EXPLORING THE USE OF ROUTINE DATA FOR RECRUITMENT AND FOLLOW-UP IN LARGE RANDOMISED TRIALS

Danielle Edwards

Green Templeton College
University of Oxford

A thesis submitted for the degree of

Doctor of Philosophy

Trinity 2019

ABSTRACT

Timely recruitment and comprehensive follow-up of patients are vital to the success of randomised controlled trials (RCTs) but both present issues. This thesis aims to investigate how routine data can be used for recruitment to clinical trials, whether there are any biases introduced by using routine data, and to determine whether routine data can be used as a follow-up method in clinical trials.

The literature on the use of routine data for recruitment and follow-up are explored in two systematic literature searches. Both reviews highlight that routine data can be used for recruitment and follow-up in RCTs. One finding is that these methodological aspects of RCTs, particularly when not pertaining to the results, are often not reported.

Using data from two large cardiovascular disease trials, the Heart Protection Study (HPS) and the Randomized EValuation of the Effects of Anacetrapib through Lipid-modification trial (REVEAL), that used electronic health records to recruit patients, the demographics of those entering the trial and declining to participate are investigated. A significant sex difference is observed, with females being less likely than males to progress through the recruitment pathway of both clinical trials. Attendance at follow-up visits and adherence to treatment were also significantly worse in females than in males. This disparity may leave studies underpowered for sex specific effects. These factors potentially limit the ability to establish the safety and efficacy of interventions in females and have the potential to have serious adverse public health consequences.

The use of routine data for follow-up is investigated by comparing the outcomes from HPS (non-fatal myocardial infarction, non-fatal stroke, revascularisation, and major vascular event) and the outcomes reported in Hospital Episode Statistics (HES) during the same time period. All of the trial outcomes were well recorded in the HES dataset, with no significant differences between males and females in their reporting of events. The use of routine data for follow-up in RCTs could reduce the frequency of study visits, which may aid recruitment.

Routine health records contain a wealth of information about patients that can be used to optimise both recruitment to trials and the follow-up of patients. The use of electronic health records (EHR) has the potential to provide more cost-effective and efficient methods for research.

ACKNOWLEDGEMENTS

This thesis would not have been possible without the support of many wonderful people.

Firstly, I would like to thank my supervisors, Dr Michael Lay, Professor Jane Armitage, Professor Martin Landray and Professor Sir Rory Collins for their help, guidance, and constructive feedback.

I would like to acknowledge the funding I received from the MRC Hubs for Trial Methodology Research for my D.Phil. without which I would not have been able to undertake this research.

Thanks also go to Karl Wallendszus for taking the time to explain some of the aspects of HPS to me and for answering my many questions about the datasets.

I would also like to thank the participants of both clinical trials used for this project, the Heart Protection Study (HPS) and the Randomized EValuation of the Effects of Anacetrapib through Lipid-modification trial (REVEAL), and acknowledge the contribution of the investigators, collaborators, and study teams who made these trials possible.

There are many friends and colleagues who have supported me throughout this process. Thank you to the inhabitants of L2/35 for all their support – I will miss our walks and tea breaks! Special thanks also go to Jenny, Louise, Ravinda, Gina, Ellie, Emily, and Karl for being there through both the good times and bad times.

I'd also like to thank Professor Paul Cullinan and Dr Siobhán Carr for all their support in the final stages of the write-up, and for generously giving me the time to complete it whilst working full-time.

Most importantly, I would like to thank my parents and my sister for their unwavering support over the last few years, for raising me to believe I can achieve anything I set my mind to, and for holding that belief for me when mine faltered. I could not have completed this without you – thank you.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

LIST OF ABBREVIATIONS

ASCEND	A Study of Cardiovascular Events iN Diabetes
BHF	British Heart Foundation
BMJ	British Medical Journal
CABG	Coronary Artery Bypass Graft
CHD	Coronary Heart Disease
CI	Confidence Interval
CTSU	Clinical Trial Service Unit & Epidemiological Studies Unit
CVD	Cardiovascular Disease
EHR	Electronic Health Record
EMR	Electronic Medical Record
FCE	Finished Consultant Episode
GP	General Practitioner
HDL	High Density Lipoprotein
HES	Hospital Episode Statistics
HPS	Heart Protection Study
HTA	Health Technology Assessment
ICD	International Statistical Classification of Diseases and Related Health Problems
IQR	Interquartile Range
LDL	Low Density Lipoprotein
MeSH	Medical Subject Headings
MI	Myocardial Infarction
MRC	Medical Research Council
MVE	Major Vascular Event
NCRAS	National Cancer Registration and Analysis Service
NHS	National Health Service
NIHR	National Institute for Health Research
ONS	Office for National Statistics
OPCS-4	Classification of Interventions and Procedures version 4
OR	Odds Ratio
PRISMA	Preferred Reporting Items for Systematic Reviews and Meta-Analyses
PTCA	Percutaneous Transluminal Coronary Angioplasty
RCT	Randomised Controlled Trial
REVEAL	Randomized EValuation of the Effects of Anacetrapib through Lipid-modification
SAE	Serious Adverse Event
SD	Standard Deviation
TIA	Transient Ischemic Attack
UKRR	UK Renal Registry

1 INTRODUCTION

1.1 INTRODUCTION

Randomised controlled trials (RCTs) are often referred to as the gold standard for conducting medical research[1] and they are essential for answering important medical research questions. Timely recruitment and comprehensive follow-up of patients are vital to the success of these trials. However, both recruitment and follow-up present issues that can prevent trials from succeeding. This thesis aims to identify how routine data can be used to recruit participants into clinical trials, whether there are any demographic differences observed between those recruited using routine data and those not, and to determine whether clinical trial follow-up can utilise routine data to improve trial efficiency.

1.2 WHAT ARE ROUTINE DATA?

The use of “big data” in health research is not a new idea. Many epidemiological studies are conducted using data from routine sources. These include demographic data sources, national reference data sources, and health event data sources. Some examples of routine data sources are census data, birth/death registration data, GP/hospital record data, pharmacy data, disease registers, insurance databases, and health surveys.

There are many benefits to conducting research using routinely collected data. These data are often cheaper to use than conducting your own surveys/studies. They also cover a larger population and datasets may cover a long time period allowing the data to

be linked longitudinally, which allows more participants to be followed up for longer periods of time.

However, there are also issues with using routinely collected data for research. One issue is that the data were not collected for the purposes of the research being conducted. This means that the dataset may not contain all the information required to answer the research question, or that information might be less detailed or accurate than desired. There have also been changes in coding practice over time which may cause confusion. It also means that care should be taken when using the data and the data collection source should be considered. An example of this is when using Hospital Episode Statistic (HES) data. These data are collected primarily for reimbursement and not for research. The data are entered by clinical coders and it may be difficult for them to accurately extract the correct information from clinical notes.[2] Another issue with using routinely collected data is that it may not be up to date or available in a timely manner. This can cause issues in clinical trials when safety information may be needed but is less of an issue for epidemiological studies.

1.3 AN INTRODUCTION TO THE TRIALS

This thesis uses data from two cardiovascular disease trials, the Heart Protection Study (HPS), and the Randomized EValuation of the Effects of Anacetrapib through Lipid-modification (REVEAL) trial. Both of these trials were conducted by the Clinical Trial Service Unit and Epidemiological Studies Unit (CTSU) in Oxford and were investigating the effects of lipid-modifying treatment among people at increased risk of cardiovascular disease.

1.3.1 THE MRC/BHF HEART PROTECTION STUDY

The HPS trial recruited 20 536 participants at high risk of cardiovascular disease between 1994 and 1997. Participants were randomised to receive daily simvastatin 40mg or matching placebo tablets and, separately, to receive antioxidant vitamin supplements (600mg vitamin E, 250mg vitamin C, and 20mg beta carotene) daily or matching placebo capsules in a 2x2 factorial design.[3] Follow-up in HPS was conducted at four, eight, and 12 months after randomisation, and then six monthly thereafter until final follow-up appointments which occurred between May and October 2001.[4]

Potentially eligible patients were identified from hospital records at 69 UK hospitals. A list of the relevant disease codes was sent to each site and the electronic discharge records were searched for patients with these codes.[3] Further information was then sought and with the permission of their doctors, patients were invited by CTSU to attend a screening appointment for the trial.

1.3.2 THE REVEAL TRIAL

The REVEAL study randomised 30 449 international participants with pre-existing atherosclerotic vascular disease between 2011 and 2013. In order to compare REVEAL with HPS, only the UK based patients form the basis of most of this research. This is because different methods were used in other global regions to identify and recruit patients and often included some pre-screening or the payment of participants. In the UK, REVEAL staff from 64 sites recruited 8381 patients and all volunteers were given atorvastatin (either 20mg or 80mg daily) as background therapy and were then randomised to receive anacetrapib 100mg or matching placebo and followed for a median of four years in nurse-led clinics.[5] Similar methods of recruitment were used to HPS; potentially eligible patients with one or more of the inclusion criteria were

identified via their electronic medical records and invitations to participate in the trial were went out in the name of the local collaborator.

1.3.3 THE RECRUITMENT PATHWAY

Throughout this thesis, the "recruitment pathway" will be discussed. In both of the trials used for analysis, this pathway starts from the identification of a potentially eligible patient via central records and the invitation to participate in the trial being sent (Figure 1.1). The recruitment pathway continues with the screening appointment at which eligibility is assessed. After screening, there is a run-in period where participants are asked to take a study treatment (active or placebo). One purpose of this is to identify participants who are likely to be non-adherent to the study treatments. This increases efficiency of the trial and aims to reduce the number of participants dropping out of the trial.[6] Another reason is to identify participants who do not tolerate the treatments or who have perceived side effects from the treatment. Once the run-in period has been completed, participants are invited back to a randomisation appointment. At this point eligibility is reassessed, and consent for participation in the main trial is sought.

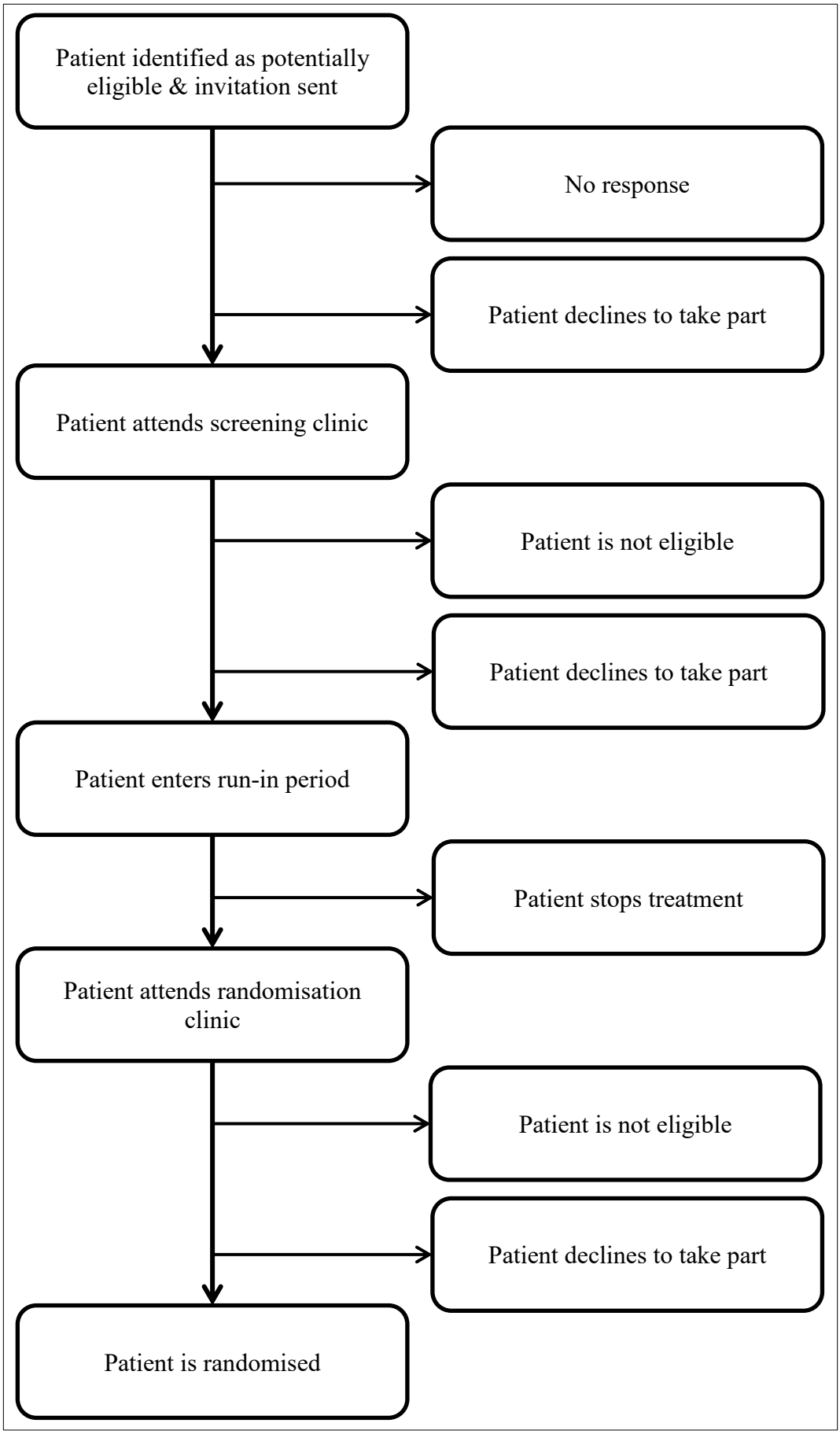


FIGURE 1.1: THE RECRUITMENT PATHWAY

1.4 ISSUES WITH RECRUITMENT IN CLINICAL TRIALS

The reliable detection of moderate treatment effects requires the randomisation of large numbers of suitable patients. This can be difficult to achieve within the necessary time constraints and delays in patient recruitment can have serious repercussions.[7] Recruitment should be conducted in a timely manner as delays in the trial can delay the general population access to potentially beneficial treatments[8] or allow the continuation of an ineffective or even harmful treatment.[9,10] The planning of a randomised controlled trial involves the calculation of a sample size. The sample size should not be too small; otherwise the study may be underpowered which increases the risk of type II error.[11] The sample size should also not be too large. Treating more patients than is necessary raises ethical issues, and also requires more resources (both financial and time).[12]

There are many reasons why patients may not be willing to take part in clinical trials. Table 1.1 lists some of these reasons, and groups them into three categories; concerns with study, inconvenience, and other.

TABLE 1.1: REASONS FOR NOT PARTICIPATING IN CLINICAL RESEARCH

Concerns with study	Inconvenience	Other
Wants standard treatment	Taking time off work for appointments	Religious beliefs
Worried about experimental treatment	Difficulties travelling to study centre	Taking part in other research
Worried about side effects	Too many visits	Relative doesn't want them taking part
Concerns about blood test/needle phobia	Visits take too long	
Want to know which treatment they're taking		
Taking too many medications already		

This list of reasons does not cover all the reasons a patient may not want to take part in research but does illustrate that there are areas that can be addressed that may improve

recruitment. Many of the reasons that are grouped under "inconvenience" could be negated by using alternative forms of follow-up, for example following participants up using their medical records. This would resolve the issues of frequent and long study visits, repeated travel to a study clinic, and reduce the amount of time participants would need to take off work. By using routine data for follow-up and improving the efficiency of clinical trials in this manner, participants may be more open to taking part in trials, thus aiding the recruitment process.

1.5 ISSUES WITH FOLLOW-UP IN CLINICAL TRIALS

One of the implications of incomplete follow-up is that failing to follow-up participants leads to missing data and potentially fewer detected outcomes. This reduces the power and lessens the ability to detect a difference between the treatment arms if one exists. If a treatment difference is not detected when it does in fact exist, patients may not receive treatments that could help them. Another problem with incomplete follow-up is that it may introduce a bias. Attrition bias is a type of selection bias caused by systematic differences between participants lost from each of the study arms.[13]

There has been considerable research into improving retention in clinical trials, with many strategies suggested. These mainly relate to methods for improving questionnaire feedback, and not into encouraging attendance at follow-up visits.[14] If routine data could be used to reduce the number of follow-up visits needed, or to remove the need for in-person follow-up altogether, then it may encourage more patients to participate in clinical trials as it would be less time-consuming for them. It also has the potential to make trials more efficient by reducing the costs for the study sponsors.

