. (10)
62 (increas$ adj2 (questionnaire$ adj3 response$)).ab,ti. (53)
63 (encourag$ adj2 (questionnaire$ adj3 response$)).ab,ti. (1)
64 (maximi$ adj2 (questionnaire$ adj3 response$)).ab,ti. (5)
65 (promot$ adj2 (questionnaire$ adj3 response$)).ab,ti. (1)
66 (improv$ adj2 (questionnaire$ adj3 response$)).ab,ti. (30)
67 (increas$ adj2 response$).ab,ti. (42610)
68 (encourag$ adj2 response$).ab,ti. (839)
69 (maximi$ adj2 response$).ab,ti. (709)
70 (promot$ adj2 response$).ab,ti. (4797)
71 (improv$ adj2 response$).ab,ti. (15843)
72 (retention adj2 strateg$).ab,ti. (723)
73 retention rate$.ab,ti. (4255)
74 (retention adj2 method$).ab,ti. (874)
75 (retention adj2 technique$).ab,ti. (198)
76 attrition rate$.ab,ti. (2200)
77 (questionnaire$ adj3 (response$ adj2 method$)).ab,ti. (62)
78 (questionnaire$ adj3 (response adj2 technique$)).ab,ti. (2)
79 (questionnaire adj response rate$).ab,ti. (950)
80 (difficult$ adj2 (retain$ or retention)).ab,ti. (324)
81 Participant Dropouts/ (0)
82 9 or 10 or 11 or 12 or 13 or 14 or 15 or 16 or 17 or 18 or 19 or 20 or 21 or 22 or 23 or 24 or 25 or 26 or 27 or 28 or 29 or 30 or 31 or 32 or 33 or 34 or 35 or 36 or 37 or 38 or 39 or 40 or 41 or 42 or 43 or 44 or 45 or 46 or 47 or 48 or 49 or 50 or 51 or 52 or 53 or 54 or 55 or 56 or 57 or 58 or 59 or 60 or 61 or 62 or 63 or 64 or 65 or 66 or 67 or 68 or 69 or 70 or 71 or 72 or 73 or 74 or 75 or 76 or 77 or 78 or 79 or 80 or 81 (116847)
83 8 or 82 (191589)
84 random$.tw. (1094465)
85 placebo$.ti,ab,sh. (365359)
86 double-blind$.tw. (164070)
87 84 or 85 or 86 (1311578)
88 eating disorder/ or anorexia nervosa/ or binge eating disorder/ or bulimia/ (39306)
89 exp suicidal behavior/ (78654)
90 automutilation/ (12559)
91 exp mood disorder/ (396029)
92 exp neurosis/ (52551)
93 exp anxiety disorder/ (176035)
94 anxiety/ (145205)
95 dental anxiety/ (2057)

96 exp psychosomatic disorder/ (37699)
97 chronic fatigue syndrome/ (8104)
98 chronic stress/ or emotional stress/ or mental stress/ (85748)
99 *mental disease/ (93464)
100 exp psychosis/ (239967)
101 schizo$.mp. (195580)
102 hebephreni$.mp. (776)
103 oligophreni$.mp. (1526)
104 psychotic$.mp. (43108)
105 psychos#s.mp. (110772)
106 (chronic$ adj mental$).ti,ab. (2162)
107 (sever$ adj mental).ti,ab. (8886)
108 (mental$ adj disorder$).ti,ab. (36208)
109 (mental$ adj ill$).ti,ab. (34176)
110 (emotion$ adj disorders$).ti,ab. (2345)
111 or/88-110 (1051509)
112 83 and 87 and 111 (2861)

PsycINFO(Ovid)

1 Patient Selection/ (186)
2 ((participat$ or recruit$ or enrol$) adj4 trial?).tw. (3138)
3 1 or 2 (3320)
4 Informed Consent/ (3704)
5 informed consent.tw. (6695)
6 4 or 5 (7730)
7 exp Clinical Trial/ (9713)
8 Experimental Subjects/ (3778)
9 (trial? or study or studies or research).tw. (2039814)
10 7 or 8 or 9 (2040228)
11 3 or (6 and 10) (8130)
12 (minimi$ adj2 attrition).ab,ti. (49)
13 (prevent$ adj2 attrition).ab,ti. (66)
14 (lessen$ adj2 attrition).ab,ti. (2)
15 (decreas$ adj2 attrition).ab,ti. (95)
16 (reduc$ adj2 attrition).ab,ti. (271)
17 (minimi$ adj2 drop-out).ab,ti. (5)
18 (prevent$ adj2 drop-out).ab,ti. (52)
19 (lessen$ adj2 drop-out).ab,ti. (0)
20 (decreas$ adj2 drop-out).ab,ti. (16)
21 (reduc$ adj2 drop-out).ab,ti. (57)
22 (minimi$ adj2 drop-out$).ab,ti. (5)
23 (prevent$ adj2 drop-out$).ab,ti. (55)
24 (lessen$ adj2 drop-out$).ab,ti. (0)
25 (decreas$ adj2 drop-out$).ab,ti. (19)
26 (reduc$ adj2 drop-out$).ab,ti. (67)
27 (minimi$ adj2 drop$-out).ab,ti. (5)
28 (prevent$ adj2 drop$-out).ab,ti. (65)
29 (lessen$ adj2 drop$-out).ab,ti. (0)
30 (decreas$ adj2 drop$-out).ab,ti. (20)
31 (reduc$ adj2 drop$-out).ab,ti. (62)
32 (minimi$ adj2 dropout$).ab,ti. (15)

33 (prevent$ adj2 dropout$).ab,ti. (317)
34 (lessen$ adj2 dropout$).ab,ti. (0)
35 (decreas$ adj2 dropout$).ab,ti. (99)
36 (reduc$ adj2 dropout$).ab,ti. (226)
37 (strateg$ adj2 drop$-out).ab,ti. (6)
38 (strateg$ adj2 dropout$).ab,ti. (27)
39 (loss adj2 follow-up).ab,ti. (355)
40 (lost adj2 follow-up).ab,ti. (597)
41 (loss adj2 followup).ab,ti. (8)
42 (lost adj2 followup).ab,ti. (7)
43 (minimi$ adj2 withdrawal).ab,ti. (18)
44 (prevent$ adj2 withdrawal).ab,ti. (139)
45 (lessen$ adj2 withdrawal).ab,ti. (9)
46 (decreas$ adj2 withdrawal).ab,ti. (298)
47 (reduc$ adj2 withdrawal).ab,ti. (457)
48 (minimi$ adj2 withdrawal$).ab,ti. (19)
49 (prevent$ adj2 withdrawal$).ab,ti. (144)
50 (lessen$ adj2 withdrawal$).ab,ti. (9)
51 (decreas$ adj2 withdrawal$).ab,ti. (303)
52 (reduc$ adj2 withdrawal$).ab,ti. (463)
53 (strateg$ adj2 attrition).ab,ti. (20)
54 (strateg$ adj2 drop-out).ab,ti. (3)
55 (strateg$ adj2 dropout).ab,ti. (26)
56 (strateg$ adj2 follow-up).ab,ti. (127)
57 (strateg$ adj2 followup).ab,ti. (2)
58 (increas$ adj2 retention).ab,ti. (1089)
59 (encourag$ adj2 retention).ab,ti. (57)
60 (maximi$ adj2 retention).ab,ti. (64)
61 (promot$ adj2 retention).ab,ti. (343)
62 (improv$ adj2 retention).ab,ti. (1261)
63 (strateg$ adj2 response$).ab,ti. (2034)
64 (strateg$ adj2 (questionnaire$ adj3 response$)).ab,ti. (7)
65 (increas$ adj2 (questionnaire$ adj3 response$)).ab,ti. (16)
66 (encourag$ adj2 (questionnaire$ adj3 response$)).ab,ti. (1)
67 (maximi$ adj2 (questionnaire$ adj3 response$)).ab,ti. (4)
68 (promot$ adj2 (questionnaire$ adj3 response$)).ab,ti. (0)
69 (improv$ adj2 (questionnaire$ adj3 response$)).ab,ti. (12)
70 (increas$ adj2 response$).ab,ti. (7838)
71 (encourag$ adj2 response$).ab,ti. (169)
72 (maximi$ adj2 response$).ab,ti. (140)
73 (promot$ adj2 response$).ab,ti. (365)
74 (improv$ adj2 response$).ab,ti. (1542)
75 (retention adj2 strateg$).ab,ti. (513)
76 retention rate$.ab,ti. (1674)
77 (retention adj2 method$).ab,ti. (172)
78 (retention adj2 technique$).ab,ti. (16)
79 attrition rate$.ab,ti. (1192)
80 (questionnaire$ adj3 (response$ adj2 method$)).ab,ti. (20)
81 (questionnaire$ adj3 (response adj2 technique$)).ab,ti. (3)
82 (questionnaire adj response rate$).ab,ti. (175)
83 (difficult$ adj2 (retain$ or retention)).ab,ti. (147)

84 exp Treatment Dropouts/ (2227)
85 12 or 13 or 14 or 15 or 16 or 17 or 18 or 19 or 20 or 21 or 22 or 23 or 24 or 25 or 26 or 27 or 28 or 29 or 30 or 31 or 32 or 33 or 34 or 35 or 36 or 37 or 38 or 39 or 40 or 41 or 42 or 43 or 44 or 45 or 46 or 47 or 48 or 49 or 50 or 51 or 52 or 53 or 54 or 55 or 56 or 57 or 58 or 59 or 60 or 61 or 62 or 63 or 64 or 65 or 66 or 67 or 68 or 69 or 70 or 71 or 72 or 73 or 74 or 75 or 76 or 77 or 78 or 79 or 80 or 81 or 82 or 83 or 84 (22941)
86 double-blind.ab,ti. (19918)
87 "random$ assigned.".ab,ti. (28515)
88 control.ab,ti. (356849)
89 86 or 87 or 88 (389728)
90 exp Experimental Attrition/ (365)
91 85 or 90 (23177)
92 89 and 91 (3791)
93 exp eating disorders/ (25722)
94 exp attempted suicide/ (8675)
95 self-mutilation/ (1105)
96 exp Self-Injurious Behavior/ (4198)
97 exp affective disorders/ (136500)
98 exp neurosis/ (7571)
99 anxiety disorders/ (15529)
100 exp somatoform disorders/ (11519)
101 exp HYSTERIA/ (1981)
102 munchausen syndrome/ or munchausen syndrome by proxy/ (321)
103 exp chronic fatigue syndrome/ (1697)
104 exp psychological stress/ (8067)
105 *mental disorders/ (58903)
106 exp psychosis/ (100398)
107 schizoaffective disorder/ (2760)
108 schizo$.mp. (119741)
109 hebephreni$.mp. (533)
110 oligophreni$.mp. (520)
111 psychotic$.mp. (41118)
112 psychos#s.mp. (54151)
113 (chronic$ adj mental$).ti,ab. (2443)
114 (severe adj mental).ti,ab. (6322)
115 (mental$ adj disorder$).ti,ab. (43030)
116 (mental$ adj ill$).ti,ab. (40799)
117 (emotion$ adj disorder$).ti,ab. (3073)
118 or/93-117 (429478)
119 92 and 118 (492)
120 exp dissociative disorders/ (4873)
121 exp Personality Disorders/ (30149)
122 118 or 120 or 121 (452579)
123 122 and 92 (508)

Cochrane Methodology Register(CMR)

#1 (minimi* near/2 attrition):ab,ti
#2 (prevent* near/2 attrition):ab,ti
#3 (lessen* near/2 attrition):ab,ti
#4 (decreas* near/2 attrition):ab,ti