1.6 OUTLINE

This thesis aims to investigate how routine data can be used for recruitment and follow-up in large clinical trials. Chapter 2 is a systematic review of randomised controlled trials that have used routine data as a recruitment method. This provides an overview of the current practices, and highlights some of the more commonly used routine data sources. It also provides insight to the work shown in Chapter 3.

In Chapter 3, two different trials that used routine data for recruitment are explored. The trials are introduced, along with the recruitment methods used. The characteristics of those who were successfully recruited are compared with those who did not progress along the recruitment pathway. The demographic differences in those receptive to the recruitment process are investigated and explored further in Chapter 4.

Chapter 4 takes the differences observed in recruitment in Chapter 3 and investigates whether these "pre-trial" recruitment differences are observed "in-trial", particularly with respect to adherence to treatment and attendance at follow-up.

Chapter 5 is a second systematic review, this time looking at randomised controlled trials that have used routine data sources as a method of follow-up. The different types of routine data are summarised, along with the type of trial, the size of the trial, and the length of the follow up period.

Chapter 6 follows on from the literature search in Chapter 5 by taking outcome events collected in a large randomised trial and comparing them to events that were captured in a routine data source. This comparison provides insight into how routine data sources can be used for follow-up in clinical trials and how accurately cardiovascular events can be detected within the HES datasets.

Finally, Chapter 7 contains a summary of all the findings presented in the thesis, along with a discussion of the strengths and limitations of these analyses and potential future areas of research.

2 THE USE OF ROUTINE DATA FOR RECRUITMENT TO CLINICAL TRIALS

2.1 INTRODUCTION

This chapter contains a review of the existing literature of studies that have used electronic health records to recruit participants to clinical trials. This will summarise the current knowledge in this field and provide some information about the use of routine data to inform the investigation of two trials using electronic health records for recruitment in Chapter 3.

2.1.1 RECRUITMENT TO CLINICAL TRIALS

Randomised controlled trials (RCTs) are essential for answering important medical research questions. Timely recruitment and comprehensive follow-up of patients are crucial to the success of these trials. However, both recruitment and follow-up present issues that can prevent trials from succeeding. The reliable detection of moderate treatment effects requires the recruitment of large numbers of suitable patients.[15] This can be difficult to achieve within the necessary time constraints and delays in patient recruitment can have serious repercussions.[7] Recruitment should be conducted in a timely manner as delays in the trial can delay the general population access to potentially beneficial treatments or allow the continuation of an ineffective or even harmful treatment. A study of 122 trials funded by UK grant agencies (MRC, NIHR HTA) found only 31% achieved/passed their recruitment target and approximately 45% of the trials failed to reach 80% of their target recruitment.[16] Another survey of

randomised trials found that 41% of trials published in the BMJ and Lancet in 2000 and 2001 experienced problems with recruitment.[17] It is likely that these figures underestimate the number of trials that did not recruit to their targets as those that did not finish or significantly struggled due to poor recruitment may not have been published. Failing to reach an adequate sample size can lead to a greater chance of encountering a type II error[15] and incorrectly retaining the null hypothesis is wasteful.[18]

One potential method for improving recruitment to clinical trials is large scale identification of potentially eligible participants using electronic medical records, or other health databases. This can be used as a stand-alone method of recruitment or used in combination with other methods. There are different types of routine data that can be used to recruit participants for clinical trials. These include searching electronic health records both hospital and primary care based, accessing national databases, for example, census data, and accessing other commercially held databases, for example using insurance databases.

2.1.2 OBJECTIVES

The objectives of this review are to identify trials where routine data have been used for recruitment to clinical trials, to describe which data have been used and in which disease areas, and to summarise the literature to provide an understanding of how these data are currently used.

2.2 METHODS

This review was undertaken using the guidelines for conducting and reporting a systematic review in the PRISMA (Preferred Reporting for Systematic Reviews and Meta-Analyses) statement.[19]

This review is focussed on recruitment methodology; therefore, there were no age, size, or disease restrictions. All types of intervention were included along with all outcomes. In order to be included in this review, the study had to be an RCT that had used a form of routine data for at least some of the recruitment period. The routine data had to have been used during the recruitment period of the trial, and not done as a retrospective review.

2.2.1 SEARCH STRATEGY

All studies published in English before January 2018 were included in this review. The search strategy used to search the electronic databases was for any of the abstracts to contain one of the following words and phrases: ("clinical trial" or "randomised controlled trial" or "randomised" or "randomisation" or "randomized controlled trial" or "randomized" or "randomization") and ("electronic health record*" or "EHR" or "electronic record*" or "electronic medical record*" or "EMR" or "routine data") and "recruit*". This search strategy was used to identify potentially eligible abstracts from MEDLINE, EMBASE, and Cochrane Review electronic databases. Abstracts of papers identified were reviewed to determine whether they met the criteria to be included in this review; those that did not meet the criteria were excluded. The full text of the remaining papers was then examined to determine whether the papers met the criteria for inclusion and to extract any relevant data from them.

The following data were identified as relevant for this review:

- Type of routine data used
- Disease area of clinical trial
- Recruitment strategy used (just routine data or in combination with others)
- Sample size (expected and achieved)

These data are summarised in a qualitative review about the use of routine data for recruitment in clinical trials. There are no meta-analyses performed given the methodological issues with the broadness of the results.

2.3 RESULTS

Figure 2.1 contains the flowchart listing the process of selection for this review, from the studies identified to those included in the final review. These 52 studies were then assessed for eligibility for this review by reading the full texts. Of these, 31 were excluded after reading the full text. The reasons for exclusion are listed in Table 2.1. There were 16 papers excluded as they were conference abstracts with no full text available, 12 papers that were excluded as they were not an RCT, four that were excluded as routine data had not been used for recruitment, two excluded as they were a retrospective analysis of a current trial, and two excluded as there were no details of the RCT given.

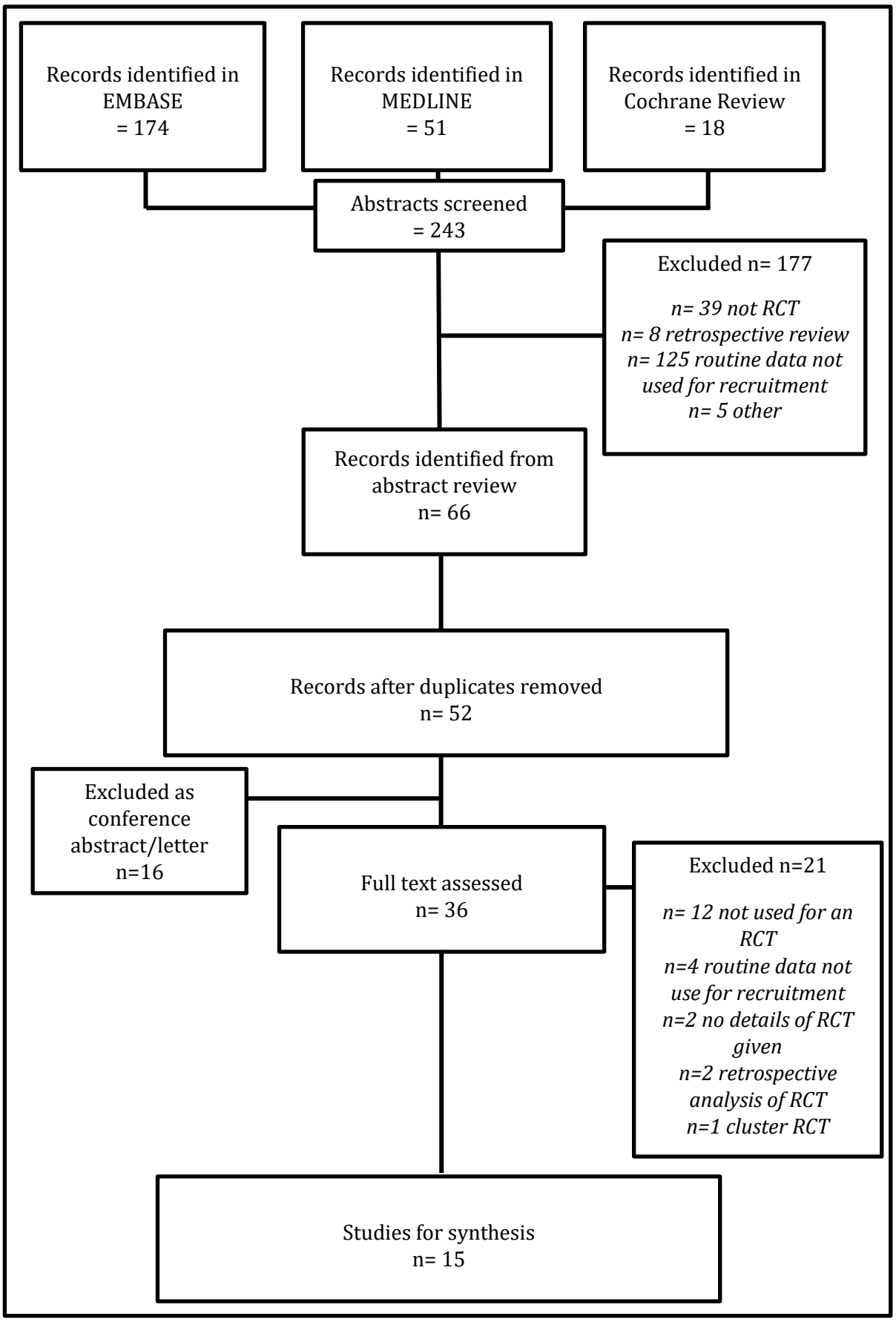


FIGURE 2.1: FLOWCHART SHOWING NUMBER OF STUDIES AT EACH STAGE OF THE REVIEW

TABLE 2.1: STUDIES EXCLUDED FROM THE REVIEW AFTER FULL TEXT REVIEW

Title	Authors	Year	Reason for exclusion
The potential of electronic health record data to optimize recruitment efficiency in cardiovascular outcome trials[20]	Nichols et al.	2017	Rejected - conference abstract, no full text available
Utilizing electronic medical record networks for identifying patients for clinical trial recruitment[21]	Spencer et al.	2015	Rejected - conference abstract, no full text available
Matching the right patient to the right clinical study: The opportunities for more efficient clinical studies using selective-recruitment designs[22]	Barrett	2017	Rejected - conference abstract, no full text available
Conformity between protocol eligibility criteria for electronic patient identification: A comparison of clinical trials[23]	Kamauu	2013	Rejected - conference abstract, no full text available
Feasibility and acceptability of a group medical visit intervention to improve HCV treatment uptake among persons who inject drugs (PWID) in a primary care setting[24]	Norton et al.	2017	Rejected - conference abstract, no full text available

Title	Authors	Year	Reason for exclusion
Recruitment strategies for community-based yoga research in a predominant minority population[25]	Keosaian et al.	2012	Rejected - conference abstract, no full text available
Smoking cessation after surviving critical illness-is it feasible?-Authors experience with implementation and recruitment[26]	Van Nostrand, Fisher & Khan	2014	Rejected - conference abstract, no full text available
Optimizing use of the electronic health record (EHR) to identify persons at high risk of diabetes in the D2D study[27]	Aroda et al.	2016	Rejected - conference abstract, no full text available
Recruitment and enrolment of asthmatics in a phase II clinical trial[28]	Parikh et al.	2010	Rejected - conference abstract, no full text available
Using electronic health records data for clinical research[29]	Wu et al.	2017	Rejected - conference abstract, no full text available
Using electronic medical records to study the safety of psychiatric medications used in pediatric patients[30]	De Luise et al.	2009	Rejected - conference abstract, no full text available
Utilizing electronic medical record networks and advanced analytics for identifying patients for clinical trial recruitment[31]	Justin et al.	2015	Rejected - conference abstract, no full text available

Title	Authors	Year	Reason for exclusion
Home blood pressure tele-monitoring and case management to control hypertension: Hyperlink design, baseline characteristics, and intervention adherence[32]	Margolis et al.	2010	Rejected - conference abstract, no full text available
Utility of electronic medical record databases using ICD-9 criteria for recruitment in clinical research: From rare to common disease[33]	Thacker & Wegele	2015	Rejected - conference abstract, no full text available
EHR-based Clinical Trial Alert Effects on Recruitment to a Neurology Trial across Institutions: Interim Analysis of a Randomized Controlled Study[34]	Khan, et al.	2013	Rejected - conference abstract, no full text available
On target: Identifying and reaching patients for public health interventions or research[35]	Cameron et al.	2010	Rejected - conference abstract, no full text available
Utility of Electronic Medical Record for recruitment in clinical research: from rare to common disease[36]	Thacker, Wegele and Richardson	2016	Rejected – not used in RCT

Title	Authors	Year	Reason for exclusion
Recruit--An Ontology Based Information Retrieval System for Clinical Trials Recruitment[37]	Patrão et al.	2015	Rejected – not used in RCT
Recruiting pregnant smokers into a clinical trial: Using a network-model managed care organization versus community-based practices[38]	Park et al.	2007	Rejected – routine databases not used for recruitment
Strategies to improve retention in randomised trials[14]	Brueton et al.	2013	Rejected – routine databases not used for recruitment
Electronic patient records to identify patients in the United Kingdom with diabetic macular oedema suitable for ILUVIEN (fluocinolone acetonide)[39]	Butt et al.	2016	Rejected – not used in RCT
Comparison of Electronic Health Record System Functionalities to support the patient recruitment process in clinical trials[40]	Schreiweis et al.	2014	Rejected – not used in RCT
Using electronic health records for clinical research: The case of the EHR4CR project[41]	De Moor et al.	2015	Rejected – not used in RCT
Electronic health records to facilitate clinical research[42]	Cowie et al.	2017	Rejected – not used in RCT

Title	Authors	Year	Reason for exclusion
Development of Best Practice Principles for Simplifying Eligibility Criteria[43]	Doods et al.	2013	Rejected – not used in RCT
Strategies designed to help healthcare professionals to recruit participants to research studies[44]	Preston et al.	2016	Rejected – routine databases not used for recruitment
First Experiences in Implementing a Software-Based Support for Patient Recruitment at Heidelberg University Hospital[45]	Schreiweis et al.	2012	Rejected – not used in RCT
Randomised studies In general practice: how to integrate the electronic record[46]	Mosis et al	2005	Rejected – not used in RCT
Database recruitment: a solution to poor recruitment in randomized trials?[47]	Stuardi, Cox & Torgerson	2010	Rejected – not used in RCT
The EHR solution to clinical trial recruitment in physician groups[48]	Miller	2006	Rejected – not used in RCT

Title	Authors	Year	Reason for exclusion
A novel patient recruitment strategy: Patient selection directly from the community through linkage to clinical data[49]	Zimmerman et al.	2017	Rejected - routine databases not used for recruitment
Effect of a clinical trial alert system on physician participation in trial recruitment[50]	Embi et al.	2005	Rejected – details of RCT not given
Cancer clinical trial enrollment of diverse and underserved patients within an urban safety net hospital[51]	Ko et al.	2015	Rejected – retrospective analysis of RCT
The correlation between the number of eligible patients in routine clinical practice and the low recruitment level in clinical trials: a retrospective study using electronic medical records[52]	Sumi et al.	2013	Rejected – retrospective analysis of RCT
Cluster randomized trials utilizing primary care electronic health records: methodological issues in design, conduct, and analysis (eCRT Study)[53]	Gulliford et al.	2014	Rejected – not randomising patients
Using web-based screening to enhance efficiency of HMO clinical trial recruitment in women aged forty and older[54]	Smith et al.	2007	Rejected – not used in RCT
The accuracy and efficiency of electronic screening for recruitment into a clinical trial on COPD[55]	Schmickl et al.	2011	Rejected – details of RCT not given

There were 15 papers that were included in this review. The details of these are listed in Table 2.2 below, along with the key information about the type of routine data used, the disease area, the recruitment strategy, and the sample size. 10 trials (67%) solely used routine data for recruitment. Others used a combination of recruitment methods. 13 (81%) used electronic health records as the only form of routine data – other types included insurance databases, and consented research databases, alongside electronic medical records. The disease area of the trials varied considerably, and the sample size ranged from 29 to 9250. All of the studies are from 2007 onwards; this is not due to a restriction on the dates used for the search, but likely due to difficulties accessing electronic records prior to this date. All of the studies identified by this review were conducted in the UK or the USA. This is likely due to the accessibility of routine data sources in these countries, along with the search terms used.