#5 (reduc* near/2 attrition):ab,ti
#6 (minimi* near/2 drop-out):ab,ti
#7 (prevent* near/2 drop-out):ab,ti
#8 (lessen* near/2 drop-out):ab,ti
#9 (decreas* near/2 drop-out):ab,ti
#10 (reduc* near/2 drop-out):ab,ti
#11 (minimi* near/2 drop-out*):ab,ti
#12 (prevent* near/2 drop-out*):ab,ti
#13 (lessen* near/2 drop-out*):ab,ti
#14 (decreas* near/2 drop-out*):ab,ti
#15 (reduc* near/2 drop-out*):ab,ti
#16 (minimi* near/2 drop*-out):ab,ti
#17 (prevent* near/2 drop*-out):ab,ti
#18 (lessen* near/2 drop*-out):ab,ti
#19 (decreas* near/2 drop*-out):ab,ti
#20 (reduc* near/2 drop*-out):ab,ti
#21 (minimi* near/2 dropout*):ab,ti
#22 (prevent* near/2 dropout*):ab,ti
#23 (lessen* near/2 dropout*):ab,ti
#24 (decreas* near/2 dropout*):ab,ti
#25 (reduc* near/2 dropout*) .ab,ti
#26 (strateg* near/2 (questionnaire* near/3 response*)):ab,ti
#27 (increas* near/2 (questionnaire* near/3 response*)):ab,ti
#28 (encourag* near/2 (questionnaire* near/3 response*)):ab,ti
#29 (maximi* near/2 (questionnaire* near/3 response*)):ab,ti
#30 (promot* near/2 (questionnaire* near/3 response*)):ab,ti
#31 (improv* near/2 (questionnaire* near/3 response*)):ab,ti
#32 (increas* near/2 response*):ab,ti
#33 (encourag* near/2 response*):ab,ti
#34 (maximi* near/2 response*):ab,ti
#35 (promot* near/2 response*):ab,ti
#36 (improv* near/2 response*):ab,ti
#37 (retention near/2 strateg*):ab,ti
#38 retention rate*:ab,ti
#39 (retention near/2 method*):ab,ti
#40 (retention near/2 technique*):ab,ti
#41 attrition rate*:ab,ti
#42 (questionnaire* near/3 (response* near/2 method*)):ab,ti
#43 (questionnaire* near/3 (response near/2 technique*)):ab,ti
#44 (questionnaire near response rate*):ab,ti (1145)
#45 (difficult* near/2 (retain* or retention)):ab,ti
#46 MeSH descriptor: [Patient Dropouts] explode all trees
#47 (strateg* near/2 drop*-out):ab,ti
#48 (strateg* near/2 dropout*):ab,ti
#49 (loss near/2 follow-up):ab,ti
#50 (lost near/2 follow-up):ab,ti
#51 (loss near/2 followup):ab,ti
#52 (lost near/2 followup):ab,ti
#53 (minimi* near/2 withdrawal):ab,ti
#54 (prevent* near/2 withdrawal):ab,ti
#55 (lessen* near/2 withdrawal):ab,ti
#56 (decreas* near/2 withdrawal):ab,ti

#57 (reduc* near/2 withdrawal):ab,ti
#58 (minimi* near/2 withdrawal*):ab,ti
#59 (prevent* near/2 withdrawal*):ab,ti
#60 (lessen* near/2 withdrawal*):ab,ti
#61 (decreas* near/2 withdrawal*):ab,ti
#62 (reduc* near/2 withdrawal*):ab,ti
#63 (strateg* near/2 attrition):ab,ti
#64 (strateg* near/2 drop-out):ab,ti
#65 (strateg* near/2 dropout):ab,ti
#66 (strateg* near/2 follow-up):ab,ti
#67 (strateg* near/2 followup):ab,ti
#68 (increas* near/2 retention):ab,ti
#69 (encourag* near/2 retention):ab,ti
#70 (maximi* near/2 retention):ab,ti
#71 (promot* near/2 retention):ab,ti
#72 (improv* near/2 retention):ab,ti
#73 (strateg* near/2 response*):ab,ti
#74 "accrual and sample size":kw or "attitudes to trials":kw or "informed consent":kw
#75 (participat* or recruit* or enrol* or select*) near/8 (trial* or research or study):ti or (participat* or recruit* or enrol* or select*) near/8 (trial* or research or study):ab
#76 #1 or #2 or #3 or #4 or #5 or #6 or #7 or #8 or #9 or #10 or #11 or #12 or #13 or #14 or #15 or #16 or #17 or #18 or #19 or #20 or #21 or #22 or #23 or #24 or #25 or #26 or #27 or #28 or #29 or #30 or #31 or #32 or #33 or #34 or #35 or #36 or #37 or #38 or #39 or #40 or #41 or #42 or #43 or #44 or #45 or #46 or #47 or #48 or #49 or #50 or #51 or #52 or #53 or #54 or #55 or #56 or #57 or #58 or #59 or #60 or #61 or #62 or #63 or #64 or #65 or #66 or #67 or #68 or #69 or #70 or #71 or #72 or #73 or #74 or #75
#77 "accrual and sample size":kw or "attitudes to trials":kw or "informed consent":kw
#78 (participat* or recruit* or enrol* or select*) near/8 (trial* or research or study):ti or (participat* or recruit* or enrol* or select*) near/8 (trial* or research or study):ab
#79 (#77 or #78 or #76)
#80 MeSH descriptor: [Anorexia Nervosa] explode all trees
#81 MeSH descriptor: [Feeding and Eating Disorders] explode all trees
#82 MeSH descriptor: [Bulimia Nervosa] explode all trees
#83 MeSH descriptor: [Suicide, Attempted] explode all trees
#84 MeSH descriptor: [Self Mutilation] explode all trees
#85 MeSH descriptor: [Self-Injurious Behavior] explode all trees
#86 MeSH descriptor: [Mood Disorders] explode all trees
#87 MeSH descriptor: [Bipolar Disorder] explode all trees
#88 MeSH descriptor: [Neurotic Disorders] explode all trees
#89 MeSH descriptor: [Depressive Disorder] explode all trees
#90 MeSH descriptor: [Depression] explode all trees
#91 MeSH descriptor: [Depressive Disorder, Major] explode all trees
#92 MeSH descriptor: [Depression, Postpartum] explode all trees
#93 MeSH descriptor: [Seasonal Affective Disorder] explode all trees
#94 MeSH descriptor: [Anxiety] explode all trees
#95 MeSH descriptor: [Anxiety Disorders] explode all trees
#96 MeSH descriptor: [Stress, Psychological] explode all trees

#97 [mh ^" Mental Disorders " [mj]]
#98 MeSH descriptor: [Schizophrenia] explode all trees
#99 MeSH descriptor: [Paranoid Disorders] explode all trees
#100 schizo*
#101 psychotic*
#102 hebephreni*
#103 oligophreni*
#104 psychos*s
#105 (chronic* near mental*):ab,ti
#106 (sever* near mental):ab,ti
#107 (mental* near disorder*):ti,ab
#108 (mental* near ill*):ti,ab
#109 MeSH descriptor: [Panic] explode all trees
#110 MeSH descriptor: [Panic Disorder] explode all trees
#111 MeSH descriptor: [Phobic Disorders] explode all trees
#112 MeSH descriptor: [Combat Disorders] explode all trees
#113 MeSH descriptor: [Stress Disorders, Post-Traumatic] explode all trees
#114 MeSH descriptor: [Somatoform Disorders] explode all trees
#115 MeSH descriptor: [Hypochondriasis] explode all trees
#116 MeSH descriptor: [Hysteria] explode all trees
#117 MeSH descriptor: [Conversion Disorder] explode all trees
#118 MeSH descriptor: [Munchausen Syndrome] explode all trees
#119 MeSH descriptor: [Munchausen Syndrome by Proxy] explode all trees
#120 MeSH descriptor: [Neurasthenia] explode all trees
#121 MeSH descriptor: [Fatigue Syndrome, Chronic] explode all trees
#122 MeSH descriptor: [Obsessive-Compulsive Disorder] explode all trees
#123 MeSH descriptor: [Obsessive Behavior] explode all trees
#124 MeSH descriptor: [Personality Disorders] explode all trees
#125 MeSH descriptor: [Dissociative Disorders] explode all trees
#126 MeSH descriptor: [Compulsive Behavior] explode all trees
#127 (emotion* near disorder*):ab,ti
#128 #80 or #81 or #82 or #83 or #84 or #85 or #86 or #87 or #88 or #89 or #90 or #91 or #92 or #93 or #94 or #95 or #96 or #97 or #98 or #99 or #100 or #101 or #102 or #103 or #104 or #105 or #106 or #107 or #108 or #109 or #110 or #111 or #112 or #114 or #115 or #116 or #117 or #118 or #119 or #120 or #121 or #122 or #123 or #126 or #127 or #124 or #125

Appendix C Risk of bias assessment for the included studies in Chapter 2

Study ID	Was true randomization used for assignment of participants to treatment groups?	Was allocation to treatment groups concealed?	Were treatment groups similar at the baseline?	Were participants blind to treatment assignment?	Were those delivering treatment blind to treatment assignment?	Were outcomes assessors blind to treatment assignment?	Were treatment groups treated identically other than the intervention of interest?	Was follow up complete and if not, were differences between groups in terms of their follow up adequately described and analysed?	Were participants analysed in the groups to which they were	Were outcomes measured in the same way for treatment groups?	Were outcomes measured in a reliable way?	Was appropriate statistical analysis used?	Was the trial design appropriate, and any deviations from the standard	Overall
Man 2015	Y	N/A	Y	N/A	N/A	Unclear	Y	N/A	Y	Y	Y	Y	Y	8
Jeste 2009	Unclear	N/A	Y	N/A	N/A	Unclear	Y	N/A	Y	Y	Y	Y	Y	7
McLean 2014	Y	N/A	Y	N/A	N/A	Unclear	Unclear	Y	Y	Y	Y	Y	Y	8
Dirmaier 2007	Y	N/A	Y	Y	Unclear	Unclear	Y	N/A	Y	Y	Y	Y	Y	9
Hughes-Morley 2016	Y	N/A	Y	N/A	Y	Unclear	Y	N/A	Y	Y	Y	Y	Y	9

Appendix D Costing and sensitivity analysis for the recruitment strategies in Chapter 2

Study ID	Cost resource kind	Description in text	Resource used for extrapolation	Cost calculation	Notes	Possible minimum	Possible maximum
DeBar 2009	coffee shop coupon $5	"those individuals completing the screener online were mailed a $5 coffee shop coupon for their efforts (no compensation was provided for response by mail) "	$5	$5*1415		7075.00	7075.00
DeBar 2009	pre-paid return envelope mail out	"Respondents had the option of completing a paper and pencil version of the screener and returning it by mail via a pre-paid return envelope or completing an online version."	USPS Priority Mail Forever Prepaid Flat Rate Padded Envelope	£0.73 per envelope Royal Mail 1st class		2513.39	2513.39
DeBar 2009	online questionnaire	"Respondents had the option of completing a paper and pencil version of the screener and returning it by mail via a pre-paid return envelope or completing an online version."	Stewart Kirk Social Research & Evaluation	$3.50 per response (however the study did not specify how many individuals completed online screening)		8519.00	8519.00
DeBar 2009	telephone interview	"(1–2 h phone interview)" + "$50 for baseline assessment"	Stewart Kirk Social Research & Evaluation	$50 per interview *3014 individuals interviewed = $150,700		301400.00	301400.00

DeBar 2009	incentives for baseline assessment	"$25 for baseline assessment"		25*2863		71575.00	71575.00
Jeste 2009	Multimedia consent	"The RA provided participants assigned to the multimedia consent with the printed consent form, and rather than listening to the RA read the consent aloud, subjects watched a DVD that explained the protocol. Subjects were encouraged to have the RA stop the DVD and repeat any segments that were unclear. "	RA manpower, DVD material development & printing of the consent form	1 RA salary Band 7 £38,786(£52/hr); DVD material development cost approximately £1,500; consent form printing (10 A4 pages)	No recruitment period mentioned, for hypothetical trial, assuming recruitment period from 4 weeks to 3 months, varies from 2 hrs/day to full time(7hrs/day)	3640.00	26130.00
Jeste 2009	Routine consent	"To control for time spent with an RA and the novelty of a DVD, we added a 10-minute control DVD (describing general information about research) to the routine consent procedure. Subjects assigned to the routine consent procedure first met individually with a trained RA to view this DVD. Next, the RA provided the subjects with the printed simulated consent form and encouraged them to read along while the RA read it aloud."	RA manpower, DVD material development & printing of the consent form	1 RA salary Band 7 £38,786; DVD material development cost approximately £1,300; consent form printing (10 A4 pages)	No recruitment period mentioned; Cost would not differ too much for the 2 strategies(B7 & B8)	3440.00	25930.00

Morgan 2013	online forum	"We approached 58 forums related to depression and other related problems, but only 25 responded with permission to post about the study."; "Although some smaller forums gave permission to post, this was usually a time-consuming process involving signing up to the forum and creating a user account, then identifying the forum moderator or administrator and contacting them for permission to post, and then finally submitting the post and monitoring responses. "	manpower	assuming from 2hrs/day to full time responsible for mailing and posting, salary Band 7 £38,786.(£52/hr)	recruitment from Feb2010 to March 2011(13 months). However no information on how many hours dedicated to this strategy.	25740.00	38786.00
Morgan 2013	links from webpage	"A new page of supporters was created to accommodate this requirement. This page thanked each organization or website that had helped promote the study to participants. Some websites were generous and included a link and blurb on their home page; others listed the website within a section of their site that contained links to other interesting websites."	more of a design feature for the website, difficult to estimate the cost.	assuming from 2hrs/day to full time responsible for mailing and posting, salary Band 7 £38,786.(£52/hr)	recruitment from Feb2010 to March 2011(13 months). However no information on how many hours dedicated to such strategy.	25740.00	38786.00

Morgan 2013	Online community noticeboards	"We explored the effectiveness of posting an invitation to participate in the study on websites that function as online community noticeboards (e.g., Craigslist and Gumtree). It is free to post a classified ad on these websites, and other survey-based studies have found them an effective recruitment source. However, although free, these websites are designed to offer products and services to local residents only."; "We posted the study notice 4 times over the recruitment period, and observed a noticeable increase in visits and enrollments coinciding with each posting."	manpower (could be an RA)	assuming from 2hrs/day to full time responsible for mailing and posting, salary Band 7 £38,786.(£52/hr)	recruitment period Feb 2010 till Mar 2011 However no information on how many hours dedicated to such strategy.	25740.00	38786.00
Morgan 2013	Email Groups or Lists	"Yahoo! and Google provide a free service in which individuals with shared interests can join an online group and share messages and information. These messages are sent to email accounts or can be viewed in a Web browser. There are hundreds of groups related to mental health conditions or risk factors, but many of these have few members or have been	manpower (could be an RA)	assuming from 2hrs/day to full time responsible for mailing and posting, salary Band 7 £38,786.(£52/hr)	recruitment period Feb 2010 till Mar 2012	25740.00	38786.00

		overtaken by spam messages. We contacted 103 relevant groups with a reasonable number of members to advertise the study and 32 gave their permission. "					
Rollman 2008	Waitroom Recruitment	"A study recruiter stationed in a practice waiting room administered the PRIME-MD patient questionnaire (PQ)11 to screen patients for the presence of an anxiety disorder. "	research assistants manpower (22 months)	RA Grade7 salary (£52/hr)£38, 786/yr, total cost £77,572	assumption: 3 hrs/day - full time	77220.00	77572.00
Rollman 2008	EMR promoting	"EpicCare® Ambulatory Electronic Medical Record, Madison, WI"; "The EMR has an integral messaging system that permits PCPs and practice staff to communicate with each other and to document care."	GP manpower for referral (22 months)	In Krusche_2014 cost per patient randomised is £407.65		407.65	407.65
Man 2015	Optimised written patient information material	"The new versions of the Healthlines information sheets evaluated in the embedded trials were presented as 4-page A4 booklets and were divided into 8 sections (as compared with 15 sections in the original versions), with contrasting colour and a	costs of developing the optimised material, reminder mail & prepaid envelope	According to commercial design service(AM PM graphics), 4-page A4		1007.90	1007.90

	larger font used for section headings, to aid navigation. The front page contained a ‘bulleted’ list of trial summary information, logos, a contents list and contact details, which were jointly intended to aid reader navigation and have visual appeal. The re-wording of the information sheets included greater use of lay terms, short sentences and paragraphs. Additionally, the covering letters were revised by shortening them by around one third, particularly by removing content that was replicated in the information sheets, as well as adding ‘bullets’ for lists and using bold, lower case text for emphasis." ; "a reminder letter was mailed out to non-responders after 2 weeks in Healthlines Depression" ; "The patient was requested to respond by returning either a valid consent form or a decline form in a pre-paid, pre-addressed envelope. "	booklet design cost £195. 1st class mail price = 64p, prepaid 1st class envelope = 73p (according to royal mail 2016) Total cost estimation 682*(0.73+ 0.64)+£200(booklet printing cost)+195= £1007.9

Man 2015	Original consent material	"The original Healthlines patient information materials for both trials were 8-page A5 patient information sheets in booklet form, together with an A4 covering letter from the general practitioner (GP)."	cost of developing original materials	approximately £110 for 700 A5 booklets from commercial service, 1st class mail price = 64p, total cost estimation 682*(0.64)+ 110= £546.48		546.48	546.48
Woolhouse 2014	Research assistant attending antenatal booking clinic	"A female research assistant (HW) approached women attending the Royal Women's Hospital antenatal clinic, and invited them to take part in the MindBabyBody pilot study, described as an evaluation of a group program designed to 'help you reduce stress and manage your mood' during pregnancy and the postnatal period. Women who expressed interest were provided with a Study Recruitment Pack, and invited to complete study materials and return them directly to the research team." ;"HW attended the antenatal clinic waiting room	RA 6*4hr + 50 information packs	RA salary £52/hr(Band 7, according to PSSRU), £60 for 50 A4 brochures. Total cost estimation = £1,308	all cost information 2016	1308.00	1308.00

		on a total of six occasions, for four hours on each occasion, distributing approximately 50 Study Information Packs, a recruitment strategy which yielded 14 participants."					
Woolhouse 2014	Recruitment via a study brochure mailed to women at time of booking in	"A study information brochure was included with information sent to women when they made initial contact with the Women's to book in to give birth at the hospital. The brochure provided information about the study and invited women to contact the study investigator directly if they were interested in participating. A Study Recruitment Pack was sent to all women who inquired about the study, with study materials then mailed directly back to the research team."; "A total of 2500 brochures were mailed out in booking in packs over the 8-month period, yielding 16 participants."	2500 brochures + mailing	about £1,500 for 2500 A4 brochures, mail price = £0.73(1st class). Total cost estimation = £3,325	all cost information 2016	3325.00	3325.00

Woolhouse 2014	Recruitment via childbirth education and physiotherapy classes	"Staff members responsible for childbirth education classes, and antenatal physiotherapy classes were provided with information about the program, and encouraged to pass on study brochures to women attending their classes. Through the study information brochure, women were invited to contact the study investigator directly if they were interested in participating, and were then sent a Study Recruitment Pack in the mail, which was returned directly to the research team via the reply paid envelope. "	prepaid envelope	£0.73 per envelope Royal Mail 1st class	no information provided on how much envelopes were sent, assuming range from 50 to 200	36.50	146.00
Daley 2008	Identified by GPs from records	"All general practices within four primary care trusts in Birmingham were informed by letter about the aims of the study. They were invited to assist with recruitment by identifying potentially eligible patients." "General practices and the mother and baby unit then sent invitation letters to eligible women and asked them to contact the trial team if they were interested in participating. "	GP manpower, mailing	GP London salary £72,414 (median); 1st class mail £64 for a sheet of 100 stamps (2016). Total mailing cost 96/100*64= £ 61.44	according to Krusche_2014, GP referral cost 407.65/patient	9845.04	9845.04

Daley 2008	Identified by ‘mother and baby unit’	" The local specialist mother and baby unit was asked to do the same." "General practices and the mother and baby unit then sent invitation letters to eligible women and asked them to contact the trial team if they were interested in participating. "	Mailing, specialist manpower	total mailing cost 28/100*64= £17.92	assuming 1 community based specialist(band 7, salary £52/hr, 38786/yr), assuming time engaged from 3 hr/day to full time	24587.92	49157.92
Daley 2008	Identified by health visitors	"Presentations about the trial were made to health visitors in each of the primary care trusts, and they were asked to refer eligible women."		difficult to estimate operational cost			
Le 2007	from bilingual and mostly bicultural family support workers	"At the Mary’s Centre, they are trained to provide intake appointments, conduct psychosocial needs assessments, and provide case management for their clients. All potential clients take a pregnancy test, and if positive, they were assigned to one of the Mary’s Centre family support workers who functions as a case manager during and after pregnancy."	family support workers	According to PSSRU 2016 cost reference, Band 3 cost £25/hr, salary £18,640, assuming similar recruitment window and level of manpower, total cost 18640*8*2= 298,240	no numbers of workers mentioned, assuming full time	298240.00	298240.00

Le 2007	from research staff who recruited potential participants in the waiting rooms	"all staff were trained to be sensitive to recruiting individuals within a busy community health clinic, and to be culturally sensitive (e.g., using personalismo) when approaching women. The research staff included two advanced undergraduate students majoring in psychology, two graduate clinical psychology students, and four post BA and MA level staff."	8 research staff	RA Salary Band 7 £38,786, total cost (£52/hr) 38786*2*8= £620,576	recruitment from Jan 2005 till Dec 2006(2 yrs.) assuming from 3 hr/day to full time	524160.00	620576.00
Schlernitzauer 1998	Advertisement in newspaper	We further expanded recruitment strategies through the use of advertisements in local newspapers after review and approval by the Biomedical Institutional Review Board and the legal department of the University of Pittsburgh Medical Center. The advertisement was published repeatedly over several years, generally appearing in the front section of the paper, within the first 10 pages.	Local newspaper: take the Evening Standard, the rates of adverts for a clinical trial is 65/scc/day. For the message in the text, a 2*9 column appears to be suitable. Assuming advertising once a week, total cost over 5 years = 52*5*65*18=304200	Assuming advertising once a week, total cost over 5 years = 52*5*65*18 =304200	rates for a column of similar size differs between newspapers; also, frequency of advertising may vary	304200.00	1000000.00
HughesMorley_2016*	PPIR leaflet	"We then organised an expert workshop involving 27 key stakeholders including 10 service users with	Organising workshop: assumed no cost. Sending leaflet to	0.73*5382+ 60+20 = 4008.86		4008.86	4088.86

		severe mental illness and two carers of people with severe mental illness, who were either EQUIP PPIR members or belonged to the EQUIP trial target population." "Once the initial version was developed, we asked for contributions from the EQUIP host trial researchers (chiefly to check for accuracy); their input did not change the content or format of the leaflet. The leaflet was then sent to a professional graphic designer in a company with significant expertise in designing patient communication materials (www.makingsense.co.uk)."	professional graphic designer : consultation costs assumed £20 to £60 /hr, 1-2 hr spent. Cost of printing leaflets (n=5382): £60 - £100 (https://www.affordableleaflets.co.uk/leaflets/a5-leaflets.html) . Mailing cost £0.73/stamp			
HughesMorley_2016*	Mailing invitation		Mailing cost £0.73/stamp	0.73 * 2800 = 2044	2044.00	2044.00

Appendix E Search strategies for Chapter 3

MEDLINE (Ovid)

1. *Mental Disorders/
2. exp schizophrenia/
3. exp paranoid disorders/
4. schizo$.mp.
5. hebephreni$.mp.
6. oligophreni$.mp.
7. psychotic$.mp.
8. psychos#s.mp.
9. (chronic$ adj mental$).ti,ab.
10. (sever$ adj mental).ti,ab.
11. (mental$ adj disorder$).ti,ab.
12. (emotion$ adj disorder$).ti,ab.
13. exp "schizophrenia spectrum and other psychotic disorders"/
14. or/1-13
15. economics/
16. exp "Costs and Cost Analysis"/
17. cost of illness/
18. exp health care costs/
19. economic value of life/
20. exp economics medical/
21. exp economics hospital/
22. economics pharmaceutical/
23. exp "Fees and Charges"/
24. (econom$or cost or costs or costly or costing or price or pricing or pharmacoeconomic$).tw.
25. (value adj1 money).tw.
26. expenditure$not energy.tw.
27. budget$.tw.
28. cost benefit analysis/
29. cost effectiveness analysis/
30. cost minimization analysis/
31. cost utility analysis/
32. economic evaluation/
33. or/15-32
34. 14 and 33

University of York Centre for Reviews and Dissemination database

1. MeSH DESCRIPTOR Mental Disorders EXPLODE ALL TREES
2. MeSH DESCRIPTOR Antipsychotic Agents EXPLODE ALL TREES IN NHSEED
3. MeSH DESCRIPTOR Schizophrenia EXPLODE 1 IN NHSEED
4. MeSH DESCRIPTOR Psychotic Disorders EXPLODE 1 IN NHSEED
5. MeSH DESCRIPTOR Antipsychotic Agents EXPLODE ALL TREES
6. MeSH DESCRIPTOR Antipsychotic Agents EXPLODE ALL TREES IN NHSEED
7. 3 OR #4 OR #6