2.3.1 TYPE OF ROUTINE DATA

All the studies included within this review used electronic health records as a source of routine data. Two of the studies also used another source of routine data to recruit to the trials. These additional sources were consented research databases and insurance claim databases. Eights of the studies searched General Practitioner (GP) or primary care databases.[56–63] Two studies searched through electronic hospital records,[64,65] six accessed USA healthcare systems,[54,65–70] and two searched through insurance databases.[54,65] These are not mutually exclusive categories as some trials used multiple sources.

TABLE 2.2: STUDIES INCLUDED IN THE FULL REVIEW

Title	Country	Years recruiting	Type of routine data	Disease area of RCT	Recruitment strategy	Sample size
Activity Increase Despite Arthritis (AÏDA): phase II randomised controlled trial of an active management booklet for hip and knee osteoarthritis in primary care[60]	UK	Jul 2008 – Aug 2009	GP electronic health records	Arthritis	Routine data	119
Randomised feasibility study of a novel experience-based internet/ intervention to support self-management in chronic asthma[62]	UK	Jun 2013 – Nov 2013	Primary care electronic medical records	Asthma	Routine data	300
Recruitment and retention of Latinos in a primary care-based physical activity and diet trial: The Resources for Health study[61]	USA	Feb 2002 – Aug 2003	Primary care clinic databases	General health - physical activity/diet	Routine data	200
The use of electronic medical records for recruitment in clinical trials: findings from the Lifestyle Intervention for Treatment of Diabetes trial[66]	USA	May 2013 – Mar 2015	Electronic heath record	Diabetes	Routine data, referral from physician	260
Boosting enrolment in neurology trials with Local Identification and Outreach Networks (LIONs)[64]	USA	Jan 2005 – Jun 2007	Hospital electronic records	Neurology	Routine data	281
Results and lessons from the spironolactone to prevent cardiovascular events in early stage chronic kidney disease (STOP-CKD) randomised controlled trial[56]	UK	Jul 2013 – May 2014	Electronic health records	Cardiovascular disease/Kidney disease	Routine data	240

Title	Country	Years recruiting	Type of routine data	Disease area of RCT	Recruitment strategy	Sample size
Desktop software to identify patients eligible for recruitment into a clinical trial: using SARMA to recruit to the ROAD feasibility trial[57]	UK	Dec 2008 – Oct 2009	GP electronic health records	Diabetes	Routine data, clinics, manual review by practice nurse	29
Multicentre, prospective, randomised, open-label, blinded end point trial of the efficacy of allopurinol therapy in improving cardiovascular outcomes in patients with ischaemic heart disease: protocol of the ALL-HEART study[58]	UK	Feb 2014 – Sep 2017	Primary care practice lists, consented research databases	Cardiovascular disease	Routine data + secondary care clinics	5212
Using electronic health record data for substance use Screening, Brief Intervention, and Referral to Treatment among adults with type 2 diabetes: Design of a National Drug Abuse Treatment Clinical Trials Network study[70]	USA	Jan 2007 – Dec 2011	Electronic health records	Diabetes/Drugs Abuse	Routine data	120
Population-based outreach versus care as usual to prevent suicide attempt: study protocol for a randomized controlled trial[65]	USA	Mar 2015 – Jul 2016	Electronic health records, insurance claim databases	Suicide	Routine data	4000

Title	Country	Years recruiting	Type of routine data	Disease area of RCT	Recruitment strategy	Sample size
Frequency of comorbid insomnia, pain, and depression in older adults with osteoarthritis: predictors of enrolment in a randomized treatment trial[69]	USA	2008 - 2010	Electronic health record	Osteoarthritis	Routine data	367
Using patient monetary incentives and electronically derived patient lists to recruit patients to a clinical trial[59]	USA	Nov 2005 – Mar 2017	Primary care electronic medical records	Risk assessment - common chronic diseases	Routine data	1305
Recruitment strategies and challenges in a large intervention trial: Systolic Blood Pressure Intervention Trial[68]	USA	Nov 2010 – Mar 2013	Electronic health records	Blood pressure	Routine data, posters, referrals	9250
The opportunities and challenges of pragmatic point-of-care randomised trials using routinely collected electronic records: Evaluations of two exemplar trials[63]	UK	Retropro: Sep 2012 – Nov 2013 eLung: Feb 2013 – Oct 2013	GP electronic health records	Cholesterol (CVD), antibiotic use in COPD	Routine data	Retropro: 301 eLung:31
Targeted recruitment of adults with type 2 diabetes for a physical activity intervention[67]	USA	Jan 2009 – Apr 2012	Electronic health records	Diabetes	Routine data, poster, clinics	89

2.3.2 DISEASE AREA OF RCT

The trials included in this review looked at a large range of diseases and treatments. Five of the trials were looking at cardiovascular disease outcomes, [56,58,63,66,68] four were looking at diabetes outcomes,[57,66,67,70] two were looking at respiratory outcomes,[62,63] and two were looking at arthritis outcomes.[60,69] There were also studies that looked at neurological outcomes[64], kidney disease[56], mental health[65], drug abuse[70], general health[61], and other common chronic diseases.[59] As before, these outcome categories are not mutually exclusive.

2.3.3 RECRUITMENT STRATEGY

Most of the studies in this review used routine database searches as their only recruitment method.[56,59–65,69]Of these, one study (Simon et al, 2016), used two different sources of routine data; electronic health records and insurance claim databases. Another study (Eakin et al, 2007) raised awareness of the trial through the media and flyers but did not use this as a method of recruitment. Six of the studies used routine data searches alongside other recruitment methods.[57,58,66–68,70] Other recruitment methods ranged from using routine data searches alongside referrals from physicians[70], manual data reviews by a practice nurse[71], hospital clinics and research databases[58], media advertising and community screening[66], and using brochures, flyers, posters[67] and mass mailings.[68]

2.3.4 SAMPLE SIZE

There was a large range in the sample size of the trials in included in this review. The smallest trial recruited 29 participants, whereas the largest trial recruited 9250 participants. The largest two trials used routine data alongside another form of

recruitement[58,68] but some of the smaller trial (n<100) also used multiple recruitment methods.[57,67]

2.4 DISCUSSION

2.4.1 FINDINGS

This review has identified different ways that routine data have been used to recruit participants to randomised controlled trials. Electronic health records are the most common data source for recruitment using routine data, and the use of medical records for recruitment has been proven to be successful for both large and small trials across a variety of disease areas.

2.4.2 LIMITATIONS

One of the main difficulties with this review was finding adequate search terms. There were trials, for example the Heart Protection Study (HPS), that are known to have used electronic health records for recruitment, but the search strategy did not identify them. This is because the MeSH terms used for the search strategy are not always assigned, particularly when relating to methodology. Additionally, many trials do not report the information about the recruitment methods, or do not provide all of the details needed to be included in this review.

Another limitation is the inability to synthesize the results. As it was a methodological review, a meta-analysis was not appropriate, and so no quantitative analysis can be presented.

2.4.3 CONCLUSIONS

As the studies in the review have demonstrated, routine data, particularly electronic health records can be used to recruit participants to clinical trials, either solely or in conjunction with other recruitment methods. Health records can be used to help recruit large numbers of participants to clinical trials in a variety of disease areas. As routine databases contain a wealth of data and it is relatively inexpensive to use them as a recruitment source, they should be considered as a method of recruitment for clinical trials, particularly when a large sample size is required.

There is considerable room for improvement in the reporting of methodology techniques, and this review has highlighted how difficult it can be to find examples of trials using routine data for recruitment. There are other large trials, for example HPS which will be investigated in Chapter 3, that were not picked up by this review. This is due to the outcomes of the trials being the important results, but care should be taken when writing papers to ensure that methodology is addressed as this is an important part of the trial process. One way in which this could be done is by writing a separate study design/methods paper in addition to the primary results paper for all trials. This would ensure that relevant methodology is available for future reviews. Another trial that was not picked up by the review was A Study of Cardiovascular Events iN Diabetes trial (ASCEND) which identified patients from central registers and GP practice records. The trial recruited 15 480 participants from a total of 423 403 invited patients.[72] Whilst the overall proportion of those randomised out of those invited was low (4%), large numbers were able to be recruited due to the relatively inexpensive method of identification and invitation on potentially eligible patients.

This chapter has described how electronic health records can be used to recruit participants into clinical trials. The literature is silent on the subject of demographic bias using routine data for recruitment, and there is considerable room for improvement in the reporting of methodological issues. The next chapter will look at how electronic health records were used to recruit participants to two large RCTs and will explore whether there are any differences in participants recruited via routine data, namely electronic health records to explore whether there are any biases introduced via this method of recruitment.

3 IDENTIFYING DIFFERENCES IN RECRUITMENT

3.1 INTRODUCTION

In this chapter, two trials that both used electronic hospital records to recruit participants are examined. The demographics of the invitees (responders and non-responders) are investigated to determine which factors are most predictive of continuation through the recruitment pathway. These analyses are presented, together with the methods used and a discussion of the results.

3.2 METHODS

Two datasets were identified as being suitable for this research as they were both large cardiovascular disease trials studying a lipid-lowering therapy. Both trials were run by the same centre (Oxford Clinical Trial Service Unit and Epidemiological Studies Unit, CTSU), employed similar methodology, and recruited patients from UK hospitals from a similar patient population. Both trials provide a true denominator of invited patients as potentially eligible trial participants were identified from routine databases. In both trials, the randomisation was preceded by a run-in period, and similar eligibility criteria were used.

3.2.1 MRC/BHF HEART PROTECTION STUDY

The Heart Protection Study (HPS) was a UK based trial looked at cholesterol lowering therapy and antioxidant vitamin supplementation in people at increased risk of heart disease.[73] At the time, there was a lot of uncertainty around the long term benefits of

statins, and they were not routinely prescribed.[74] The study recruited 20 536 volunteers at high risk of vascular disease between 1994 and 1997 and followed volunteers up through six-monthly nurse-led clinic visits for a median of five years over the scheduled treatment period.[3] Volunteers were randomised to receive daily simvastatin 40mg or matching placebo tablets and, separately, to receive antioxidant vitamin supplements (600mg vitamin E, 250mg vitamin C, and 20mg beta carotene) daily or matching placebo capsules in a 2x2 factorial design.[3] Potentially eligible patients were identified by searching hospital computerised records (from particular wards and clinics e.g. the coronary care unit, diabetes clinics) at 69 UK hospital sites for relevant codes; a list of the relevant disease codes was sent to each site and the electronic discharge records were searched for patients with these codes.[3] Information was then obtained on the patient's name, sex, date of birth, vital status at discharge, date of discharge/death, all diagnostic and procedural codes recorded for them, their consultant's name, and their GP's name and address.[3] Particular types of patients were over-sampled by CTSU (those with diabetes, females, older patients) to ensure sufficient numbers in various sub-groups were recruited to allow direct comparisons to be made within these groups.[4] This does not affect the concept of the true denominator but it does mean that there are more invitees in these subgroups than might normally be expected. With the permission of their doctors, patients were invited centrally by CTSU in the name of the local collaborator to attend a local clinic for a health check and a screening appointment.[4] In total, invitations were sent to 130 873 potentially eligible patients, of whom 63 603 attended one of the study screening clinics.[75] At screening, the eligibility of patients was assessed and additional demographic information recorded (Appendix I).

A clinical assessment was conducted, and baseline total cholesterol measured using a desktop analyser.[75] Eligible patients were then asked to consent to enter an 8-10 week

pre-randomisation run-in period and provided with ten weeks of treatment: an initial four weeks of placebo statin and active vitamins, followed by six weeks of both active treatments.[75] The pre-randomisation run-in phase was intended to limit the randomisation to those who would be likely to take their allocated treatment for the following five years.[75] In addition, the patient's centrally measured lipid profile was sent to their general practitioner, and in cases where the GP considered there to be a clear indication or contraindication to statin therapy, that person was not offered randomisation. Patients who were compliant with the study treatments during the run-in and still eligible for the trial were then asked to continue and take part in the randomised trial. Randomised participants were seen in study clinics for routine appointments at four, eight, and 12 months, and then six monthly thereafter until final follow-up appointments which occurred between May and October 2001.[4]

3.2.2 REVEAL TRIAL

The REVEAL (Randomized EValuation of the Effects of Anacetrapib through Lipid-modification) study was international and aimed to recruit at least 30 000 participants with pre-existing atherosclerotic vascular disease to assess the clinical efficacy and safety of anacetrapib 100mg daily.[5] Whilst HPS looked at the use of simvastatin to lower LDL cholesterol, REVEAL looked at how anacetrapib raised HDL cholesterol whilst also lowering LDL cholesterol. In order to compare REVEAL with HPS, only the UK based patients form the basis of most of this research. This is because different methods were used in other global regions to identify and recruit patients and often included some pre-screening or the payment of participants.

In the UK, REVEAL staff from 64 sites recruited 8381 patients with pre-existing atherosclerotic vascular disease (and a total of 30 449 volunteers worldwide).[5]

Recruitment took place between 2011 and 2013 and all volunteers were given atorvastatin (either 20mg or 80mg daily) as background therapy and were then randomised to receive anacetrapib 100mg or matching placebo and followed for a median of four years in nurse-led clinics.[5] Potentially eligible UK patients were identified using similar methods to HPS from a largely overlapping set of UK hospitals; patients with one or more of the inclusion criteria were selected from electronic medical records and invitations to a local screening clinic were sent out in the name of the local collaborator. In REVEAL, there was a small element of pre-screening after invitation; when patients telephoned to confirm their screening appointment, a brief eligibility check was done to remove patients that were obviously ineligible.[76] At the screening visit, relevant medical history, current medications, and other factors relating to eligibility were recorded (Appendix II). Eligible participants had the study explained to them and were asked to consent to take part in the run-in period. The run-in period consisted of 12 weeks taking active atorvastatin (at the appropriate dose) and placebo anacetrapib, with one tablet of each to be taken daily.[5] This single-blind run-in phase was intended to help identify and exclude individuals who would be unlikely to comply with long-term study treatment. As in HPS, the participant's doctor was consulted about their views as to whether the participant was suitable for randomisation. The appropriate dose of atorvastatin was determined by issuing participants with a dose that was intended to reduce their LDL cholesterol to less than 2 mmol/L and to be at least as intensive as their current treatment.[5] Participants who were compliant with the study treatments during the run-in and still eligible for the trial were then asked to continue and take part in the randomised trial. Randomised participants were seen in study clinics for routine appointments at two, six, and 12 months, and then six monthly thereafter until final follow-up appointments. [5]

3.2.3 STATISTICAL METHODS

In order to assess potential associations between demographic factors and progression through the recruitment pathway, a logistic regression analysis was performed for each stage of the process (screening, run-in, and randomisation). Models were adjusted for age, sex, socioeconomic status (assessed using Townsend Index), distance from centre, and ethnicity (where available). The results from the logistic regression analyses are reported as Odds Ratios (OR) and 95% Wald confidence intervals. Analysis of deviance was used to select the best model, and all interactions were considered. All analyses were performed in SAS 9.3 (SAS Institute, Cary NC). Results were considered to be statistically significant when $p<0.05$.

In the HPS dataset, ethnicity was grouped into the three largest categories (White, Indian/Pakistani, and West Indian/Guyanese) and "other". This was due to very small numbers of other ethnic groups. These slightly unusual groupings were in line with practice at the time and have therefore not been renamed in these analyses. In REVEAL, ethnicity was grouped into White, Black, Asian, and "other" for the same reason.