Appendix F Analysis code script for the cost-effectiveness model in Chapter 4 and Chapter 5 (implemented in R)

R code for Chapter 4

```
nYr <- 10
nPt <- 1000
##Setting the model structure / reset model parameters
timeseq <- seq(0, nYr, 1/12) #cycle sequence
nCycle <- length(timeseq) #number of cycles
nStates <- c(1,2,3) #Health States 'Stable' == 1, 'Relapsed' == 2,'Death' == 3
CVDstates <- c(0,1,2) #Cardiovascular events 'no CVD' == 0, 'Primary CHD' ==
1, 'Primary CAV' == 2
CHDstates <- c(1,2,3,4,5) #CHD subtype "stable angina" == 1, "unstable angina"
== 2, "MI" == 3, "Surgery"==4, others == 5, fatal==6
CVAstates <- c(1,2,3,4) #CVA subtype "TIA"==1, "haemorrhagic stroke" == 2,
"ischaemic stoke" == 3, others == 4, fatal == 5s
trans_array <- array(0,dim = c(nCycle,length(nStates),1000))#setting initial
transition probability matrix, using a 3-d array
Stateno <- matrix(0, nrow=nCycle+1, ncol = 1000) #retruns which health state
individual is in at current cycle
Stateno [1,] <- 1
Stateno [2,] <- 1 #all individual starts at stable
i_compli <- rep(0,1000) #compliance indicator

#initiating other comorbidities - CJS
r_DMII <- rep(0,1000)
r_CJS <- rep(0,1000)
r_court <- rep(0,1000)
r_prison <- rep(0,1000)
r_EPS <- rep(0,1000)
r_relapse <- rep(0,1000)
state_CVD <- matrix(0, nrow=1000, ncol = nCycle)
state_CHD <- matrix(0, nrow=1000, ncol = nCycle)
state_CVA <- matrix(0, nrow=1000, ncol = nCycle)
state_EPS <-rep(0,1000)
state_DMII <- rep(0,1000)
state_CJS <- rep(0,1000)
state_court <- rep(0,1000)
state_prison <- rep(0,1000)
cycle_prison <- rep(0,1000)
rel_prison <- rep(0,1000)
r_CVA <- rep(0,1000)
r_CHD <- rep(0,1000)
servicecost_month <- matrix(0,nrow=nPt, ncol=nCycle)
societalcost_month <- matrix(0,nrow=nPt, ncol=nCycle)
relapsecost_month <- matrix(0, nrow=nPt, ncol=nCycle)
CVDcost_month <- matrix(0,nrow=nPt, ncol=nCycle)
CVDcost <- matrix(0,nrow=nPt, ncol=nYr)
relapsecost <- matrix(0,nrow=nPt, ncol=nYr)
servicecost <- matrix(0,nrow=nPt, ncol=nYr)
societalcost <- matrix(0,nrow=nPt, ncol=nYr)
psychinpat_month <- matrix(0,nrow=nPt, ncol=nCycle)
psychinpat_yr <- matrix(0,nrow=nPt, ncol=nYr)
utility <- matrix(0, nrow=1000, ncol = nCycle)
QALY<-matrix(0, nrow=1000, ncol=nYr)
discount <- 0.035
r_radar <- matrix(0, nrow=nPt, ncol=nCycle+1)
extra<-as.data.frame(matrix(0,1,1000))
extracol <- rep(0,1000)


################################################################################
################################################################################
#################################### Variable list
################################################################################
################################################################################
```

```
############################  Transition probabilities
#########################

#monthly relapse risk
  r_relapse <- rep(0.008558231, 1000) #base case maintenance relapse
  #r_relapse <- rep(0.009414,1000) #base case RADAR relapse (best case
scenario, 10% non-inferiority margin)
  #r_relapse <- rep(0.0242348, 1000) #base case RADAR relapse (worst case
scenario)
  #monthly mortality (adjusted from annual, predefined and update each year. *
non-PSA)
  alpha <- log(r_relapse[1]/0.00855)
  r_morta <- rep(0,1000)#morality matrix


#compliance probability
r_compli <- 1
  #monthly risk of Diebetes
  r_DMII <- rep(0.01, 1000)
  #criminal justice contact risk (monthly, adjusted from annual)
  r_CJS <- rep(0.035,1000)
  #chance of a court charge after criminal justice contact
  r_court <- rep(0.203,1000)
  #chance of imprisonment after court charge
  r_prison <- rep(0.394,1000) #base case Lin 2014
  #chance of being discharged from hospitalisation within a month (1 cycle)
  r_recover <- rep(0.8,1000) #recover rates (predefined, vector)

  ############################  Utilities & disutilities
###################################

#Stable schizophrenia
  u_stable <- 0.799
  #Serious relapse schizophrenia (hospitalisation)
  u_relapse <- 0.27
  #disutility with CHD (all absolute value, positive)
  u_SF <- 0
  #social fuction improvement (purported, assumption)
  du_UA <- 0.216
  du_SA <- 0.216
  du_MI <- 0.072
  du_CAS <- 0.072
  du_unclassCHD <- 0.101
  #disutility with CVA
  du_TIA <- 0.088
  du_HS <- 0.185
  du_IS <- 0.185
  du_unclassCVA <- 0.153
  #disutility with DMII
  du_DMII <- 0.151
  #disutility with EPS
  du_EPS <- 0.074

  disu_EPS <- 0
  disu_DMII <- 0
  disu_CVD <- 0

  ############################ Cost components ############################
#For PSA, all cost take 70% of the mean as SE for Gamma distribution
#CVD event cost (PRIMROSE economic model)
##CHD
c_UA <- 220
  ci_UA <- 566
  c_SA <- 220
  ci_SA <- 220
  ci_MI <- 5720
  c_MI <- 220
  ci_CAS <- 6008
  ci_unclassCHD <- 2169
  c_unclassCHD <- 220
```

```
  ##CVA
  ci_TIA <- 1368
  c_TIA <- 340
  ci_HS <- 10347
  c_HS <- 2782
  ci_IS <- 10347
  c_IS <- 2782
  ci_unclassCVA <- 5858
  c_unclassCVA <- 1561
  #DMII treatment cost (UKPDS cost)
  c_DMII <- 199
  #EPS treatment cost (Procyclidine, 5mg/day for 3 months + 1 psychiatist
visit)
  c_EPS <- 177
  #Relapse cost (acute hospital)
  ##Acute hospital (per day)
  c_hospital <-259
  ##Olanzapine (included in DATA)
  #Service use cost ()
  ##Outpatient psychiatric visits
  c_outpatpsyvist <- 283.97
  ##Outpatient other visits
  c_outpatothvist <- 119.84
  ##Day hospital visits
  c_Dayhospvist <- 119.84
  ## Community mental health centre visits
  c_Commhcvist  <- 133
  ## Day care centre visits
  c_dayccvist <-54
  ## Specialist ed54ucation
  c_spedu <-54
  ## Psychiatrist visits
  c_psycvit <-300
  ## CPN visits
  c_cpnvist <- 26
  ## GP visits <-
  c_gpvist <- 66
  #AP cost ()
  #Societal cost (criminal justice)
  c_prison <- 3518
  c_court <- 500
  c_CJS <- 540



  aprelapse <- read.csv(file = "AP_nice_para.csv", header = T)


  AP.ab <- as.vector(aprelapse$AP)
  popAP1<- as.vector(pop$AP1.abbr)
  r_EPS_ola <- grep(AP.ab[1], popAP1, value = FALSE)
  r_EPS_ami <- grep(AP.ab[2], popAP1, value = FALSE)
  r_EPS_zot <- grep(AP.ab[3], popAP1, value = FALSE)
  r_EPS_ari <- grep(AP.ab[4], popAP1, value = FALSE)
  r_EPS_pal <- grep(AP.ab[5], popAP1, value = FALSE)
  r_EPS_ris <- grep(AP.ab[6], popAP1, value = FALSE)
  r_EPS_hal <- grep(AP.ab[7], popAP1, value = FALSE)
  r_EPS[r_EPS_ola] <- aprelapse$monthly[1]
  r_EPS[r_EPS_ami] <- aprelapse$monthly[2]
  r_EPS[r_EPS_zot] <- aprelapse$monthly[3]
  r_EPS[r_EPS_ari] <- aprelapse$monthly[4]
  r_EPS[r_EPS_pal] <- aprelapse$monthly[5]
  r_EPS[r_EPS_ris] <- aprelapse$monthly[6]
  r_EPS[r_EPS_hal] <- aprelapse$monthly[7]
  r_lai <- grep("-D", popAP1, value = FALSE)
  r_EPS[r_lai] <- aprelapse$monthly[9]

  #Assuming other APs has same relapse risk to Paliperidone
```

```
  for (i in 1:1000){
    if (r_EPS[i] == 0){
      r_EPS[i] <- aprelapse$monthly[4]
    }
  }

  APstatus <- data.frame(pop$Latest_atypical_drug, pop$Latest_typical_drug)


  ##############################################################################
  ##############################################################################
  ##########  10 years Simulated Patient-Level Markov Model  Maintenance v 3.0
  ##############################################################################
  ##############################################################################

  set.seed(135)
  for (i in 1:1000) { #individual simulation from data

    for (j in 2:nCycle){

      #determining patient path in the model
      #algorithm for transition probability matrix
      r_CVA <- 0.000181546271 *1.151 * ((j-
1)/12)^(0.151)*(exp(0.072*pop$Latest_age-
0.004*pop$Sex_recoded+0.011*pop$Latest_sbp+0.296*pop$Latest_antiht+

0.043*pop$Lastest_tc-1.108*pop$Latest_hdl+1.241*pop$Latest_lipid_drug-
0.015*pop$Latest_weight-

0.399*pop$Latest_height+0.171*pop$Latest_diabetes-
0.021*pop$Ex_Smoker+0.176*pop$Smoker+0.123*pop$Latest_any_antidepressant+

0.380*pop$Latest_Hx_of_heavy_drinking-
0.051+0.270*pop$Latest_typical_drug+0.129*pop$Latest_atypical_drug))

      ### CVA, TIA, haemorrhagic stroke, ischaemic/unclassified stroke and
unspecified CVA

      r_CHD <- 0.00092658942 * 1.058* ((j-1)/12) ^(0.058) *
(exp(0.05*pop$Latest_age-
0.764*pop$Sex_recoded+0.006*pop$Latest_sbp+0.629*pop$Latest_antiht+

0.282*pop$Lastest_tc-0.771*pop$Latest_hdl+1.287*pop$Latest_lipid_drug-
0.002*pop$Latest_weight-

1.59*pop$Latest_height+0.492*pop$Latest_diabetes+0.113*pop$Ex_Smoker+0.416*pop
$Smoker+0.194*pop$Latest_any_antidepressant+

0.327*pop$Latest_Hx_of_heavy_drinking+0.006+0.045*pop$Latest_typical_drug-
0.303*pop$Latest_atypical_drug))
      #r_CHD <- 1-(1-r_CHD)^(1/12)
      if (Stateno[j,i]==1) { #if patient is stable
        if(pop$Sex_men[i]==1){ #calculating mortality based on sex and gender,
by the CVD death mutiplyer
          r_morta[i] <- lifetb[pop$Latest_age[i],(pop$Sex_recoded[i]+2)]*2.5
#mortality rate from the life table by age&sex
        }
        else if(pop$Sex_men[i]==0){ #calculating mortality
          r_morta[i] <- lifetb[pop$Latest_age[i],(pop$Sex_recoded[i]+2)]*1.78
        }
        i_compli[i]<-rbinom(1,1,r_compli)
        trans_array[j,,i] <- c(1-r_relapse[i]*exp(1-i_compli[i])-r_morta[i],
r_relapse[i]*exp(1-i_compli[i]), r_morta[i])
        #CVD risk already predefined in the primrose model
        if (pop$latest_CHD[i] + pop$latest_CVA[i] == 0){
          state_CVD[i,j] <- sample(CVDstates, size = 1, replace = TRUE, prob =
c(1-r_CHD[i]-r_CVA[i], r_CHD[i], r_CVA[i]))
        }
        else {state_CVD[i,j] <- state_CVD[i, j-1]}
        if(pop$Latest_diabetes[i] == 1){
```

```
          state_DMII[i]<-1
        }
        else if(pop$Latest_diabetes[i] == 0){
          state_DMII[i] <- rbinom(1,1,r_DMII[i])
        }

        state_EPS[i] <- rbinom(1,1,r_EPS[i])
        state_CJS[i] <- rbinom(1,1,r_CJS[i])

        #calculate the utility and cost

        utility[i,j] <- u_stable/12 #schizophrenia stable utility
        if (state_CVD[i,j] == 1) {
          pop$latest_CHD[i] <- 1 #marking in baseline data CHD
          if (state_CHD[i,j-1] == 0){ #check whether this is initiation of
events

            state_CHD[i,j] <- sample(CHDstates, size=1, replace = T, prob =
c(r_UA, r_SA, r_MI, r_CAS, r_unclassCHD))
          }
          else {
            state_CHD[i,j] <- state_CHD[i,j-1]
          }
          if(state_CHD[i,j] == 1){
            disu_CVD <- du_UA/12

            if(state_CHD[i,j-1]==0){
              servicecost_month[i,j] <- servicecost_month[i,j] + ci_UA #if
initiation, treatment cost applies
            }
            else {servicecost_month[i,j]<- servicecost_month[i,j] + c_UA/12}
          }
          if(state_CHD[i,j] == 2){
              disu_CVD <- du_SA/12
              if(state_CHD[i,j-1]==0){
                servicecost_month[i,j] <- servicecost_month[i,j] + ci_SA #if
initiation, treatment cost applies
              }
              else {servicecost_month[i,j]<- servicecost_month[i,j] + c_SA/12}
          }
          if(state_CHD[i,j] == 3){
            disu_CVD <- du_MI/12
            if(state_CHD[i,j-1]==0){
              servicecost_month[i,j] <- servicecost_month[i,j] + ci_MI #if
initiation, treatment cost applies
            }
            else {servicecost_month[i,j]<- servicecost_month[i,j] + c_MI/12}
          }
          if(state_CHD[i,j] == 4){
            disu_CVD <- du_CAS/12
            if(state_CHD[i,j-1]==0){
              servicecost_month[i,j] <- servicecost_month[i,j] + ci_CAS #if
initiation, treatment cost applies
            }
            else {servicecost_month[i,j]<- servicecost_month[i,j]}
          }
          if(state_CHD[i,j] == 5){
            disu_CVD <- du_unclassCHD/12
            if(state_CHD[i,j-1]==0){
              servicecost_month[i,j] <- servicecost_month[i,j] + ci_unclassCHD
#if initiation, treatment cost applies
            }
            else {servicecost_month[i,j]<- servicecost_month[i,j] +
c_unclassCHD/12}
          }
        }
        else if (state_CVD[i,j] == 2){
          pop$latest_CVA [i] <- 1
          if (state_CVA[i,j-1] == 0){ #check whether this is initiation of
events
```

```
            state_CVA[i,j] <- sample(CVAstates, 1, replace=T, prob = c(r_TIA,
r_HS, r_IS, r_unclassCVA))
          }
          else {
            state_CVA[i,j] <- state_CVA[i,j-1]
          }

          if(state_CVA[i,j] == 1){
            disu_CVD<-du_TIA/12
            if(state_CVA[i,j-1]==0){
              servicecost_month[i,j] <- servicecost_month[i,j] + ci_TIA #if
initiation, treatment cost applies
            }
            else {servicecost_month[i,j]<- servicecost_month[i,j] + c_TIA/12}
          }
          if(state_CVA[i,j] == 2){
            disu_CVD<-du_HS/12
            if(state_CVA[i,j-1]==0){
              servicecost_month[i,j] <- servicecost_month[i,j] + ci_HS #if
initiation, treatment cost applies
            }
            else {servicecost_month[i,j]<- servicecost_month[i,j] + c_HS/12}
          }
          if(state_CVA[i,j] == 3){
            disu_CVD<-du_IS/12
            if(state_CVA[i,j-1]==0){
              servicecost_month[i,j] <- servicecost_month[i,j] + ci_IS #if
initiation, treatment cost applies
            }
            else {servicecost_month[i,j]<- servicecost_month[i,j] + c_IS/12}
          }
          if(state_CVA[i,j] == 4){
            disu_CVD<-du_unclassCVA/12
            if(state_CVA[i,j-1]==0){
              servicecost_month[i,j] <- servicecost_month[i,j] + ci_unclassCVA
#if initiation, treatment cost applies
            }
            else {servicecost_month[i,j]<- servicecost_month[i,j] +
c_unclassCVA/12}
          }
        }
        if (state_DMII[i] == 1) {
          pop$Latest_diabetes[i]<-1
          servicecost_month[i,j]<- servicecost_month[i,j]+c_DMII/12
          disu_DMII <- du_DMII/12
        }
        if (state_EPS[i] == 1){
          disu_EPS <- du_EPS/12
          servicecost_month[i,j]<- servicecost_month[i,j]+c_EPS/12
        }

        utility[i,j]<- utility[i,j] - max(disu_CVD,disu_DMII,disu_EPS)
#(absolute value here)

        if (state_prison[i]==1){ #registering whether patient's in Prison
          if (j == rel_prison[i]){
            state_prison[i] <- 0
          }
          else {state_prison[i] <- 1}
        }
        else if (state_prison[i]==0){
          if (state_CJS[i] == 1) {
            societalcost_month[i,j] <- societalcost_month[i,j] + c_CJS
            state_court[i]<-rbinom(1,1, r_court[i])
            if (state_court[i] == 1) {
              societalcost_month[i,j] <- societalcost_month[i,j] + c_court
              state_prison[i] <- rbinom(1,1, r_prison[i])
              if (state_prison[i] ==1){
                cycle_prison[i] <- round(rgamma(n = 1,shape = 3, scale = 16))
#asuumption average 4 years length sentences #UK Prison Population Statistics
```

```
                rel_prison[i] <- j+cycle_prison[i]
                societalcost_month[i,j] <- societalcost_month[i,j] +
c_prison*cycle_prison[i]
              }
            }
          }
        }#print(utility[i,j])
        servicecost_month[i,j] <-
servicecost_month[i,j]+(pop$Annual.cost.1[i]+pop$Annual.cost.2[i])/12 +

(c_outpatpsyvist/6*1.4+c_outpatothvist/6*0.1+c_Dayhospvist/6*2.3+c_Commhcvist/
6*2.4+c_dayccvist/6*5.9+c_spedu/6*2.9+ c_cpnvist/6*12.6 + c_psycvit/6*2.5
+c_gpvist/6*1.8)
      }


      else if (Stateno[j,i]==2) {
        if(pop$Sex_men[i]==1){
          r_morta[i] <- lifetb[pop$Latest_age[i],(pop$Sex_recoded[i]+2)]*6.91
#mortality rate from the life table by age&sex
        }
        else if(pop$Sex_men[i]==0){
          r_morta[i] <- lifetb[pop$Latest_age[i],(pop$Sex_recoded[i]+2)]*7.9
        }
        trans_array[j,,i] <- c(r_recover[i], 1-r_recover[i]-r_morta[i],
r_morta[i])
        if (1-r_recover[i]-r_morta[i] <0) {trans_array[j,2,i] <- 0}  ####
check if the remaining probabilty smaller than 0
        #criminal justice

        if (pop$latest_CHD[i] + pop$latest_CVA[i] == 0){

        state_CVD[i,j] <- sample(CVDstates, size = 1, replace = T, prob = c(1-
r_CHD[i]-r_CVA[i], r_CHD[i], r_CVA[i]))
        }
        else {state_CVD[i,j] <-state_CVD[i, j-1]}

        if(pop$Latest_diabetes[i] == 1){
          state_DMII[i]<-1
        }
        else if(pop$Latest_diabetes[i] == 0){
          state_DMII[i] <- rbinom(1,1,r_DMII[i])
        }

        state_EPS[i] <- rbinom(1,1,r_EPS[i])
        state_CJS[i] <- rbinom(1,1,(r_CJS[i]* 1.53))

        #calculate the utility (how to get a composite utility for
comorbidities?)

        utility[i,j] <- u_relapse/12 #schizophrenia relapse utility
        if (state_CVD[i,j] == 1) {
          pop$latest_CHD[i] <- 1
          if (state_CHD[i,j-1] == 0){ #check whether this is initiation of
events
            state_CHD[i,j] <- sample(CHDstates, 1, replace = T, prob = c(r_UA,
r_SA, r_MI, r_CAS, r_unclassCHD))
          }
          else {
            state_CHD[i,j] <- state_CHD[i,j-1]
          }
          if(state_CHD[i,j] == 1){
            disu_CVD <- du_UA/12

            if(state_CHD[i,j-1]==0){
              servicecost_month[i,j] <- servicecost_month[i,j] + ci_UA #if
initiation, treatment cost applies
            }
            else {servicecost_month[i,j]<- servicecost_month[i,j] + c_UA/12}
          }
```

```
          if(state_CHD[i,j] == 2){
            disu_CVD <- du_SA/12
            if(state_CHD[i,j-1]==0){
              servicecost_month[i,j] <- servicecost_month[i,j] + ci_SA #if
initiation, treatment cost applies
            }
            else {servicecost_month[i,j]<- servicecost_month[i,j] + c_SA/12}
          }
          if(state_CHD[i,j] == 3){
            disu_CVD <- du_MI/12
            if(state_CHD[i,j-1]==0){
              servicecost_month[i,j] <- servicecost_month[i,j] + ci_MI #if
initiation, treatment cost applies
            }
            else {servicecost_month[i,j]<- servicecost_month[i,j] + c_MI/12}
          }
          if(state_CHD[i,j] == 4){
            disu_CVD <- du_CAS/12
            if(state_CHD[i,j-1]==0){
              servicecost_month[i,j] <- servicecost_month[i,j] + ci_CAS #if
initiation, treatment cost applies
            }
            else {servicecost_month[i,j]<- servicecost_month[i,j]}
          }
          if(state_CHD[i,j] == 5){
            disu_CVD <- du_unclassCHD/12
            if(state_CHD[i,j-1]==0){
              servicecost_month[i,j] <- servicecost_month[i,j] + ci_unclassCHD
#if initiation, treatment cost applies
            }
            else {servicecost_month[i,j]<- servicecost_month[i,j] +
c_unclassCHD/12}
          }
        }
        else if (state_CVD[i,j] == 2){
          pop$latest_CVA [i] <- 1
          if (state_CVA[i,j-1] == 0){ #check whether this is initiation of
events
            state_CVA[i,j] <- sample(CVAstates, 1, replace=T, prob = c(r_TIA,
r_HS, r_IS, r_unclassCVA))
          }
          else {
            state_CVA[i,j] <- state_CVA[i,j-1]
          }

          if(state_CVA[i,j] == 1){
            disu_CVD<-du_TIA/12
            if(state_CVA[i,j-1]==0){
              servicecost_month[i,j] <- servicecost_month[i,j] + ci_TIA #if
initiation, treatment cost applies
            }
            else {servicecost_month[i,j]<- servicecost_month[i,j] + c_TIA/12}
          }
          if(state_CVA[i,j] == 2){
            disu_CVD<-du_HS/12
            if(state_CVA[i,j-1]==0){
              servicecost_month[i,j] <- servicecost_month[i,j] + ci_HS #if
initiation, treatment cost applies
            }
            else {servicecost_month[i,j]<- servicecost_month[i,j] + c_HS/12}
          }
          if(state_CVA[i,j] == 3){
            disu_CVD<-du_IS/12
            if(state_CVA[i,j-1]==0){
              servicecost_month[i,j] <- servicecost_month[i,j] + ci_IS #if
initiation, treatment cost applies
            }
            else {servicecost_month[i,j]<- servicecost_month[i,j] + c_IS/12}
          }
          if(state_CVA[i,j] == 4){
```

```
            disu_CVD<-du_unclassCVA/12
            if(state_CVA[i,j-1]==0){
              servicecost_month[i,j] <- servicecost_month[i,j] + ci_unclassCVA
#if initiation, treatment cost applies