Deprivation was defined using Townsend Index and was calculated using the patient's postcode at invitation and the data from the 1991 census and grouped into tertiles. The Townsend Index uses four variables to assess deprivation; unemployment (as a percentage of those aged 16 and over), non-car ownership (as a percentage of all households), non-home ownership (as a percentage of all households), and household overcrowding. These variables are then standardised using a Z-score and totalled to obtain the Townsend deprivation Index. Positive scores indicate areas with higher deprivation, negative scores indicate areas with lower deprivation.[77] The tertile

boundaries calculated for HPS were also used for REVEAL in order for them to be comparable.

3.3 RESULTS

3.3.1 HPS UNIVARIATE ANALYSIS

A total of 130 873 patients were invited to participate in HPS from 69 UK sites. Of these, 63 603 (48·6%) attended a screening clinic. 32 145 (50·5%) of those attending screening entered the run-in period, and 20 536 (63·9%) of those entering run-in were randomised (see Figure 3.1). The baseline characteristics at each stage contain some missing data for each of the categories. This explains the non-additive nature of the tables, and the details of this are shown in Table 3.1. The characteristics at each stage are listed along with the screening, run-in and randomisation yields (calculated as percentage of previous stage) in Table 3.2. Randomised patients were more likely to be male, older, white, and have a higher socioeconomic status. The largest difference in randomisation yield is associated with sex.

TABLE 3.1: MISSING DEMOGRAPHIC DATA AT EACH STAGE OF RECRUITMENT (HPS)

	Invited n(%)	Screened n(%)	Entered run-in n(%)	Randomised n(%)
Missing sex	416 (0.32%)	1 (<0.01%)	-	-
Missing age	29 (<0.01%)	14 (0.02%)	10 (0.03%)	6 (0.03%)
Missing SES[+]	590 (0.50%)	256 (0.40%)	134 (0.42%)	73 (0.36%)
Missing ethnicity	-	224 (0.35%)	48 (0.15%)	-
Missing all information	27 (0.02%)	-	-	-

[+]*SES – socioeconomic status*

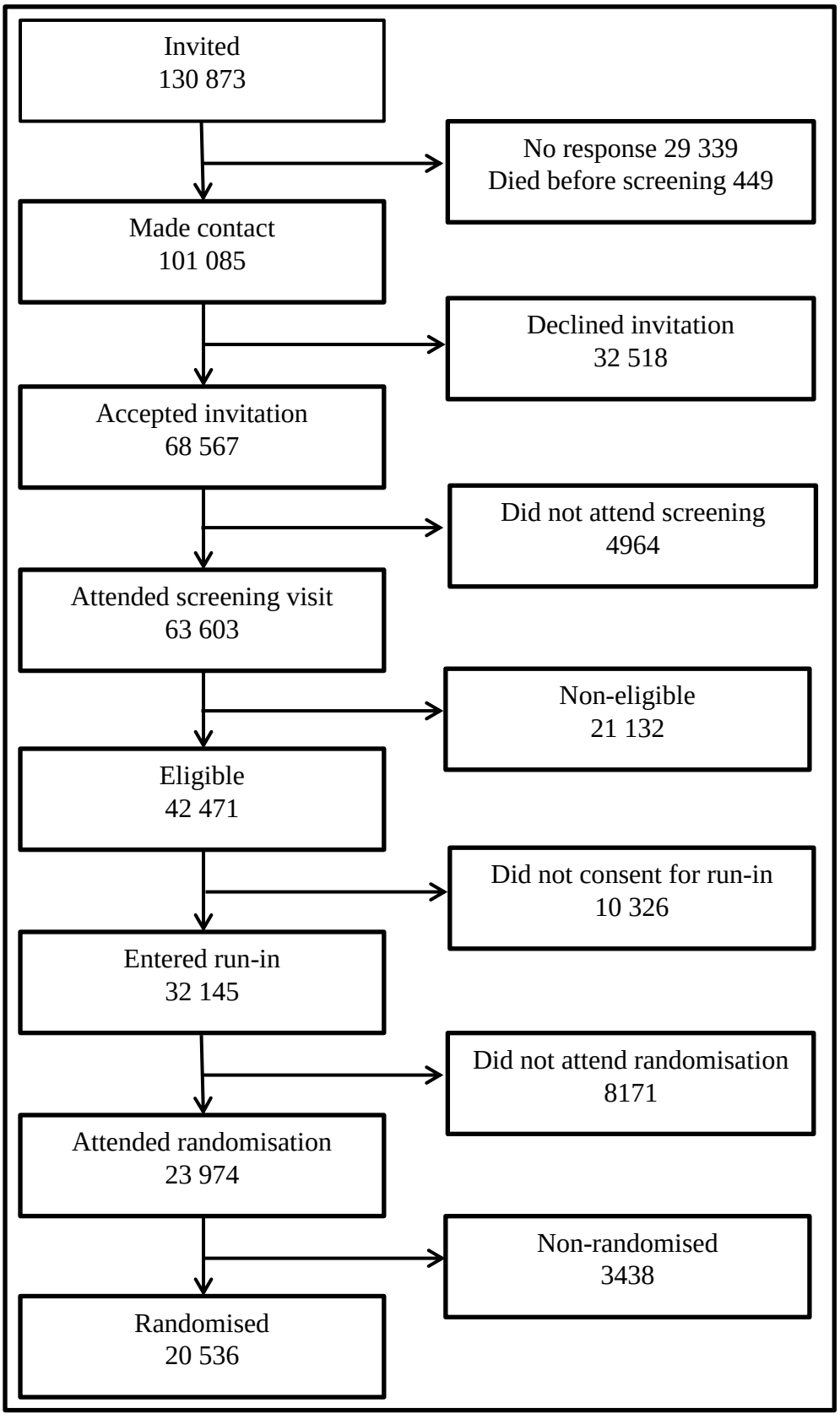


FIGURE 3.1: HPS RECRUITMENT FLOWCHART

TABLE 3.2: BASELINE CHARACTERISTICS BY TIMEPOINT (HPS)

		Invited (N=130 873) n (%)	Screened (N=63 603) n (%)	Entered run-in (N=32 145) n (%)	Randomised (N=20 536) n (%)	Screening yield[+] (%)	Run-in yield[++] (%)	Randomisation yield[+++] (%)	% Randomised of those invited
Sex	Male	82 404 (63.2)	42 162 (66.3)	23 302 (72.5)	15 454 (75.2)	51.2***	55.3***	66.3***	18.8
	Female	48 053 (36.8)	21 440 (33.7)	8843 (27.5)	5082 (24.8)	44.6	41.2	57.5	10.6
Age	Mean (SD)	63.7 (8.9)	64.0 (8.5)	63.6 (8.5)	63.7 (8.4)				
	<50	12 036 (9.2)	5048 (7.9)	2628 (8.2)	1606 (7.8)	41.9	52.1	61.1*	13.3
	≥50 <60	29 094 (22.3)	13 900(21.9)	7494 (23.3)	4773 (23.2)	47.7	53.9	63.7	16.4
	≥60 <70	51 692 (39.6)	26 458 (41.6)	13 465 (41.9)	8633 (42.0)	51.1	50.9	64.1	16.7
	≥70	38 022 (29.1)	18 183 (28.6)	8548 (26.6)	5558 (27.0)	47.8	47.0	65.0	14.6
Ethnicity	White		59 502 (93.9)	30 792 (95.9)	19 901 (96.9)	-	51.7***	64.6***	33.4
	Indian/Pakistani		1234 (1.9)	418 (1.3)	198 (1.0)	-	33.9	47.4	16.0
	West Indian/Guyanese		2095 (3.3)	678 (2.1)	332 (1.6)	-	32.4	49.0	15.8
	Other		548 (0.9)	209 (0.7)	105 (0.5)	-	38.1	50.2	19.2
Deprivation Score	Median (IQR)	0.3 (-2.4, 3.2)	-0.4 (-2.7, 2.6)	-0.7 (-2.9, 2.3)	-0.9 (-3.0, 2.1)				
(Townsend Index)	Advantaged	43441 (33.4)	24165 (38.1)	13001 (40.6)	8724 (42.6)	55.6***	53.8***	67.1***	20.1
	Intermediate	43430 (33.3)	21207 (33.5)	10661 (33.3)	6815 (33.3)	48.8	50.3	63.9	15.7
	Deprived	43412 (33.3)	17975 (28.4)	8349 (26.1)	4924 (24.1)	41.4	46.4	59.0	11.3
Baseline Total Cholesterol	<5.5 mmol/L		24 119 (41.4)	12 020 (39.7)	8001 (41.0)	-	49.8	66.6***	33.2
	≥5.5 < 7.0 mmol/L		28 377 (48.7)	15 255 (50.4)	9780 (50.1)	-	53.8	64.1	34.5
	≥7.0 mmol/L		5797 (9.9)	2978 (9.9)	1750 (8.9)	-	51.4	58.8	30.2

*+ screening yield = percentage screened out of invited, ++ run-in yield = percentage entering run-in out of those screened, +++randomisation yield = percentage randomised out of those entering run-in, * $\chi 2$ $p<0.05$, ** $\chi 2$ $p<0.01$, *** $\chi 2$ $p<0.001$ – $\chi 2$ tests were used for categorical data and $\chi 2$ tests for trend were performed for ordinal data*

The recruitment pathway was investigated further to determine where the difference between the sexes occurred (Table 3.3). Invited females were less likely to attend a screening visit than males (44·6% vs. 51·1%) having been less likely to respond to the invitation or to respond indicating that they did not wish to participate. Males were more likely to die between the date of invitation and the appointment date and were less likely to attend an appointment they had previously accepted. Both these differences were much smaller though. A total of 63 603 patients attended a screening clinic. Of those who attended screening, more females than males were ineligible (40·8% vs. 29·4%: $p<0.0001$) for a variety of medical and non-medical reasons. These reasons included no previous cardiovascular disease, taking contraindicated drugs, and likely non-attendance at clinics. A full list of eligibility criteria is presented in Appendix I, along with a breakdown of the inclusion criteria met by those eligible for the trial by sex (Appendix III). Females were more likely to not meet the inclusion criteria or to be ineligible due to likely non-attendance.

Of those deemed eligible at screening, fewer females than males consented to enter the run-in (69·7% vs. 78·2%: $p<0.0001$). A total of 23 302 males and 8843 females entered the run-in but females were less likely than males to complete it and attend a randomisation visit (69·7% vs. 76·4%: $p<0.0001$). This was largely due to differences in volunteers' wishes and non-attendance. Conversely, drop-out due to abnormal blood results was lower in females than in males. Further investigation revealed that males were more likely to have low cholesterol ($p<0.0001$) and females were significantly more likely to be excluded due to a high liver function test ($p<0.0001$). After completing the run-in period, 23 974 males and females attended their randomisation appointment, but females were less likely than males to be randomised at the visit (82·4% vs. 86·8%: $p<0.0001$). The main reasons for this difference were reluctance to

continue, likely problems with clinic visits, non-compliance with run-in treatment, and unexplained muscle pain/weakness during the run-in period.

TABLE 3.3: RECRUITMENT PATHWAY BY SEX (HPS)

	Male n (%)	Female n (%)	P value
Total invited	82 404	48 053	
No response	17 980 (21.8)	11 156 (23.2)	<0.001
Replied no	18 757 (22.8)	13 581 (28.3)	<0.001
Replied yes – did not attend screening	3191 (3.9)	1741 (3.6)	0.002
Died before screening	314 (0.4)	135 (0.3)	0.003
Attended screening	42162	21440	<0.001
Non-eligible	12 383 (29.4)	8748 (40.8)	<0.001
Non-eligible, no inclusion met[+]	1680 (13.6)	1328 (15.2)	<0.001
Non-eligible, exclusion criteria[+]	6715 (54.2)	4709 (53.8)	<0.001
Non-eligible, likely non-compliance[+]	5877 (47.5)	4238 (48.4)	<0.001
Eligible	29 779 (70.6)	12 692 (59.2)	<0.001
Eligible, non-consent	6477 (21.8)	3849 (30.3)	<0.001
Eligible, entered run-in	23 302	8843	<0.001
Drop-out (GP/collaborator veto)	764 (3.3)	309 (3.5)	0.37
Died during run-in	52 (0.2)	15 (0.2)	1.00
Abnormal bloods	918 (3.9)	207 (2.3)	<0.001
Participant choice	3760 (16.1)	2146 (24.3)	<0.001
Attended randomisation appointment	17 808	6166	<0.001
No longer eligible[+]	730 (4.1)	302 (4.9)	<0.001
Non-compliance[+]	2772 (15.6)	1303 (21.1)	<0.001
Other reasons[+]	560 (3.1)	267 (4.3)	<0.001
Non-consent	28 (0.2)	9 (0.1)	0.10
Randomised	15454	5082	<0.001

[+]*Not mutually exclusive*

3.3.2 HPS MULTIVARIATE ANALYSIS

In order to determine and quantify the relationship between recruitment status and demographic factors, a logistic regression analysis was performed for each of the recruitment stages (screening, run-in, and randomisation). The factors that were significant for predicting whether patients attended screening were sex, age, and

socioeconomic status (Table 3.4). Females were significantly less likely to attend a screening visit than males (OR: 0·79, 95% CI 0·77-0·81). Volunteers who attended their screening appointment were more likely to be male and to have a higher socioeconomic status.

Ethnicity was also recorded at the screening visit and the factors that were significant for predicting whether screened volunteers entered run-in were sex, age, socioeconomic status, and ethnicity (Table 3.5). Although white volunteers were significantly more likely to enter the run-in period than other ethnic groups, it should be noted that there were very small numbers of volunteers from other groups. Females were much less likely than males to enter the run-in period (OR: 0·57, 95% CI 0·55-0·59).

Females were also significantly less likely than males to be randomised (OR: 0·69, 95% CI 0·66-0·73). As seen at each of the previous stages, randomised volunteers were more likely to be male and to have a higher socioeconomic status (Table 3.6). There were no significant interactions between sex and socioeconomic status at screening, run-in, or randomisation.

TABLE 3.4: FACTORS AFFECTING ATTENDANCE AT SCREENING (HPS)

	Attendance at screening		
	Est OR	OR 95% CI	p-value
Sex			
Male	-	-	-
Female	0.79	0.77-0.81	<0.001
Age Group			
<50	-	-	-
≥50 <60	0.88	0.85-0.90	0.004
≥60 <70	0.69	0.66-0.72	<0.001
≥70	0.86	0.85-0.90	0.057
SES[+]			
Advantaged	-	-	-
Intermediate	0.77	0.75-0.79	0.14
Deprived	0.57	0.56-0.59	<0.001

[+]*SES – socioeconomic status*

TABLE 3.5: FACTORS AFFECTING ENTERING RUN-IN (HPS)

	Entering run-in		
	Est OR	OR 95% CI	p-value
Sex			
Male	-	-	-
Female	0.57	0.55-0.59	<0.001
Age Group			
<50	-	-	-
≥50 <60	0.84	0.81-0.88	<0.001
≥60 <70	1.12	1.05-1.19	0.01
≥70	1.16	1.11-1.21	<0.001
SES[+]			
Advantaged	-	-	-
Intermediate	0.91	0.88-0.94	0.92
Deprived	0.82	0.79-0.86	<0.001
Ethnicity			
White	-	-	-
Indian/Pakistani	0.53	0.47-0.59	0.02
West Indian/Guyanese	0.42	0.38-0.47	<0.001
Other	0.56	0.47-0.67	0.41

[+]*SES – socioeconomic status*

TABLE 3.6: FACTORS AFFECTING RANDOMISATION (HPS)

	Randomisation		
	Est OR	OR 95% CI	p-value
Sex			
Male	-	-	-
Female	0.69	0.66-0.73	<0.001
Age Group			
<50	-	-	-
≥50 <60	1.01	0.95-1.07	0.2
≥60 <70	0.92	0.84-1.00	0.04
≥70	1	0.95-1.65	0.33
SES[+]			
Advantaged	-	-	-
Intermediate	0.89	0.84-0.94	0.2
Deprived	0.75	0.71-0.79	<0.001
Ethnicity			
White	-	-	-
Indian/Pakistani	0.58	0.48-0.70	0.15
West Indian/Guyanese	0.54	0.46-0.63	0.01
Other	0.59	0.45-0.77	0.31

[+]*SES – socioeconomic status*

The different proportions of males and females reaching each stage of the recruitment process are shown in Figure 3.2. At each stage, fewer females agreed to continue taking part in the study resulting in an absolute 8% difference in the proportion randomised.