            }
            else {servicecost_month[i,j]<- servicecost_month[i,j] +
c_unclassCVA/12}
          }
        }
        if (state_DMII[i] == 1) {
          pop$Latest_diabetes[i]<-1
          servicecost_month[i,j]<- servicecost_month[i,j]+c_DMII/12
          disu_DMII <- du_DMII/12
        }
        if (state_EPS[i] ==1){
          disu_EPS <- du_EPS/12
          servicecost_month[i,j]<- servicecost_month[i,j]+c_EPS/12
        }

        utility[i,j]<- utility[i,j] + min(disu_CVD,disu_DMII,disu_EPS)

        if (state_prison[i]==1){ #registering whether patient's in Prison
          if (j == rel_prison[i]){
            state_prison[i] <- 0
          }
          else {state_prison[i] <- 1}
        }
        else if (state_prison[i]==0){
          if (state_CJS[i] == 1) {
            societalcost_month[i,j] <- societalcost_month[i,j] + c_CJS
            state_court[i]<-rbinom(1,1, r_court[i])
            if (state_court[i] == 1) {
              societalcost_month[i,j] <- societalcost_month[i,j] + c_court
              state_prison[i] <- rbinom(1,1, r_prison[i])
              if (state_prison[i] ==1){
                cycle_prison[i] <- round(rgamma(n = 1,shape = 3, scale = 16))
#asuumption average 4 years length sentences #UK Prison Population Statistics
                rel_prison[i] <- j+cycle_prison[i]
                societalcost_month[i,j] <- societalcost_month[i,j] +
c_prison*cycle_prison[i]
              }
            }
          }
        }#print(utility[i,j])
        servicecost_month[i,j] <-
servicecost_month[i,j]+(pop$Annual.cost.1[i]+pop$Annual.cost.2[i]+as.numeric(A
Plist$Annual.cost....[10]))/12 +

(c_outpatpsyvist/6*2.1+c_outpatothvist/6*0.3+c_Dayhospvist/6*2.1+c_Commhcvist/
6*1.4+c_dayccvist/6*0.9+c_psycvit/6*2.3+ c_cpnvist/6*5.2 + c_gpvist/6*1.6) +
c_hospital *30
        relapsecost_month[i,j] <-
(pop$Annual.cost.1[i]+pop$Annual.cost.2[i]+as.numeric(APlist$Annual.cost....[1
0]))/12 +

(c_outpatpsyvist/6*2.1+c_outpatothvist/6*0.3+c_Dayhospvist/6*2.1+c_Commhcvist/
6*1.4+c_dayccvist/6*0.9+c_psycvit/6*2.3+ c_cpnvist/6*5.2 + c_gpvist/6*1.6) +
c_hospital *30
      }
      else if (Stateno[j,i]==3) {utility[i,j] <- 0 ; break}

      #print(utility[i,j])
      #sampling the next health states from trans vector
      #print(trans_array[j,,i])
      if (j >= nCycle) break
      set.seed(135)
      Stateno[j+1,i] <- sample(nStates, size = 1, replace = TRUE, prob =
trans_array[j,,i])
      #print(Stateno[j+1,i])
```

```
      #creating parameters of co-morbidity -- a different set of probabilities
from which it is sample every circle
      if (j %% 12 == 0) {
        pop$Latest_age[i] <- pop$Latest_age[i] + 1 #every calendar year,
calculating the yearly cost and utility
        societalcost[,j/12]<-(rowSums(societalcost_month[,c((j-
11):j)]))/(1+discount)^(j/12)
        servicecost[,j/12]<-(rowSums(servicecost_month[,c((j-
11):j)]))/(1+discount)^(j/12)
        QALY[ ,j/12] <- rowSums(utility[ ,c((j-11):j)])/(1+discount)^(j/12)
        #cost_maintenance <- societalcost + servicecost
        print(c(i,j))#print(QALY)
      }
    }#treatment allocation --> antipsychotic mantainance

    #print(i)
    write.csv(Stateno, file = '~/Documents/Markov
Tutorial/results/Transtrajec.csv', row.names = T)
    write.csv(servicecost, file = '~/Documents/Markov
Tutorial/results/servicecost.csv', row.names = T)
    write.csv(societalcost, file = '~/Documents/Markov
Tutorial/results/societalcost.csv', row.names = T)
    write.csv(QALY, file ='~/Documents/Markov Tutorial/results/QALY.csv',
row.names = T)
    #}
  }
  totalcost <- servicecost+societalcost
  cost_10yr_M<- rowSums(totalcost)
  QALY_10yr_M<- rowSums(QALY)



  ##############################################################################
  ##############################################################################
  ################ 10 YEARS  Simulated Patient-Level Markov Model -  RADAR
  ##############################################################################
  ##############################################################################

  c_radar<-c(0.2,0.3,0.4,0.5,0.75,0.8,1,0)
  set.seed(122)
  for (i in 1:1000) { #individual simulation from data

    for (j in 2:nCycle){
      if (j == 2){
      r_radar[i,2] <- 1 #r_radar, radar indicator 1 full maintanance, 0
discontinuation
      }
      r_CVA <- 1-(1-(0.000181546271 *1.151 * ((j-
1)/12)^(0.151)*(exp(0.072*pop$Latest_age-
0.004*pop$Sex_recoded+0.011*pop$Latest_sbp+0.296*pop$Latest_antiht+

0.043*pop$Lastest_tc-1.108*pop$Latest_hdl+1.241*pop$Latest_lipid_drug-
0.015*pop$Latest_weight-

0.399*pop$Latest_height+0.171*pop$Latest_diabetes-
0.021*pop$Ex_Smoker+0.176*pop$Smoker+0.123*pop$Latest_any_antidepressant+

0.380*pop$Latest_Hx_of_heavy_drinking-
0.051+0.270*pop$Latest_typical_drug+0.129*pop$Latest_atypical_drug))))^(1/12)
      #r_CVA <- 1-(1-r_CVA)^(1/12)
      ### CVA, TIA, haemorrhagic stroke, ischaemic/unclassified stroke and
unspecified CVA

      r_CHD <- 1-(1-(0.00092658942 * 1.058* ((j-1)/12) ^(0.058) *
(exp(0.05*pop$Latest_age-
0.764*pop$Sex_recoded+0.006*pop$Latest_sbp+0.629*pop$Latest_antiht+

0.282*pop$Lastest_tc-0.771*pop$Latest_hdl+1.287*pop$Latest_lipid_drug-
0.002*pop$Latest_weight-
```

```
1.59*pop$Latest_height+0.492*pop$Latest_diabetes+0.113*pop$Ex_Smoker+0.416*pop
$Smoker+0.194*pop$Latest_any_antidepressant+

0.327*pop$Latest_Hx_of_heavy_drinking+0.006+0.045*pop$Latest_typical_drug-
0.303*pop$Latest_atypical_drug))))^(1/12)
      #r_CHD <- 1-(1-r_CHD)^(1/12)
      if (Stateno[j,i]==1) {
        if(pop$Sex_men[i]==1){
          r_morta[i] <- lifetb[pop$Latest_age[i],(pop$Sex_recoded[i]+2)]*2.5
#mortality rate from the life table by age&sex
        }
        else if(pop$Sex_men[i]==0){
          r_morta[i] <- lifetb[pop$Latest_age[i],(pop$Sex_recoded[i]+2)]*1.78
        }
        trans_array[j,,i] <- c(1-r_relapse[i]*exp(alpha*(1-r_radar[i,j]))-
r_morta[i], r_relapse[i]*exp(alpha*(1-r_radar[i,j])), r_morta[i])  #exp
        #trans_array[j,,i] <- c(1-(r_relapse[i]-0.0156*r_radar[i,j])-
r_morta[i], r_relapse[i]-0.0156*r_radar[i,j], r_morta[i]) #linear
        #trans_array[j,,i] <- c(1-(r_relapse[i]-0.0156*sqrt(r_radar[i,j]))-
r_morta[i], r_relapse[i]-0.0156*sqrt(r_radar[i,j]), r_morta[i]) #sqrt
        #trans_array[j,,i] <- c(1-(r_relapse[i]-0.0156*r_radar[i,j]^2)-
r_morta[i], r_relapse[i]-0.0156*r_radar[i,j]^2, r_morta[i]) #quadratic
        #trans_array[j,,i] <- c(1-(r_relapse[i]-0.0156*r_radar[i,j]^3)-
r_morta[i], r_relapse[i]-0.0156*r_radar[i,j]^3, r_morta[i]) #cubed

        #CVD risk already predefined in the primrose model
        if (pop$latest_CHD[i] + pop$latest_CVA[i] == 0){
          state_CVD[i,j] <- sample(CVDstates, size = 1, replace = TRUE, prob =
c(1-r_CHD[i]-r_CVA[i], r_CHD[i], r_CVA[i]))
        }
        else {state_CVD[i,j] <-state_CVD[i, j-1]}
        if(pop$Latest_diabetes[i] == 1){
          state_DMII[i]<-1
        }
        else if(pop$Latest_diabetes[i] == 0){
          if (r_radar[i,j] != 0 ) {state_DMII[i] <- rbinom(1,1,r_DMII[i])}
          else if (r_radar[i,j] == 0) {state_DMII[i] <- 0}
        }

        if (r_radar[i,j] != 0 ){
        state_EPS[i] <- rbinom(1,1,r_EPS[i])}
        else if (r_radar[i,j]==0){
          state_EPS[i] <- 0
        }

        state_CJS[i] <- rbinom(1,1,r_CJS[i])

        #calculate the utility and cost

        utility[i,j] <- u_stable/12 #schizophrenia stable utility
        if (state_CVD[i,j] == 1) {#CHD events
          pop$latest_CHD[i] <- 1 #marking in baseline data CHD
          if (state_CHD[i,j-1] == 0){ #check whether this is initiation of
events
            state_CHD[i,j] <- sample(CHDstates, size=1, replace = T, prob =
c(r_UA, r_SA, r_MI, r_CAS, r_unclassCHD))
          }
          else {
            state_CHD[i,j] <- state_CHD[i,j-1]
          }
          if(state_CHD[i,j] == 1){
            disu_CVD <- du_UA/12

            if(state_CHD[i,j-1]==0){
              servicecost_month[i,j] <- servicecost_month[i,j] + ci_UA #if
initiation, treatment cost applies
              CVDcost_month[i,j] <- CVDcost_month[i,j]+ci_UA
            }
            else {servicecost_month[i,j]<- servicecost_month[i,j] + c_UA/12
```

```
            CVDcost_month[i,j] <- CVDcost_month[i,j]+c_UA/12}
          }
          if(state_CHD[i,j] == 2){
              disu_CVD <- du_SA/12
              if(state_CHD[i,j-1]==0){
                servicecost_month[i,j] <- servicecost_month[i,j] + ci_SA #if
initiation, treatment cost applies
                CVDcost_month[i,j] <- CVDcost_month[i,j] + ci_SA
                            }
              else {servicecost_month[i,j]<- servicecost_month[i,j] + c_SA/12
              CVDcost_month[i,j] <- CVDcost_month[i,j]+c_SA/12}
          }
          if(state_CHD[i,j] == 3){
            disu_CVD <- du_MI/12
            if(state_CHD[i,j-1]==0){
              servicecost_month[i,j] <- servicecost_month[i,j] + ci_MI #if
initiation, treatment cost applies
            }
            else {servicecost_month[i,j]<- servicecost_month[i,j] + c_MI/12}
          }
          if(state_CHD[i,j] == 4){
            disu_CVD <- du_CAS/12
            if(state_CHD[i,j-1]==0){
              servicecost_month[i,j] <- servicecost_month[i,j] + ci_CAS #if
initiation, treatment cost applies
              CVDcost_month[i,j] <- CVDcost_month[i,j] + ci_CAS
            }
            else {servicecost_month[i,j]<- servicecost_month[i,j]}
          }
          if(state_CHD[i,j] == 5){
            disu_CVD <- du_unclassCHD/12
            if(state_CHD[i,j-1]==0){
              servicecost_month[i,j] <- servicecost_month[i,j] + ci_unclassCHD
#if initiation, treatment cost applies
              CVDcost_month[i,j] <- CVDcost_month[i,j] + ci_unclassCHD
            }
            else {servicecost_month[i,j]<- servicecost_month[i,j] +
c_unclassCHD/12
            CVDcost_month[i,j] <- CVDcost_month[i,j] +c_unclassCHD/12}
          }
        }
        else if (state_CVD[i,j] == 2){
          pop$latest_CVA [i] <- 1