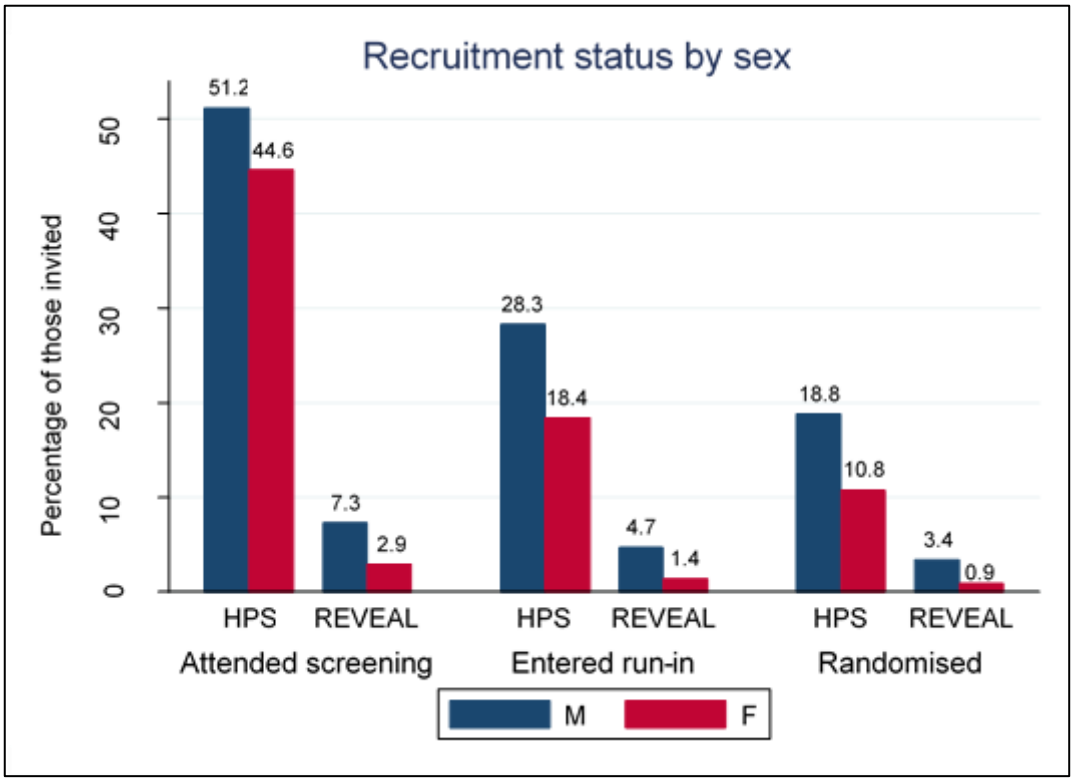
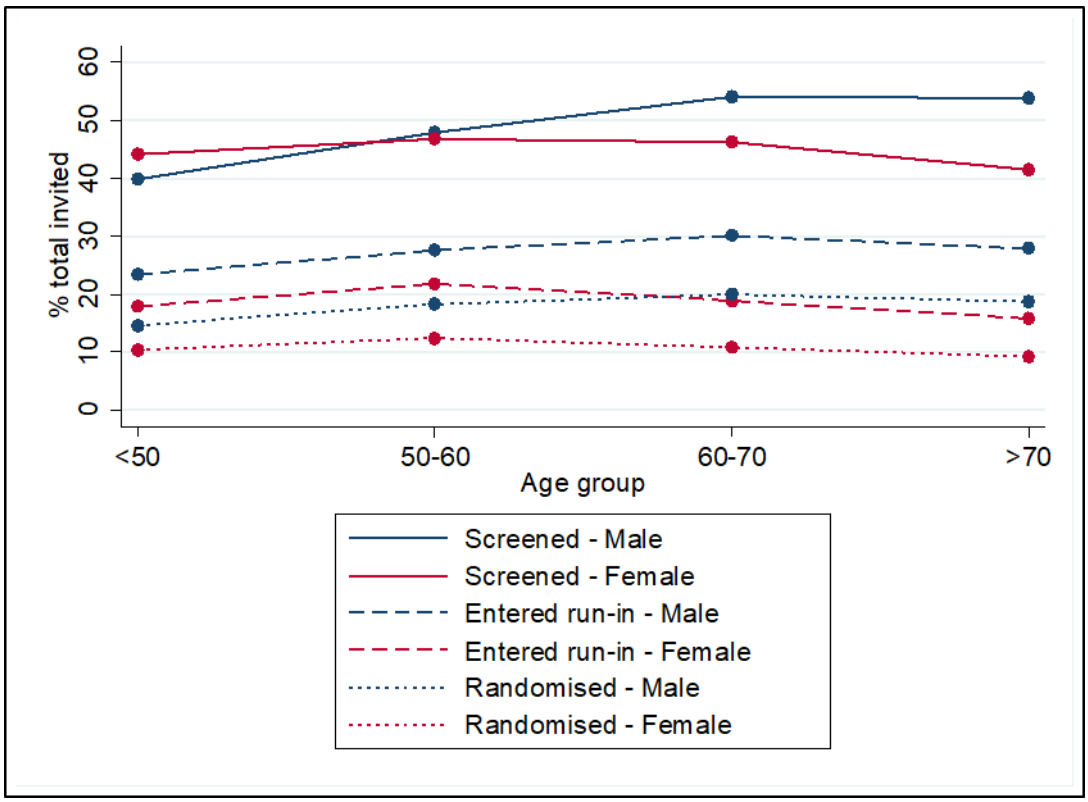


FIGURE 3.2: RECRUITMENT STATUS BY SEX COMPARING HPS AND REVEAL

The effects of age were looked at as it was hypothesised that the observed differences in sex could be due to the ages of participants. Figure 3.3 shows the percentage of participants progressing to the next stage of the recruitment pathway by age and sex. Age appears to have little effect on the percentage of participants moving along the pathway and similar differences are observed in sex within each age group. There is a slight difference in those attending screening (with a larger percentage of females aged <50 attending compared to males), however this is likely due to comparatively small numbers of people in this age group being invited.



Point estimates and 95% CIs plotted for each age group

FIGURE 3.3: PERCENTAGE OF PARTICIPANTS PROGRESSING TO NEXT STAGE BY AGE AND SEX (HPS)

3.3.3 REVEAL (UK VOLUNTEERS) UNIVARIATE ANALYSIS

A total of 336 490 UK patients were invited to participate in REVEAL. Of those invited 19 125 (5·7%) attended a screening appointment, 11 772 (3·5%) entered the pre-randomisation run-in period, and 8382 (2·5%) were randomised into the main trial. Overall, the proportion of those that were randomised of those invited was much lower in REVEAL than in HPS. The proportion of females invited reaching each stage of the recruitment pathway is less than the proportion of invited males resulting 0·9% of invited females being randomised compared to 3·4% of invited males (Figure 3.2).

The characteristics at each stage of the recruitment process are listed in Table 3.7. The patterns in recruitment are similar to those seen in HPS: volunteers who agreed to

continue at each stage were more likely to be male, older, white and of a higher socioeconomic status.

TABLE 3.7: BASELINE CHARACTERISTICS BY TIME POINT (REVEAL)

		Invited (N=336 490) n (%)	Screened (N=19 125) n (%)	Entered run-in (N=11 772) n (%)	Randomised (N=8382) n (%)	Screening yield[+] (%)	Run-in yield[++] (%)	Randomisation yield[+++] (%)	% Randomised of those invited
Sex	Male	213 701 (63.5)	15 572 (81.4)	10 007 (85.5)	7310 (87.2)	7.3***	64.3***	73.0***	3.4
	Female	122 789 (36.5)	3553 (18.6)	1765 (14.5)	1072 (12.8)	2.9	49.7	60.7	0.9
Age	Mean (SD)								
	≥50 <60	45 338 (13.4)	3032 (15.8)	1780 (15.1)	1303 (15.6)	6.7	58.7	73.2	2.9
	≥60 <70	86 297 (25.7)	7360 (38.5)	4652 (39.5)	3480 (41.5)	8.5	63.2	74.8	4.0
	≥70	204 855 (60.9)	8733 (45.7)	5340 (45.4)	3599 (42.9)	4.3	61.1	67.4	1.8
Ethnicity	White		15 149 (97.0)	10 805 (97.5)	8198 (97.8)	-	71.3***	75.9***	-
	Black		59 (0.4)	28 (0.3)	13 (0.2)	-	47.5	46.4	-
	Asian		350 (2.2)	213 (1.8)	136 (1.6)	-	60.9	63.8	-
	Other		67 (0.4)	45 (0.4)	34 (0.4)	-	67.2	75.6	-
Deprivation Score (Townsend Index)	Median (IQR)	-0.8 (-2.9, 2.1)	-1.8 (-3.4, 0.9)	-1.9 (-3.5, 0.6)	-1.9 (-3.5, 0.5)				
	Advantaged	118 241 (35.1)	8627 (45.1)	5543 (47.1)	4020 (48.0)	7.3***	64.3***	72.5***	3.4
	Intermediate	115 084 (34.2)	6466 (33.8)	3961 (33.6)	2808 (33.5)	5.6	61.3	70.9	2.4
	Deprived	99 133 (31.2)	4032 (21.1)	2268 (19.3)	1553 (18.5)	4.1	56.3	68.5	1.6

*+ screening yield = percentage screened out of invited, ++ run-in yield = percentage entering run-in out of those screened, +++randomisation yield = percentage randomised out of those entering run-in, * $\chi 2$ p<0.05, ** $\chi 2$ p<0.01, *** $\chi 2$ p<0.001 – χ^2 tests were used for categorical data and χ^2 tests for trend were performed for ordinal data*

Further detail was sought to explain where females discontinued the study recruitment pathway (Table 3.8). Fewer invited females were eligible and consented to enter the run-in period (1·4% invited females vs. 4·7% males: p<0.0001). Of those that did attend a screening visit, more females than males were not eligible to participate in the trials (44·7% compared to 31·4%) for a variety of reasons (mainly likely non-compliance or taking a contra-indicated drug). Of those screened who were eligible, 89·8% of females consented to enter the run-in phase compared to 93·7% of males. A total of 9784 volunteers completed run-in and attended their randomisation appointment. A significantly higher proportion of females dropped out during run-in and did not attend their randomisation appointment (23·6% of females compared to 15·7% of males: p<0.0001). Of those attending a randomisation appointment, females were less likely to still be eligible than males (84·6% vs. 91·9%: p<0.0001) and were also less likely to consent to randomisation. As a consequence, of the 10 007 males and 1765 females who entered run-in, a total of 8382 volunteers were randomised; females were less likely to be randomised than males (79·3% vs 86·7%: p<0.0001).

TABLE 3.8: RECRUITMENT PATHWAY BY SEX (REVEAL)

	Male n (%)	**Female n (%)**	**P value**
Total invited	213 685	122 785	
Did not attend screening	198 114 (92.7)	119 231 (97.1)	<0.001
Attended screening	15 571	3554	
Non-eligible	4889 (31.4)	1589 (44.7)	<0.001
Eligible	10682 (68.6)	1965 (55.3)	<0.001
Eligible, non-consent	675 (6.3)	200 (10.2)	<0.001
Eligible, entered run-in	10 007	1765	
Withdrew during run-in	1575 (15.7)	413 (23.6)	<0.001
Attended randomisation appointment	8432	1352	
No longer eligible	751 (8.9)	208 (15.4))	<0.001
Non-consent	371 (4.4)	72 (5.3)	0.13
Randomised	7310	1072	

3.3.4 REVEAL (UK VOLUNTEERS) MULTIVARIATE ANALYSIS

As in HPS, in order to determine the relationship between recruitment status and the demographic factors a logistic regression analysis was performed for each of the recruitment stages (screening, run-in, and randomisation). The factors that were significant for predicting whether a patient attended screening were sex, age, and socioeconomic status (Table 3.9). The results show that females were significantly less likely to attend a screening visit than males (OR: 0.43, 95%CI 0.41-0.44); this is a stronger association than was seen in HPS. As seen in the univariate analysis, participants who attended their screening appointment were more likely to be male, aged between 60 and 70, and to have a higher socioeconomic status (determined using Townsend index). Significant factors for predicting whether screened patients entered run-in were sex, age, socioeconomic status, and ethnicity (Table 3.10). Females were much less likely than males to enter the run-in period (OR: 0·53, 95% CI 0·49-0·58). Again, as in HPS, white patients were significantly more likely enter the run-in period and be randomised than other ethnic groups but there were very small numbers of patients from the other ethnic groups. As seen at the other stages of the recruitment process, females were significantly less likely than males to be randomised (OR: 0·59, 95% CI 0·53-0·66). Randomised volunteers were more likely to be male, white, and have a higher socioeconomic status (Table 3.11). There were no significant interactions between sex and socioeconomic status at screening, run-in, or randomisation.

TABLE 3.9: FACTORS AFFECTING ATTENDANCE AT SCREENING (REVEAL)

		Attendance at screening		
		Est OR	OR 95% CI	p-value
Sex				
	Male	-	-	-
	Female	0.43	0.41-0.44	<0.001
Age Group				
	≥50 <60	-	-	-
	≥60 <70	1.26	1.20-1.31	<0.001
	≥70	0.65	0.62-0.67	<0.001
SES[+]				
	Advantaged	-	-	-
	Intermediate	0.76	0.73-0.78	<0.001
	Deprived	0.52	0.50-0.54	<0.001

[+]*SES – socioeconomic status*

TABLE 3.10: FACTORS AFFECTING ENTERING RUN-IN (REVEAL)

		Entering run-in		
		Est OR	OR 95% CI	p-value
Sex				
	Male	-	-	-
	Female	0.53	0.49-0.58	<0.001
Age Group				
	≥50 <60	-	-	-
	≥60 <70	1.21	1.09-1.34	<0.001
	≥70	1.08	0.97-1.19	0.54
SES[+]				
	Advantaged	-	-	-
	Intermediate	0.92	0.85-1.00	0.23
	Deprived	0.79	0.72-0.86	<0.001
Ethnicity				
	White	-	-	-
	Black	0.4	0.24-0.68	0.01
	Asian	0.62	0.50-0.78	0.55
	Other	0.82	0.49-1.36	0.36

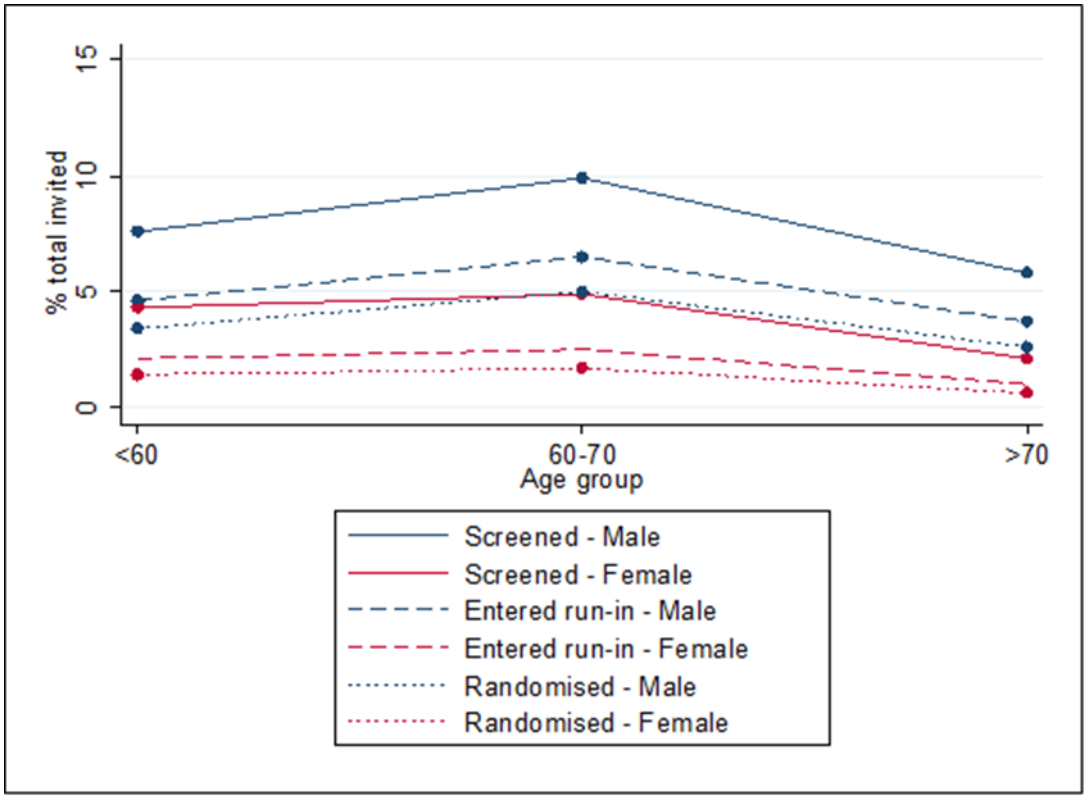
[+]*SES – socioeconomic status*

TABLE 3.11: FACTORS AFFECTING RANDOMISATION (REVEAL)

		Randomisation		
		Est OR	OR 95% CI	p-value
Sex				
	Male	-	-	-
	Female	0.59	0.53-0.66	<0.001
Age Group				
	≥50 <60	-	-	-
	≥60 <70	1.07	0.93-1.22	<0.001
	≥70	0.75	0.66-0.86	<0.001
SES[+]				
	Advantaged	-	-	-
	Intermediate	0.93	0.84-1.03	0.66
	Deprived	0.83	0.74-0.94	0.01
Ethnicity				
	White	-	-	-
	Black	0.29	0.14-0.63	0.01
	Asian	0.53	0.39-0.70	0.37
	Other	0.93	0.47-1.85	0.14

[+]*SES – socioeconomic status*

As in HPS, the relationship between age and sex was explored (Figure 3.4). Age does not appear to be causing the sex difference as there are similar differences between the sexes observed in each age group.



Point estimates and 95% CIs plotted for each age group

FIGURE 3.4: PERCENTAGE OF PARTICIPANTS PROGRESSING TO NEXT STAGE BY AGE AND SEX (REVEAL)

3.4 DISCUSSION

3.4.1 FINDINGS

Sex was the main demographic factor that determined successful recruitment in these cardiovascular disease trials, with significantly fewer females agreeing to participate/progress to the next stage than males. This was the case in HPS and also in REVEAL nearly 20 years later. Although the overall proportion of invited patients who agreed to attend a screening appointment was lower in the more recent REVEAL trial, similar sex differences were seen in both trials. Additionally, the proportion of patients identified from electronic health records that respond to postal invitations has declined substantially over this time period. This suggests the main issue is encouraging patients to attend a screening appointment in the first place. However, the differences in sex are

not driven solely by one stage of the recruitment process as at every stage, females were less likely to continue. Fewer females responded to the invitation, and those that did respond were less likely to accept the invitation than males. At the screening visit, females were less likely to be eligible, and less likely to consent to enter the run-in period. A smaller proportion of females attended the randomisation visit than males, and those that did were less likely to be randomised.

The striking decline in response to the initial postal invitation between these two trials 15-20 years apart is in keeping with trends in other cardiovascular disease trials run by CTSU [9,78] but represents an additional challenge for those recruiting into trials when only a postal address is known. Wider use of email or other recruitment methods may help alleviate this but also present their own difficulties.[7] While these data do not allow further exploration of the underlying causes of these sex differences, recognition by investigators and consideration of these issues when designing recruitment materials, in particular to ensure that there are no hidden sex biases which might be putting females off attending and taking part, and when discussing adherence with participants may help to mitigate some of these effects. Better understanding from qualitative studies of the reasons for these differences in trial related behaviour between males and females may also be of value.

One of the reasons females may be more likely to decline an invitation to take part in a research trial is that they are potentially more likely to be care-givers and the time taken attending appointments (either frequency or duration) may be unappealing. A solution to this would be a greater range of appointment times, shorter study visits, or less frequent study visits. Another reason may be that females perceive themselves to be at less risk of cardiovascular disease than males and therefore may not want to take part in

cardiovascular disease trials. A further reason is that female participants may be more risk-adverse than males and therefore may be less likely to participate in trials. Clear literature about the risks of cardiovascular disease and the potential benefits of taking part in the trial should be included in the trial information letters inviting patients to take part. To ensure that there are no hidden biases, focus groups and other qualitative research should be conducted with both male and female participants.

3.4.2 LIMITATIONS

There are some limitations to this research. Firstly, this research has focussed on two similar trials, HPS and REVEAL, which studied the effects of lipid-modifying treatment among people with cardiovascular disease. The observed sex imbalance may not be generalisable to studies of different designs, interventions or diseases. These observations on trial recruitment are based on UK patients only as this method of unselected invitation of potentially eligible patients based on electronic hospital record data was not deployed elsewhere. However, these results ensure that data on response to invitation are based on a true known denominator and the use of a common method across very large number of patients. The differences observed in recruitment in these two trials may be specific to trials, or even more specifically cardiovascular disease trials. A review of 156 trials looking at cardiovascular disease prevention found that whilst the recruitment of females to cardiovascular disease (CVD) trials has improved in the last 30 years, there is still a significant sex imbalance. However, a large cohort study (UK BIOBANK) found that response rates were higher in females than in males[79], indicating this may be trial or disease specific.

3.4.3 CONCLUSIONS

In summary, sex has a significant effect on recruitment at each stage of the process with relatively fewer females continuing through the trial recruitment pathway at each step. This is important as trial populations should aim to be representative of the overall treatment population, and this sex imbalance throughout recruitment could limit the ability to establish the safety and efficacy of interventions in females.

This chapter has looked at the demographics of responders and non-responders, with sex being the most predictive factor for progression through the recruitment pathway. The next chapter will look at the differences in adherence to treatment by sex, and also assess whether sex affects the likelihood of a participant attending a study clinic in person.

4 ADHERENCE TO TREATMENT IN HPS AND REVEAL

4.1 INTRODUCTION

The previous chapter demonstrated far fewer females being randomised into these two cardiovascular disease trials (The Heart Protection Study, HPS, and the Randomized EValuation of the Effects of Anacetrapib through Lipid-modification study, REVEAL), so this chapter aims to investigate the in-trial differences between males and females. This will allow the pathway from randomisation to end of study to be explored and will investigate the sex effect on method of follow-up, and with adherence to treatment.

4.1.1 FOLLOW-UP METHODS

Follow-up is vital to the assessment of outcomes in clinical trials.[80] There are many different types of follow-up in clinical trials including face-to-face visits, telephone follow-up, follow-up via a GP, follow-up via medical records, and follow-up via postal questionnaire. Clinical trials use a variety of methods to follow participants up[81] but others only use one.[82] The statistical power of a randomised controlled trial can be reduced substantially by incomplete follow-up, and it can be difficult to accurately measure the trial outcomes when the follow-up is incomplete.[83] There have been many studies looking at methods of follow-up and how to improve response rates[84–87] but these have focussed on changes to the follow-up and not on the demographics of those not attending. It is difficult to find examples of sex differences in follow-up in the literature as methods of follow-up are rarely reported by subgroups.

4.1.2 MEASURES OF ADHERENCE

Adherence is defined as the extent to which a patient acts in accordance with the prescribed timing, dosing, and frequency of medication administration.[88] There is currently no gold-standard for assessing treatment adherence, and there are a variety of methods that can be used. These can be divided into direct methods and indirect methods. An example of a direct method is taking a blood or urine sample to assess the levels of a biomarker.[89] This is the most accurate form of measuring adherence, but is not available for all medications, is often expensive, is time-consuming, and is more invasive for the participant. This in turn can discourage patients from wanting to take part in the research.[90] Indirect methods of measuring adherence include pill counts, patient diaries, interviews with participants, and looking at whether the prescription was filled.[91,92] These methods are often cheaper ways of assessing treatment adherence, but are subject to bias.[93] Interviews and other self-reported methods of adherence are likely to overestimate the adherence, whereas prescription filling is more likely to be close to the true adherence. However, prescription filling does not measure whether a participant has taken their treatments, just whether they collected them.[90] Interestingly, research has shown that patients adherent to their study treatments have better outcomes, even if assigned to a placebo group in a controlled trial. This suggests that adherent participants may differ in other health behaviours compared to their non-adherent counterparts.[94]

4.1.3 ADHERENCE BY SEX

There are some observed differences in treatment adherence by sex. However, adherence seems to vary across different disease areas (Table 4.1). A Danish study looking at first episode psychosis in the schizophrenia spectrum found that females

were more compliant with their treatments than males.[95] A large study looking at linking pharmacy data for prescription refills in the US found large amounts of variation by medical condition with significantly fewer females collecting their medications for hypertension, hyperlipidaemia, and diabetes.[96] Conversely, the same study demonstrated that females were more adherent than males to treatments for osteoporosis, cancer, asthma, and multiple sclerosis (MS). Furthermore, a study in The Netherlands found that females were significantly less likely than males to be taking an optimal medical therapy one year after experiencing a myocardial infarction (MI).[97] There is little to suggest why there may be a difference in adherence to treatment between males and females.

TABLE 4.1: MEDICATION ADHERENCE BY STUDY AND TREATMENT

Study	Country	Study type	Number of patients	Disease area	% adherent male	% adherent female
Gender differences in first-episode psychosis at 5-year follow-up – two different courses of disease? Results from the OPUS study at 5-year follow-up[95]	Denmark	RCT	547	Psychosis	81%	87%
Patient characteristics associated with medication adherence[96]	USA	Retrospective health system analysis	31 636	Hypertension	70.5%	68.8%
				Depression	59.7%	58.0%
				Hyperlipidaemia	70.8%	67.7%
				Asthma/COPD	31.0%	32.2%
				Diabetes	54.9%	50.2%
				Osteoporosis	61.7%	65.1%
				Cancer	61.0%	69.2%
				MS	69.6%	76.6%
Age and gender differences in medical adherence after myocardial infarction: Women do not receive optimal treatment – The Netherlands claims database[97]	The Netherlands	Retrospective cohort study	59 534	MI	51%	44%

4.2 METHODS

4.2.1 FOLLOW-UP IN HPS AND REVEAL

In HPS, participants were invited to follow-up visits in clinic at four, eight and twelve months post randomisation and six-monthly thereafter.[3] At each follow-up visit in HPS, information was collected about any suspected heart attacks, strokes, vascular procedures, cancers, or any other serious adverse events (SAEs).[3] Volunteers who were unwilling/unable to attend clinics in person were contacted via telephone (or alternatively follow-up was maintained via their GP), but the statin/placebo tablet had to be stopped if blood samples could not be taken for monitoring.[98] Participants were asked for information on any serious adverse events and their current statin use. Additionally, lipid profile assays were sought from 1500 randomly selected participants.[75]

In REVEAL, participants were invited to follow-up visits at two and six months post randomisation and then six monthly until the end of the study.[5] Similarly to HPS, information was collected on all SAEs (including study outcomes and reasons for hospitalization), any non-serious adverse events resulting in discontinuation of study treatment, and any symptoms of muscle pain or weakness or suggestive of hepatitis (loss of appetite, nausea, jaundice, lethargy, or malaise).[5] Again, participants could switch to a non-direct method of follow-up (telephone follow-up or follow-up via their GP), but the study treatments had to be stopped if blood samples could not be taken for safety monitoring.[99] Both trials had the possibility of "early recall" visits for participants who required any extra review.[3,99]

4.2.2 ADHERENCE TO TREATMENT IN HPS AND REVEAL

Adherence to study treatment was assessed at each trial follow-up visit. In HPS, participants were asked what the approximate percentages of their scheduled study treatments they had taken, ≥90%, 89-80%, 79-10%, or <10%.[3] Participants were classified as being adherent to the study treatments if they indicated they had taken >80% of their tablets/capsules. 80% is a commonly used threshold in health research.[96] If a participant stopped taking their treatment, the reasons for doing so were recorded. In REVEAL, similar criteria were used with participants being asked had they taken if they had taken "all" (≥90%), "most" (89-80%), "some" (79-10%) or "none" (<10%) of their atorvastatin and their anacetrapib or matching placebo tablets.[5] Similarly to HPS, participants in REVEAL were classed as adherent to their study treatments if they reported taking >80% of their tablets.

4.2.3 STATISTICAL METHODS

The percentage of patients attending an in-clinic follow-up visit at each time point was calculated as the proportion of those still remaining in the trial (not withdrawn from the trial or dead at study follow-up visit). Adherence to treatment at each time point was calculated as a percentage of those attending each study visit. This means that the denominator is different for these figures as patients who switched away from in-clinic follow-up visits were no longer taking their treatments and were therefore not included.

In order assess the potential associations between demographic factors, namely sex, and attendance at follow-up, a logistic regression analysis was performed. Models were adjusted for age, sex, socioeconomic status (assessed using Townsend Index), and ethnicity. The results from the logistic regression analyses are reported as Odds Ratios (OR) and 95% Wald confidence intervals. Analysis of deviance was used to select the

best model, and all interactions were considered. All analyses were performed in SAS 9.3 (SAS Institute, Cary NC). Results were considered to be statistically significant when $p<0.05$.

Similarly, logistic regression was used to assess the relationship between sex and adherence to treatment. Models were adjusted for age, sex, socioeconomic status, and ethnicity. The results from the logistic regression analyses are reported as odds ratios and 95% Wald confidence intervals and as before, results were considered to be statistically significant when $p<0.05$.

4.3 RESULTS

4.3.1 METHOD OF FOLLOW-UP

A follow-up visit in HPS could be conducted in one of several ways; in-clinic follow-up, in hospital follow-up, at home follow-up, telephone follow-up, GP follow-up, and telephone and hospital follow-up. The proportion of participants attending a clinic follow-up at each time point is shown in Figure 4.1. More male participants attended their follow-up visits in clinic. The difference between the sexes occurs from the first follow-up visit (four months after randomisation) and continues throughout the follow-up period reaching a 5% difference after five years of follow-up.

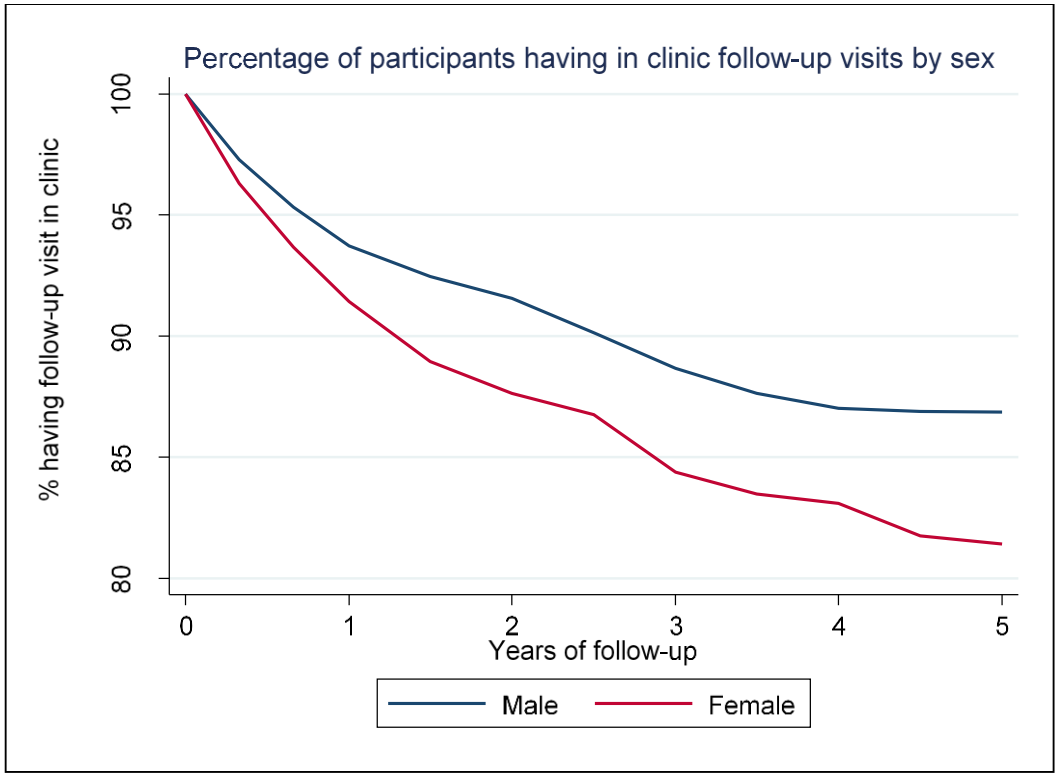


FIGURE 4.1: PERCENTAGE OF IN CLINIC FOLLOW-UP VISITS BY SEX (HPS)

To quantify this relationship, and to determine whether the differences observed are significantly different when adjusted for age, socioeconomic status, and ethnicity, logistic regression analyses were performed. At the first follow-up visit (four months after randomisation), the factors that were significantly associated with having an in-person clinic visit were sex and socioeconomic status (Table 4.2). Females were significantly less likely than males to have an in-person follow-up (OR: 0.73, 95% CI: 0.61-0.87). Additionally, those from more deprived areas were less likely to have an in-person follow-up visit with those in the intermediate tertile having an odds ratio of 0.77 (95% CI: 0.63-0.94) and those in the most deprived tertile having an odds ratio of 0.66 (95% CI: 0.53-0.81) compared to those in the more advantaged tertile.