          if (state_CVA[i,j-1] == 0){ #check whether this is initiation of
events
            state_CVA[i,j] <- sample(CVAstates, 1, replace=T, prob = c(r_TIA,
r_HS, r_IS, r_unclassCVA))
          }
          else {
            state_CVA[i,j] <- state_CVA[i,j-1]
          }

          if(state_CVA[i,j] == 1){
            disu_CVD<-du_TIA/12
            if(state_CVA[i,j-1]==0){
              servicecost_month[i,j] <- servicecost_month[i,j] + ci_TIA #if
initiation, treatment cost applies
              CVDcost_month[i,j] <- CVDcost_month[i,j] +ci_TIA
            }
            else {servicecost_month[i,j]<- servicecost_month[i,j] + c_TIA/12
            CVDcost_month[i,j] <- CVDcost_month[i,j] + c_TIA/12}
          }
          if(state_CVA[i,j] == 2){
            disu_CVD<-du_HS/12
            if(state_CVA[i,j-1]==0){
              servicecost_month[i,j] <- servicecost_month[i,j] + ci_HS #if
initiation, treatment cost applies
              CVDcost_month[i,j] <- CVDcost_month[i,j] +ci_HS
            }
            else {servicecost_month[i,j]<- servicecost_month[i,j] + c_HS/12
```

```
            CVDcost_month[i,j] <- CVDcost_month[i,j] + c_HS/12}
          }
          if(state_CVA[i,j] == 3){
            disu_CVD<-du_IS/12
            if(state_CVA[i,j-1]==0){
              servicecost_month[i,j] <- servicecost_month[i,j] + ci_IS #if
initiation, treatment cost applies
              CVDcost_month[i,j] <- CVDcost_month[i,j] +ci_IS
            }
            else {servicecost_month[i,j]<- servicecost_month[i,j] + c_IS/12
            CVDcost_month[i,j] <- CVDcost_month[i,j] + ci_IS/12}
          }
          if(state_CVA[i,j] == 4){
            disu_CVD<-du_unclassCVA/12
            if(state_CVA[i,j-1]==0){
              servicecost_month[i,j] <- servicecost_month[i,j] + ci_unclassCVA
#if initiation, treatment cost applies
              CVDcost_month[i,j] <- CVDcost_month[i,j] + ci_unclassCVA
            }
            else {servicecost_month[i,j]<- servicecost_month[i,j] +
c_unclassCVA/12
            CVDcost_month[i,j] <- CVDcost_month[i,j] + c_unclassCVA/12}
          }
        }
        if (state_DMII[i] == 1) {
          pop$Latest_diabetes[i]<-1
          servicecost_month[i,j]<- servicecost_month[i,j]+c_DMII/12
          disu_DMII <- du_DMII/12
        }
        if (state_EPS[i] == 1){
          disu_EPS <- du_EPS/12
          servicecost_month[i,j]<- servicecost_month[i,j]+c_EPS/12
        }

        if (r_radar[i,j]!=0){
        utility[i,j]<- utility[i,j] - max(disu_CVD,disu_DMII,disu_EPS)
#(absolute value here)
      }
        else if (r_radar[i,j]==0){
          utility[i,j]<- min((utility[i,j] - max(disu_CVD,disu_DMII,disu_EPS)
+ u_SF/12),1) #(RADAR social function improvement)
        }

        if (state_prison[i]==1){ #registering whether patient's in Prison
          if (j == rel_prison[i]){
            state_prison[i] <- 0
          }
          else {state_prison[i] <- 1}
        }
        else if (state_prison[i]==0){
          if (state_CJS[i] == 1) {
            societalcost_month[i,j] <- societalcost_month[i,j] + c_CJS
            state_court[i]<-rbinom(1,1, r_court[i])
            if (state_court[i] == 1) {
              societalcost_month[i,j] <- societalcost_month[i,j] + c_court
              state_prison[i] <- rbinom(1,1, r_prison[i])
              if (state_prison[i] ==1){
                cycle_prison[i] <- round(rgamma(n = 1,shape = 3, scale = 16))
#asuumption average 4 years length sentences #UK Prison Population Statistics
                rel_prison[i] <- j+cycle_prison[i]
                societalcost_month[i,j] <- societalcost_month[i,j] +
c_prison*cycle_prison[i]
              }
            }
          }
        }#print(utility[i,j])
        servicecost_month[i,j] <-
servicecost_month[i,j]+(pop$Annual.cost.1[i]+pop$Annual.cost.2[i])/12 +

(c_outpatpsyvist/6*1.4+c_outpatothvist/6*0.1+c_Dayhospvist/6*2.3+c_Commhcvist/
```

```
6*2.4+c_dayccvist/6*5.9+c_spedu/6*2.9+ c_cpnvist/6*12.6 + c_psycvit/6*2.5
+c_gpvist/6*1.8)
        if (r_radar[i,j] == sort(c_radar,partial=length(c_radar)-
1)[2]){#second last
          r_radar[i,j+1] <- sample(c_radar[c_radar<=r_radar[i,j]],size = 1)
        }
        else if (r_radar[i,j] == 0) {
          r_radar[i, j+1] <- 0
        }
        else {
          r_radar[i,j+1] <- sample(c_radar[c_radar<=r_radar[i,j]],size = 1,
prob = c_radar[c_radar<=r_radar[i,j]])
          }
      }

      else if (Stateno[j,i]==2) {
        if(pop$Sex_men[i]==1){
          r_morta[i] <- lifetb[pop$Latest_age[i],(pop$Sex_recoded[i]+2)]*6.91
#mortality rate from the life table by age&sex
        }
        else if(pop$Sex_men[i]==0){
          r_morta[i] <- lifetb[pop$Latest_age[i],(pop$Sex_recoded[i]+2)]*7.9
        }
        trans_array[j,,i] <- c(r_recover[i], 1-r_recover[i]-r_morta[i],
r_morta[i])
        if (1-r_recover[i]-r_morta[i] <0) {trans_array[j,2,i] <- 0}  ####
check if the remaining probabilty smaller than 0

        if (pop$latest_CHD[i] + pop$latest_CVA[i] == 0){
        state_CVD[i,j] <- sample(CVDstates, size = 1, replace = T, prob = c(1-
r_CHD[i]-r_CVA[i], r_CHD[i], r_CVA[i]))
        }
        else {state_CVD[i,j] <-state_CVD[i, j-1]}

        if(pop$Latest_diabetes[i] == 1){
          state_DMII[i]<-1
        }
        else if(pop$Latest_diabetes[i] == 0){
          if (r_radar[i,j] != 0 ) {state_DMII[i] <- rbinom(1,1,r_DMII[i])}
          else if (r_radar[i,j] == 0) {state_DMII[i] <- 0}
        }

        if (r_radar[i,j] != 0 ){
          state_EPS[i] <- rbinom(1,1,r_EPS[i])}
        else if (r_radar[i,j]==0){
          state_EPS[i] <- 0
        }

        state_CJS[i] <- rbinom(1,1,(1.53 *r_CJS[i]))#base case according to
Lin 2014 (US data) increased 53% when hospitalised

        #calculate the utility (how to get a composite utility for
comorbidities?)

        utility[i,j] <- u_relapse/12 #schizophrenia relapse utility
        if (state_CVD[i,j] == 1) {
          pop$latest_CHD[i] <- 1
          if (state_CHD[i,j-1] == 0){ #check whether this is initiation of
events
            state_CHD[i,j] <- sample(CHDstates, 1, replace = T, prob = c(r_UA,
r_SA, r_MI, r_CAS, r_unclassCHD))
          }
          else {
            state_CHD[i,j] <- state_CHD[i,j-1]
          }
          if(state_CHD[i,j] == 1){
            disu_CVD <- du_UA/12

            if(state_CHD[i,j-1]==0){
```

```
              servicecost_month[i,j] <- servicecost_month[i,j] + ci_UA #if
initiation, treatment cost applies
              CVDcost_month[i,j] <- CVDcost_month[i,j] +ci_UA
            }
            else {servicecost_month[i,j]<- servicecost_month[i,j] + c_UA/12
            CVDcost_month[i,j] <- CVDcost_month[i,j]+ci_UA/12}
          }
          if(state_CHD[i,j] == 2){
            disu_CVD <- du_SA/12
            if(state_CHD[i,j-1]==0){
              servicecost_month[i,j] <- servicecost_month[i,j] + ci_SA #if
initiation, treatment cost applies
              CVDcost_month[i,j] <- CVDcost_month[i,j]+ci_SA
            }
            else {servicecost_month[i,j]<- servicecost_month[i,j] + c_SA/12
            CVDcost_month[i,j] <- CVDcost_month[i,j]+c_SA/12}
          }
          if(state_CHD[i,j] == 3){
            disu_CVD <- du_MI/12
            if(state_CHD[i,j-1]==0){
              servicecost_month[i,j] <- servicecost_month[i,j] + ci_MI #if
initiation, treatment cost applies
              CVDcost_month[i,j] <- CVDcost_month[i,j] +ci_MI
            }
            else {servicecost_month[i,j]<- servicecost_month[i,j] + c_MI/12
            CVDcost_month[i,j] <- CVDcost_month[i,j]+ c_MI/12}
          }
          if(state_CHD[i,j] == 4){
            disu_CVD <- du_CAS/12
            if(state_CHD[i,j-1]==0){
              servicecost_month[i,j] <- servicecost_month[i,j] + ci_CAS #if
initiation, treatment cost applies
              CVDcost_month[i,j] <- CVDcost_month[i,j] +ci_CAS
            }
            else {servicecost_month[i,j]<- servicecost_month[i,j]}
          }
          if(state_CHD[i,j] == 5){
            disu_CVD <- du_unclassCHD/12
            if(state_CHD[i,j-1]==0){
              servicecost_month[i,j] <- servicecost_month[i,j] + ci_unclassCHD
#if initiation, treatment cost applies
              CVDcost_month[i,j] <- CVDcost_month[i,j] + ci_unclassCHD
            }
            else {servicecost_month[i,j]<- servicecost_month[i,j] +
c_unclassCHD/12
            CVDcost_month[i,j] <- CVDcost_month[i,j] +c_unclassCHD/12}
          }
        }
        else if (state_CVD[i,j] == 2){
          pop$latest_CVA [i] <- 1
          if (state_CVA[i,j-1] == 0){ #check whether this is initiation of
events
            state_CVA[i,j] <- sample(CVAstates, 1, replace=T, prob = c(r_TIA,
r_HS, r_IS, r_unclassCVA))
          }
          else {
            state_CVA[i,j] <- state_CVA[i,j-1]
          }

          if(state_CVA[i,j] == 1){
            disu_CVD<-du_TIA/12
            if(state_CVA[i,j-1]==0){
              servicecost_month[i,j] <- servicecost_month[i,j] + ci_TIA #if
initiation, treatment cost applies
              CVDcost_month[i,j] <- CVDcost_month[i,j]+ci_TIA
            }
            else {servicecost_month[i,j]<- servicecost_month[i,j] + c_TIA/12
            CVDcost_month[i,j] <- CVDcost_month[i,j]+c_TIA/12}
          }
          if(state_CVA[i,j] == 2){
```

```
            disu_CVD<-du_HS/12
            if(state_CVA[i,j-1]==0){
              servicecost_month[i,j] <- servicecost_month[i,j] + ci_HS #if
initiation, treatment cost applies
              CVDcost_month[i,j] <- CVDcost_month[i,j] + ci_HS
            }
            else {servicecost_month[i,j]<- servicecost_month[i,j] + c_HS/12
            CVDcost_month[i,j] <- CVDcost_month[i,j] + c_HS/12}
          }
          if(state_CVA[i,j] == 3){
            disu_CVD<-du_IS/12
            if(state_CVA[i,j-1]==0){
              servicecost_month[i,j] <- servicecost_month[i,j] + ci_IS #if
initiation, treatment cost applies
              CVDcost_month[i,j] <- CVDcost_month[i,j]+ci_IS
            }