TABLE 4.2: FACTORS AFFECTING TYPE OF FIRST FOLLOW-UP VISIT (HPS)

	'In-person' follow-up visit - visit 1		
	Est OR	OR 95% CI	p-value
Sex			
Male	-	-	-
Female	0.73	0.61-0.87	<0.001
Age Group			
<50	-	-	-
≥50 <60	0.97	0.67-1.40	0.86
≥60 <70	0.75	0.53-1.06	0.10
≥70	0.72	0.51-1.04	0.07
SES[+]			
Advantaged	-	-	-
Intermediate	0.77	0.63-0.94	0.01
Deprived	0.66	0.53-0.81	<0.001
Ethnicity			
White	-	-	-
Indian/Pakistani	1.11	0.49-2.53	0.80
West Indian/Guyanese	0.75	0.41-1.34	0.33
Other	0.74	0.27-2.03	0.56

[+]*SES – socioeconomic status*

At the midpoint of the follow-up period (two and a half years after randomisation), the factors that were significantly associated with having an in-person clinic visit were sex and socioeconomic status (Table 4.3). One of the age groups was also significantly associated with the type of follow-up visit.

Females were significantly less likely than males to have an in-person follow-up (OR: 0.73, 95% CI: 0.66-0.81). Additionally, those from more deprived areas were less likely to have an in-person follow-up visit with those in the intermediate tertile having an odds ratio of 0.81 (95% CI: 0.73-0.91) and those in the most deprived tertile having an odds ratio of 0.64 (95% CI: 0.57-0.72) compared to those in the more advantaged tertile. Those in the ≥50 <60 age group were significantly more likely to have an in-person follow-up visit at the midpoint (OR: 1.29, 95% CI: 1.06-1.56). There were no significant differences in type of follow-up for ethnicity.

TABLE 4.3: FACTORS AFFECTING TYPE OF FOLLOW-UP VISIT AT MIDPOINT (HPS)

	'In-person' follow-up visit – midpoint visit		
	Est OR	OR 95% CI	p-value
Sex			
Male	-	-	-
Female	0.73	0.66-0.81	<0.001
Age Group			
<50	-	-	-
≥50 <60	1.29	1.06-1.56	0.01
≥60 <70	1.12	0.94-1.34	0.20
≥70	0.90	0.75-1.08	0.27
SES[+]			
Advantaged	-	-	-
Intermediate	0.81	0.73-0.91	<0.001
Deprived	0.64	0.57-0.72	<0.001
Ethnicity			
White	-	-	-
Indian/Pakistani	0.87	0.56-1.35	0.54
West Indian/Guyanese	0.80	0.57-1.13	0.21
Other	1.36	0.66-2.82	0.41

[+]*SES – socioeconomic status*

After five years of follow-up, the factors that were significantly associated with having an in-person clinic visit were sex and socioeconomic status (Table 4.4). Females were significantly less likely than males to have an in-person follow-up (OR: 0.68, 95% CI: 0.61-0.76). Those from more deprived areas were also less likely to have an in-person follow-up visit with those in the intermediate tertile having an odds ratio of 0.78 (95% CI: 0.70-0.87) and those in the most deprived tertile having an odds ratio of 0.65 (95% CI: 0.57-0.73) compared to those in the more advantaged tertile. There were no significant differences in type of follow-up for ethnicity, but those in the oldest age group ≥70 were significantly less likely to have an in-person follow-up at five years than those in the youngest age group <50 (OR: 0.71, 95% CI: 0.59-0.87).

TABLE 4.4: FACTORS AFFECTING TYPE OF FOLLOW-UP VISIT AT FIVE YEAR FOLLOW-UP (HPS)

		'In-person' follow-up visit – five years post randomisation		
		Est OR	OR 95% CI	p-value
Sex				
	Male	-	-	-
	Female	0.68	0.61-0.76	<0.001
Age Group				
	<50	-	-	-
	≥50 <60	1.11	0.91-1.36	0.29
	≥60 <70	0.99	0.82-1.19	0.91
	≥70	0.71	0.59-0.87	0.001
SES[+]				
	Advantaged	-	-	-
	Intermediate	0.78	0.70-0.87	<0.001
	Deprived	0.65	0.57-0.73	<0.001
Ethnicity				
	White	-	-	-
	Indian/Pakistani	1.08	0.62-1.88	0.78
	West Indian/Guyanese	0.82	0.56-1.20	0.31
	Other	1.28	0.61-2.68	0.52

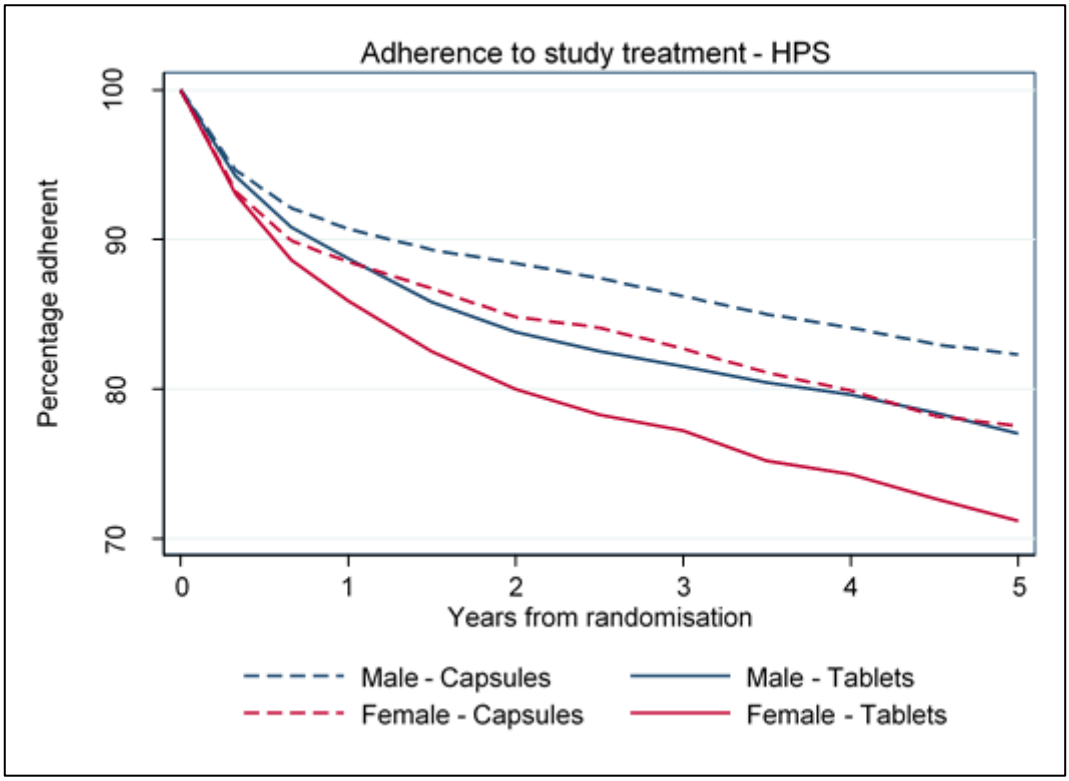
[+]*SES – socioeconomic status*

Similar data for the REVEAL trial were not available as the trial is still in post-trial follow-up and so these analyses could not be performed for this trial.

4.3.2 ADHERENCE TO TREATMENT

The mean duration of follow-up in HPS was five years for all randomised volunteers. Adherence to study treatment was assessed at each follow-up visit and was defined as at least 80% of the scheduled doses having been taken since the last follow-up visit. Females were less likely to be adherent to either study treatment (simvastatin/placebo tablet and vitamin/placebo capsules) than males (Figure 4.2). At the first follow-up visit at four months post randomisation, adherence was similar for both tablets and capsules with 93% of females and 94% of males being compliant. By one-year post randomisation, 86% of females were taking their tablets compared to 89% of males, and

89% of females were taking their capsules compared to 91% of males. After five years of treatment, 71% of females were still adherent to their allocated tablets compared to 77% of males, and 76% of females were still taking their capsules compared to 82% of males.



Tablets=simvastatin or matching placebo, capsules=antioxidant vitamins or matching placebo
Adherent to treatment = taking ≥80% of study treatments since last visit

FIGURE 4.2: ADHERENCE TO STUDY TREATMENT BY SEX (HPS)

The most common reasons for stopping treatment were "Patient wishes to stop (other than to start statin)" and "Patient unable or unwilling to attend clinics" and these differences were not due to differences in reported side effects.[3] There were no apparent differences between the treatment groups in any reported side-effects or in any biochemical abnormalities.[3] Across all study visits, females were more likely than males to stop taking their study treatments due to patient wishes (other than to start a

statin), (34.3% vs. 29.3%) and were more likely to stop due to problems attending clinics (16.8% vs. 12.1%).

In order to determine whether the differences observed are significantly different when adjusted for age, socioeconomic status, and ethnicity, logistic regression analyses were performed. At the first follow-up visit (four months after randomisation), the factors that were significantly associated with adherence to treatment were sex, socioeconomic status, and ethnicity (Table 4.5). Females were significantly less likely than males to report being adherent to their study treatments (statin/placebo OR: 0.81, 95% CI: 0.71-0.92, vitamin/placebo OR: 0.78, 95% CI: 0.68-0.89). Those from more deprived areas were less likely to have been adherent to their treatments with those in the most deprived tertile having an odds ratio of 0.80 (95% CI: 0.69-0.93) for their statin/placebo tablets and an odds ratio of 0.78 (95% CI: 0.67-0.91) for their vitamin/placebo capsules compared to those in the more advantaged tertile. Additionally, those with either Indian/Pakistani or West Indian/Guyanese ethnicity were less likely to have been adherent either of their treatments.

TABLE 4.5: FACTORS AFFECTING ADHERENCE AT FIRST FOLLOW-UP VISIT (HPS)

	Adherence to treatment - first follow-up visit (HPS)					
	Statin/placebo tablets			Vitamin/placebo capsules		
	Est OR	OR 95% CI	p-value	Est OR	OR 95% CI	p-value
Sex						
Male	-	-	-	-	-	-
Female	0.81	0.71-0.92	<0.001	0.78	0.68-0.89	<0.001
Age Group						
<50	-	-	-	-	-	-
≥50 <60	0.99	0.78-1.26	0.91	1.03	0.80-1.32	0.84
≥60 <70	0.96	0.76-1.21	0.74	0.96	0.75-1.22	0.75
≥70	0.98	0.77-1.25	0.89	0.97	0.75-1.24	0.78
SES[+]						
Advantaged	-	-	-	-	-	-
Intermediate	0.95	0.82-1.09	0.45	0.94	0.82-1.09	0.43
Deprived	0.80	0.69-0.93	0.003	0.78	0.67-0.91	<0.001
Ethnicity						
White	-	-	-	-	-	-
Indian/Pakistani	0.44	0.29-0.68	<0.001	0.43	0.28-0.67	<0.001
West Indian/Guyanese	0.56	0.38-0.81	0.002	0.54	0.37-0.80	0.002
Other	0.98	0.49-2.96	0.69	1.09	0.44-2.70	0.846

[+]*SES – socioeconomic status*

At the midpoint of the follow-up period (two and a half years after randomisation), the factors that were significantly associated with being adherent to the statin/placebo treatment were sex, socioeconomic status, and ethnicity (Table 4.6). These were also significantly associated with adherence to the vitamin/placebo capsules, along with being in the oldest age group (≥70). Females were significantly less likely than males to report being adherent to their study treatments (statin/placebo OR: 0.78, 95% CI: 0.72-0.85, vitamin/placebo OR: 0.77, 95% CI: 0.70-0.85). Those from more deprived areas were less likely to have been adherent to their treatments with those in the most deprived tertile having an odds ratio of 0.80 (95% CI: 0.73-0.87) for their statin/placebo tablets. A similar difference was observed in the vitamin/placebo capsules. Additionally, those with either Indian/Pakistani or West Indian/Guyanese ethnicity were

less likely to have been adherent either of their treatments, and those in the oldest age group were less adherent to their vitamin/placebo capsules compared to the youngest age group (OR: 0.62-0.88, 95% CI: 0.62-0.88).

TABLE 4.6: FACTORS AFFECTING ADHERENCE AT MIDPOINT FOLLOW-UP VISIT (HPS)

	Adherence to treatment - midpoint follow-up visit (HPS)					
	Statin/placebo tablets			**Vitamin/placebo capsules**		
	Est OR	OR 95% CI	p-value	Est OR	OR 95% CI	p-value
Sex						
Male	-	-	-	-	-	-
Female	0.78	0.72-0.85	<0.001	0.77	0.70-0.85	<0.001
Age Group						
<50	-	-	-	-	-	-
≥50 <60	0.97	0.83-1.13	0.66	0.93	0.78-1.12	0.44
≥60 <70	0.95	0.82-1.09	0.46	0.85	0.71-1.01	0.06
≥70	0.93	0.80-1.08	0.32	0.74	0.62-0.88	0.001
SES[+]						
Advantaged	-	-	-	-	-	-
Intermediate	0.90	0.83-0.99	0.02	0.89	0.81-0.99	0.03
Deprived	0.80	0.73-0.87	<0.001	0.77	0.69-0.85	<0.001
Ethnicity						
White	-	-	-	-	-	-
Indian/Pakistani	0.64	0.46-0.89	0.01	0.56	0.39-0.81	0.002
West Indian/Guyanese	0.67	0.51-0.87	0.003	0.59	0.43-0.78	<0.001
Other	0.80	0.49-1.31	0.38	0.81	0.46-1.41	0.45

[+]*SES – socioeconomic status*

After five years of follow-up, the factors that were significantly associated with being adherent to the statin/placebo treatment were sex and socioeconomic status (Table 4.7). Females were significantly less likely than males to still be adherent to their statin/placebo treatment (OR: 0.74, 95% CI: 0.68-0.81). Those from more deprived areas were also less likely to have been adherent with those in the intermediate tertile having an odds ratio of 0.87 (95% CI: 0.80-0.96) and those in the most deprived tertile having an odds ratio of 0.77 (95% CI: 0.70-0.85) compared to those in the more advantaged tertile.

There were no significant differences in adherence to the tablets for age, but those of Indian/Pakistani ethnicity were less likely to have been adherent than those with white ethnicity (OR:0.61, 95% CI: 0.45-0.82). It should be noted that there were very small numbers of non-white participants in this study. Similar differences were observed for the vitamin/placebo capsules.