            else {servicecost_month[i,j]<- servicecost_month[i,j] + c_IS/12
            CVDcost_month[i,j] <- CVDcost_month[i,j]+ c_IS/12}
          }
          if(state_CVA[i,j] == 4){
            disu_CVD<-du_unclassCVA/12
            if(state_CVA[i,j-1]==0){
              servicecost_month[i,j] <- servicecost_month[i,j] + ci_unclassCVA
#if initiation, treatment cost applies
              CVDcost_month[i,j] <- CVDcost_month[i,j] + ci_unclassCVA
            }
            else {servicecost_month[i,j]<- servicecost_month[i,j] +
c_unclassCVA/12
            CVDcost_month[i,j] <- CVDcost_month[i,j]+ c_unclassCVA/12}
          }
        }
        if (state_DMII[i] == 1) {
          pop$Latest_diabetes[i]<-1
          servicecost_month[i,j]<- servicecost_month[i,j]+c_DMII/12
          disu_DMII <- du_DMII/12
        }
        if (state_EPS[i] ==1){
          disu_EPS <- du_EPS/12
          servicecost_month[i,j]<- servicecost_month[i,j]+c_EPS/12
        }

        utility[i,j]<- utility[i,j] + min(disu_CVD,disu_DMII,disu_EPS)

        if (state_prison[i]==1){ #registering whether patient's in Prison
          if (j == rel_prison[i]){
            state_prison[i] <- 0
          }
          else {state_prison[i] <- 1}
        }
        else if (state_prison[i]==0){
          if (state_CJS[i] == 1) {
            societalcost_month[i,j] <- societalcost_month[i,j] + c_CJS
            state_court[i]<-rbinom(1,1, r_court[i])
            if (state_court[i] == 1) {
              societalcost_month[i,j] <- societalcost_month[i,j] + c_court
              state_prison[i] <- rbinom(1,1, r_prison[i])
              if (state_prison[i] ==1){
                cycle_prison[i] <- round(rgamma(n = 1,shape = 3, scale = 16))
#asuumption average 4 years length sentences #UK Prison Population Statistics
                rel_prison[i] <- j+cycle_prison[i]
                societalcost_month[i,j] <- societalcost_month[i,j] +
c_prison*cycle_prison[i]
              }
            }
          }
        }#print(utility[i,j])
        servicecost_month[i,j] <-
servicecost_month[i,j]+(pop$Annual.cost.1[i]+pop$Annual.cost.2[i]+as.numeric(A
Plist$Annual.cost....[10]))/12 +
```

```
(c_outpatpsyvist/6*2.1+c_outpatothvist/6*0.3+c_Dayhospvist/6*2.1+c_Commhcvist/
6*1.4+c_dayccvist/6*0.9+c_psycvit/6*2.3+ c_cpnvist/6*5.2 + c_gpvist/6*1.6) +
c_hospital *30
        relapsecost_month[i,j] <-
(pop$Annual.cost.1[i]+pop$Annual.cost.2[i]+as.numeric(APlist$Annual.cost....[1
0]))/12 +

(c_outpatpsyvist/6*2.1+c_outpatothvist/6*0.3+c_Dayhospvist/6*2.1+c_Commhcvist/
6*1.4+c_dayccvist/6*0.9+c_psycvit/6*2.3+ c_cpnvist/6*5.2 + c_gpvist/6*1.6) +
c_hospital *30
        r_radar[i,j+1] <- sample(c_radar[c_radar>=r_radar[i,j]],size = 1, prob
= c_radar[c_radar>=r_radar[i,j]])
        if (r_radar[i,j+1]==0){
          pop$Latest_typical_drug[i] <- 0
          pop$Latest_atypical_drug[i] <- 0
        }
        else {
          pop$Latest_atypical_drug[i]<-APstatus$pop.Latest_atypical_drug[i]
          pop$Latest_typical_drug[i]<-APstatus$pop.Latest_typical_drug[i]
        }
      }
      else if (Stateno[j,i]==3) {utility[i,j] <- 0 ; break}

      #print(utility[i,j])
      #sampling the next health states from trans vector
      #print(trans_array[j,,i])
      if (j >= nCycle) break
      Stateno[j+1,i] <- sample(nStates, size = 1, replace = TRUE, prob =
trans_array[j,,i])

      #print(Stateno[j+1,i])
      #creating parameters of co-morbidity -- a different set of probabilities
from which it is sample every circle
      if (j %% 12 == 0) {
        pop$Latest_age[i] <- pop$Latest_age[i] + 1 #every calendar year,
calculating the yearly cost and utility
        societalcost[,j/12]<-(rowSums(societalcost_month[,c((j-
11):j)]))/(1+discount)^(j/12)
        servicecost[,j/12]<-(rowSums(servicecost_month[,c((j-
11):j)]))/(1+discount)^(j/12)
        relapsecost[,j/12]<-(rowSums(relapsecost_month[,c((j-
11):j)]))/(1+discount)^(j/12)
        CVDcost[,j/12]<-(rowSums(CVDcost_month[,c((j-
11):j)]))/(1+discount)^(j/12)
        QALY[ ,j/12] <- rowSums(utility[ ,c((j-11):j)])/(1+discount)^(j/12)
        #print(QALY)
      }
      print(r_radar[i,j])
      print(c(i,j))
    }#treatment allocation --> antipsychotic mantainance

    #print(QALY)
    #}
  }
  write.csv(Stateno, file = '~/Documents/Markov
Tutorial/results/TranstrajecRADAR.csv', row.names = T)
  write.csv(servicecost, file = '~/Documents/Markov
Tutorial/results/servicecostRADAR.csv', row.names = T)
  write.csv(societalcost, file = '~/Documents/Markov
Tutorial/results/societalcostRADAR.csv', row.names = T)
  write.csv(QALY, file ='~/Documents/Markov Tutorial/results/QALYRADAR.csv',
row.names = T)
  write.csv(r_radar, file='~/Documents/Markov Tutorial/results/reduction10yr',
row.names = T)

  totalcost <- servicecost+societalcost
  cost_10yr_R<- rowSums(totalcost)
  QALY_10yr_R<- rowSums(QALY)
```

```
#######################EXAMPLE -- Single arm
collection###################################
x <- NULL
evsi <- NULL
evsi_t <- NULL
evsi_n <- NULL
for (n in 1:1000){
for (j in seq(10, 1000, 10)){
for (i in 1:length(r_compli)){ #generating an observational study sample size
of 50 with the compliance probability in the parameter
  x[i] <- rbinom(1, j, r_compli[i])
  }
  model <- gam(BCEApsa$NMB ~ s(x)) #fit spline with the Net benefit and x
  evsi<- mean(pmax(0,model$fitted))-max(0,mean(model$fitted))
  evsi_t <- c(evsi_t, evsi)
}
}

 evsi_n <- matrix(evsi_t,nrow = 100, ncol = 1000)
 mean_evsi <- rowMeans(evsi_n)
 #max_evsi <- rowMaxs(evsi_n)
 #min_evsi <- rowMins(evsi_n)

CI <- apply(evsi_n, 1, quantile, probs = c(0.025,0.975))

N <- seq(from=10, to=1000, by = 10)
data.frame(evsi_n)

total_EVSI <- rowMeans(evsi_n)*(210450-N)
sample_c <- 604*N + 1115841.3 #(183.24 / patient recruited from the systematic
review, 769637 median cost of setting up a trial in UK),  Hind paper)

plot(x = N, y=total_EVSI, type = 'l', main = "EVSI with regard to sample
size", xlab = 'Sample size N', ylab='EVSI(N)')
lines(N, CI[1,], col="red",lty=3)
lines(N, CI[2,], col="red", lty=3)

ENGS_n <- total_EVSI-sample_c

plot(x = N, y=total_EVSI, type = 'l', main = "ENGS and EVSI with regard to
sample size", xlab = 'Sample size N', ylab = 'Value (£)', col='red', lty=2)
lines(x = N, y=ENGS_n)
lines(x = N, y=sample_c, col='blue', lty=2)
abline(v = ENGS$N[which.max(ENGS$ENGS_n)], lty=2)
max(ENGS_n)
ENGS$N[which.max(ENGS$ENGS_n)]
ENGS<- data.frame(cbind(N, ENGS_n))

ENGS$N[which.max(ENGS$ENGS_n)]


modmean <- gam(mean_evsi ~ N)


modU <- gam(CI[1,]~N)
modL <- gam(CI[2,]~N)

predmean<-predict(modmean, newdata=N)
predU<-predict(modU, newdata=N)
predL<-predict(modL, newdata=N)

###################EXAMPLE -- parallel group on safety outcome
############################
x <- NULL
x_M <- NULL
x_R <- NULL
evsi <- NULL
evsi_t <- NULL
```

```
for (j in seq(5000,50000, 200)){
  for (i in 1:length(r_compli)){ #generating an observational study sample
size of 50 with the compliance probability in the parameter
    x_M[i] <- rbinom(1, j, r_relapse_M[i])
    x_R[i] <- rbinom(1, j, r_relapse_R[i])
    x[i] <- log(((x_R[i])/(j-x_R[i]))/((x_M[i])/(j-x_M[i])))
  }
  impute.mean <- function(x) replace(x, is.na(x) | is.nan(x) | is.infinite(x),
mean(x[!is.na(x) & !is.nan(x) & !is.infinite(x)])) #cleaning log odds ratio
due to some zeros (perfect prediction)
  x <- sapply(x, impute.mean)
  model <- gam(BCEApsa$NMB ~ s(x)) #fit spline with the Net benefit and x
  evsi<- mean(pmax(0,model$fitted))-max(0,mean(model$fitted))
  evsi_t<- c(evsi_t, evsi)
}

N <- seq(5000,50000, 200)

evsi <- cbind(N, evsi_t)

plot(x = N, y=evsi_t, type = 'l', main = "EVSI with regard to sample size",
xlab = 'Sample size N', ylab='EVSI(N)')

############################## two key statistics
#####################################################

x <- NULL
x_M <- NULL
x_R <- NULL
r_M <- NULL
r_R <- NULL
evsi <- NULL
evsi_t <- NULL
for (n in 1:500){
for (j in seq(10,3000, 10)){
  for (i in 1:length(r_compli)){ #generating an observational study sample
size of 50 with the compliance probability in the parameter
    x_M[i] <- rbinom(1, j, r_relapse_M[i])
    x_R[i] <- rbinom(1, j, r_relapse_R[i])
    r_M[i] <- x_M[i]/j
    r_R[i] <- x_R[i]/j
  }
  model <- gam(BCEApsa$NMB ~ te(r_M, r_R)) #fit spline with the Net benefit
and x
  evsi<- mean(pmax(0,model$fitted))-max(0,mean(model$fitted))
  evsi_t<- c(evsi_t, evsi)
}
  print(n)
}

evsi_n <- matrix(evsi_t,nrow = length(seq(10,3000,10)), ncol = 500)

N <- seq(10,3000, 10)

mean_evsi <- rowMeans(evsi_n)
#max_evsi <- rowMaxs(evsi_n)
#min_evsi <- rowMins(evsi_n)

CI <- apply(evsi_n, 1, quantile, probs = c(0.025,0.975))

df <- data.frame(N, evsi_n)
plot(x = N, y=rowMeans(evsi_n)*((210450-2*N)), type = 'l', main = "EVSI with
regard to sample size", xlab = 'Sample size N', ylab='EVSI(N)')
lines(N, CI[1,], col="red",lty=3)
lines(N, CI[2,], col="red", lty=3)

sample_c <-  640.1*N*2 + 1115841.3+abs(N*mean(BCEApsa$NMB)) #(183.24 / patient
recruited from the systematic review, 769637 median cost of setting up a trial
in UK),  Hind paper)
```

```
total_EVSI <- rowMeans(evsi_n)*(210450-2*N)

max(total_EVSI)

ENGS_n <- total_EVSI-sample_c #For a parallel group trial, total evsi-sampling
cost + oppotunity costs for the control arm (mean Incremental net benefit*N)

plot(x = N, y=ENGS_n, type = 'l', main = "EVSI, TC and ENGS with regard to
sample size", xlab = 'Sample size N', ylab='Value (£)', xlim=c(0,1000),
ylim=c(-4000000, 4000000))
lines(x = N, y=total_EVSI, col='red', lty=2)
lines(x = N, y= sample_c, col='blue', lty=2)
abline(h=0, lty=2)
abline(v=ENGS$N[which.max(ENGS$ENGS_n)], lty=2)
max(ENGS_n)


ENGS<- data.frame(cbind(N, ENGS_n))

ENGS$N[which.max(ENGS$ENGS_n)]

model <- gam(BCEApsa$NMB ~ s(x)) #fit spline with the Net benefit and x
model$fitted #fitted value of NMB

############## ploygon plot with
interval########################################

# predicts + interval
newx <- seq(min(df$x), max(df$x), length.out=100)
preds <- predict(mod, newdata = data.frame(x=newx),
                 interval = 'confidence')
```