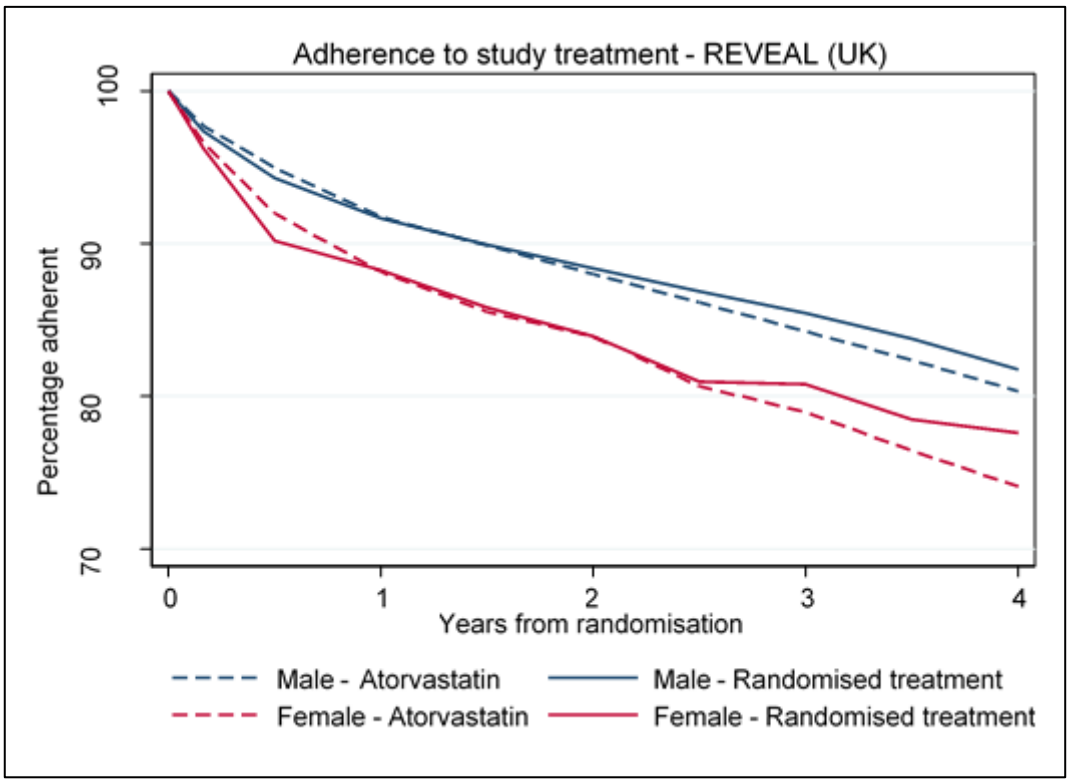
TABLE 4.7: FACTORS AFFECTING ADHERENCE AT FIVE YEAR FOLLOW-UP VISIT (HPS)

	Adherence to treatment – five-year follow-up visit (HPS)					
	Statin/placebo tablets			**Vitamin/placebo capsules**		
	Est OR	OR 95% CI	p-value	Est OR	OR 95% CI	p-value
Sex						
Male	-	-	-	-	-	-
Female	0.74	0.68-0.81	<0.001	0.75	0.68-0.83	<0.001
Age Group						
<50	-	-	-	-	-	-
≥50 <60	1.08	0.92-1.27	0.35	1.07	0.90-1.28	0.44
≥60 <70	1.03	0.89-1.20	0.71	0.96	0.81-1.14	0.64
≥70	0.91	0.78-1.07	0.25	0.80	0.67-0.96	0.01
SES[+]						
Advantaged	-	-	-	-	-	-
Intermediate	0.87	0.80-0.96	0.003	0.84	0.76-0.93	0.001
Deprived	0.77	0.70-0.85	<0.001	0.75	0.67-0.83	<0.001
Ethnicity						
White	-	-	-	-	-	-
Indian/Pakistani	0.83	0.54-1.27	0.39	0.68	0.44-1.06	0.09
West Indian/Guyanese	0.61	0.45-0.82	0.001	0.53	0.39-0.72	<0.001
Other	0.78	0.46-1.31	0.35	0.89	0.49-1.61	0.70

[+]*SES – socioeconomic status*

The mean duration of follow-up in REVEAL was four years for all randomised volunteers. Adherence to study treatments (atorvastatin and anacetrapib/placebo) was assessed at each follow-up visit and as in HPS, was defined as at least 80% of the scheduled study medications having been taken since the last follow-up visit. Females were less likely to have adhered to either study treatments than males (Figure 4.3). At the first follow-up visit at two months post randomisation, adherence was similar for

both treatments with 96% of females being compliant, and 96% of males. By one-year post randomisation, 88% of females were taking their treatments compared to 92% of males, and after four years of treatment, 77% of females were still adherent to their allocated randomised treatment compared to 82% of males, and 74% of females were still taking their atorvastatin compared to 80% of males.



Adherent to treatment = taking ≥80% of study treatments since last visit

FIGURE 4.3: ADHERENCE TO STUDY TREATMENT BY SEX (REVEAL - UK)

In order to determine whether the differences observed are significantly different when adjusted for age, socioeconomic status, and ethnicity, logistic regression analyses were performed. At the first follow-up visit (two months after randomisation), the only factors that were significantly associated with adherence to treatment were age and ethnicity (Table 4.8). Those in the oldest age group ≥70 were less likely to have been adherent to their treatments compared to those in the youngest age group <60 with an odds ratio of 0.63 (95% CI: 0.40-0.97) for their atorvastatin but no significant difference

in the anacetrapib/placebo treatment. There were no significant differences in adherence to either treatment for socioeconomic status at the first visit.

TABLE 4.8: FACTORS AFFECTING ADHERENCE AT FIRST FOLLOW-UP VISIT (REVEAL)

	Adherence to treatment - first follow-up visit (REVEAL)					
	Atorvastatin			Anacetrapib/placebo		
	Est OR	OR 95% CI	p-value	Est OR	OR 95% CI	p-value
Sex						
Male	-	-	-	-	-	-
Female	0.75	0.52-1.07	0.11	0.72	0.50-1.06	0.10
Age Group						
<60	-	-	-	-	-	-
≥60 <70	0.74	0.47-1.15	0.18	0.83	0.52-1.32	0.44
≥70	0.63	0.40-0.97	0.03	0.66	0.42-1.03	0.07
SES[+]						
Advantaged	-	-	-	-	-	-
Intermediate	0.71	0.48-1.07	0.10	0.66	0.42-1.03	0.07
Deprived	0.80	0.54-1.19	0.28	0.71	0.46-1.10	0.12
Ethnicity						
White	-	-	-	-	-	-
Black	*	*	*	*	*	*
Asian	0.44	0.20-0.96	0.04	0.54	0.22-1.35	0.19
Other	0.41	0.09-1.75	0.23	0.36	0.09-1.53	0.17

[+]*100% adherent – predicts success perfectly*
[+]*SES – socioeconomic status*

At the midpoint of the follow-up period (two years after randomisation), the factors that were significantly associated with being adherent to both treatments were sex and age (Table 4.9). Females were significantly less likely than males to report being adherent to their study treatments (atorvastatin OR: 0.73, 95% CI: 0.60-0.88, anacetrapib/placebo OR: 0.75, 95% CI: 0.62-0.91). Additionally, those in the oldest age group (≥70) were less likely to have been adherent to either treatment than those in the youngest age group (<60) with an odds ratio of 0.54 (95% CI: 0.43-0.68) for their atorvastatin and an odds ratio of 0.48 (95% CI: 0.38-0.60) for their anacetrapib/placebo treatment. There

were no significant differences in adherence to treatment for socioeconomic status or ethnicity.

TABLE 4.9: FACTORS AFFECTING ADHERENCE AT MIDPOINT FOLLOW-UP VISIT (REVEAL)

	Adherence to treatment - midpoint follow-up visit (REVEAL)					
	Atorvastatin			**Anacetrapib/placebo**		
	Est OR	OR 95% CI	p-value	Est OR	OR 95% CI	p-value
Sex						
Male	-	-	-	-	-	-
Female	0.73	0.60-0.88	0.001	0.75	0.62-0.91	0.003
Age Group						
<60	-	-	-	-	-	-
≥60 <70	0.84	0.67-1.06	0.14	0.73	0.57-0.92	0.008
≥70	0.54	0.43-0.68	<0.001	0.48	0.38-0.60	<0.001
SES[+]						
Advantaged	-	-	-	-	-	-
Intermediate	1.06	0.87-1.29	0.58	0.94	0.77-1.15	0.56
Deprived	1.10	0.91-1.32	0.34	0.98	0.81-1.19	0.82
Ethnicity						
White	-	-	-	-	-	-
Black	0.76	0.17-3.45	0.72	0.77	0.17-3.52	0.74
Asian	1.01	0.57-1.81	0.96	1.02	0.57-1.83	0.94
Other	1.13	0.34-3.73	0.84	0.63	0.24-1.67	0.36

[+]*SES – socioeconomic status*

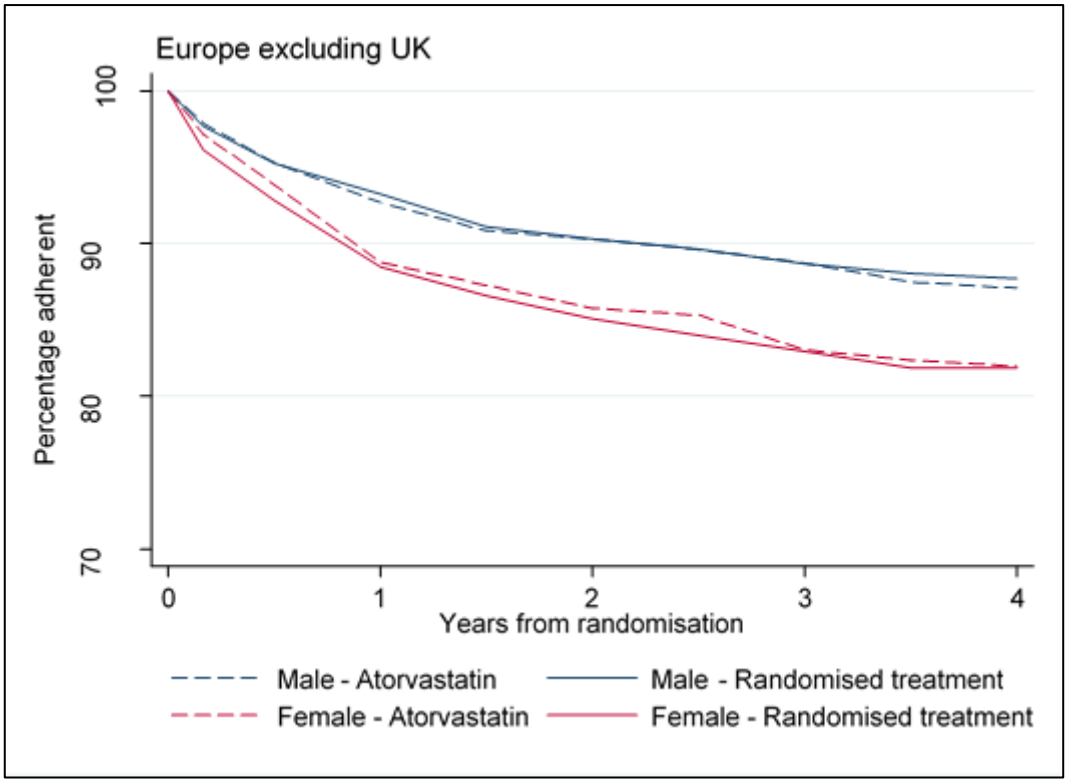
After four years of follow-up, the factors that were significantly associated with being adherent to either treatment were sex and age (Table 4.10). Females were significantly less likely than males to still be adherent to both their atorvastatin treatment (OR: 0.76, 95% CI: 0.60-0.94) and their anacetrapib/placebo treatment (OR: 0.67, 95% CI: 0.55-0.84). Participants in the oldest age group (>70) were significantly likely to be adherent to either treatment than those in the youngest age group (atorvastatin OR: 0.56, 95% Ci: 0.44-0.72, anacetrapib/placebo OR: 0.49, 95% CI: 0.39-0.63). There were no significant differences in adherence to either treatment for socioeconomic status or ethnicity.

TABLE 4.10: FACTORS AFFECTING ADHERENCE AT FOUR YEAR FOLLOW-UP VISIT (REVEAL)

	Adherence to treatment - midpoint follow-up visit (REVEAL)					
	Atorvastatin			Anacetrapib/placebo		
	Est OR	OR 95% CI	p-value	Est OR	OR 95% CI	p-value
Sex						
Male	-	-	-	-	-	-
Female	0.76	0.60-0.94	0.01	0.67	0.55-0.84	<0.001
Age Group						
<60	-	-	-	-	-	-
≥60 <70	0.93	0.72-1.20	0.59	0.81	0.63-1.04	0.09
≥70	0.56	0.44-0.72	<0.001	0.49	0.39-0.63	<0.001
SES[+]						
Advantaged	-	-	-	-	-	-
Intermediate	1.00	0.79-1.27	0.99	0.98	0.78-1.24	0.89
Deprived	1.03	0.83-1.29	0.76	1.03	0.83-1.28	0.76
Ethnicity						
White	-	-	-	-	-	-
Black	0.55	0.11-2.90	0.48	0.63	0.12-3.34	0.59
Asian	0.61	0.34-1.07	0.09	0.61	0.35-1.06	0.08
Other	3.32	0.44-25.01	0.25	1.75	0.40-7.66	0.46

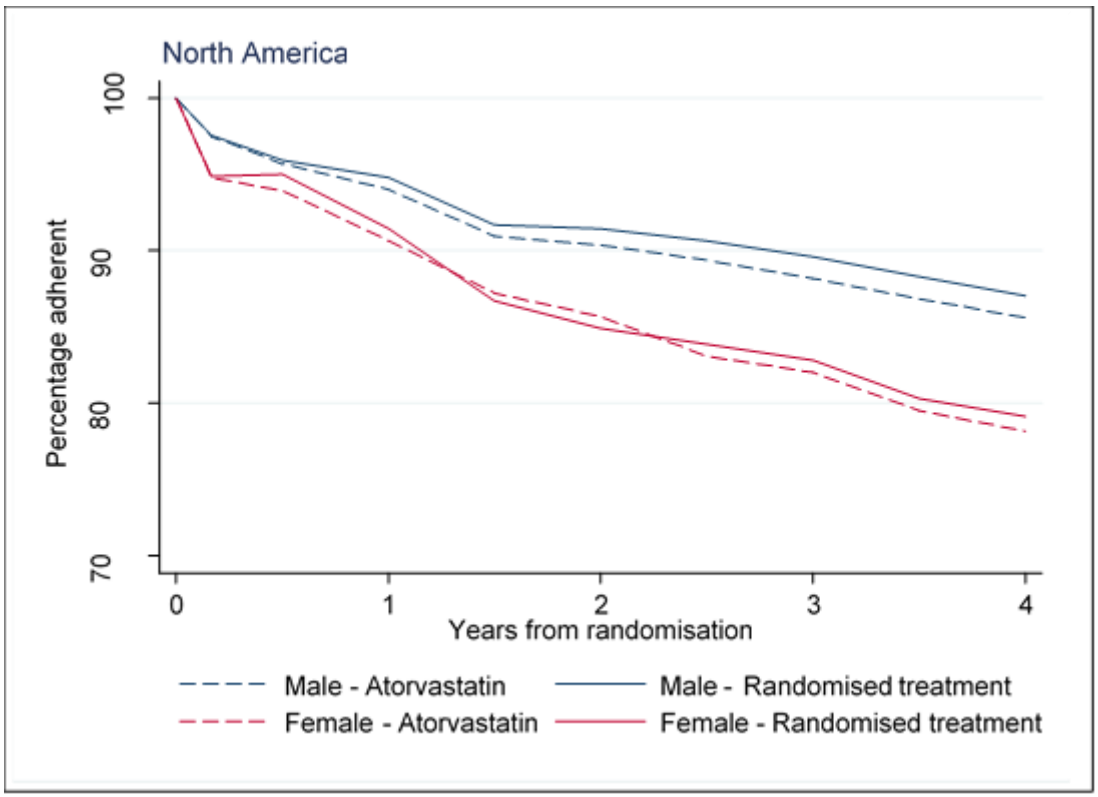
[+]*SES – socioeconomic status*

Adherence to study treatment by sex was also explored for each of the three other geographical regions in REVEAL (Europe excluding the UK, China, and North America). Female participants had lower adherence to both atorvastatin and randomised anacetrapib/placebo treatment in each geographic region, although the overall rates of treatment adherence varied by region (Figures 4.4, 4.5 & 4.6).



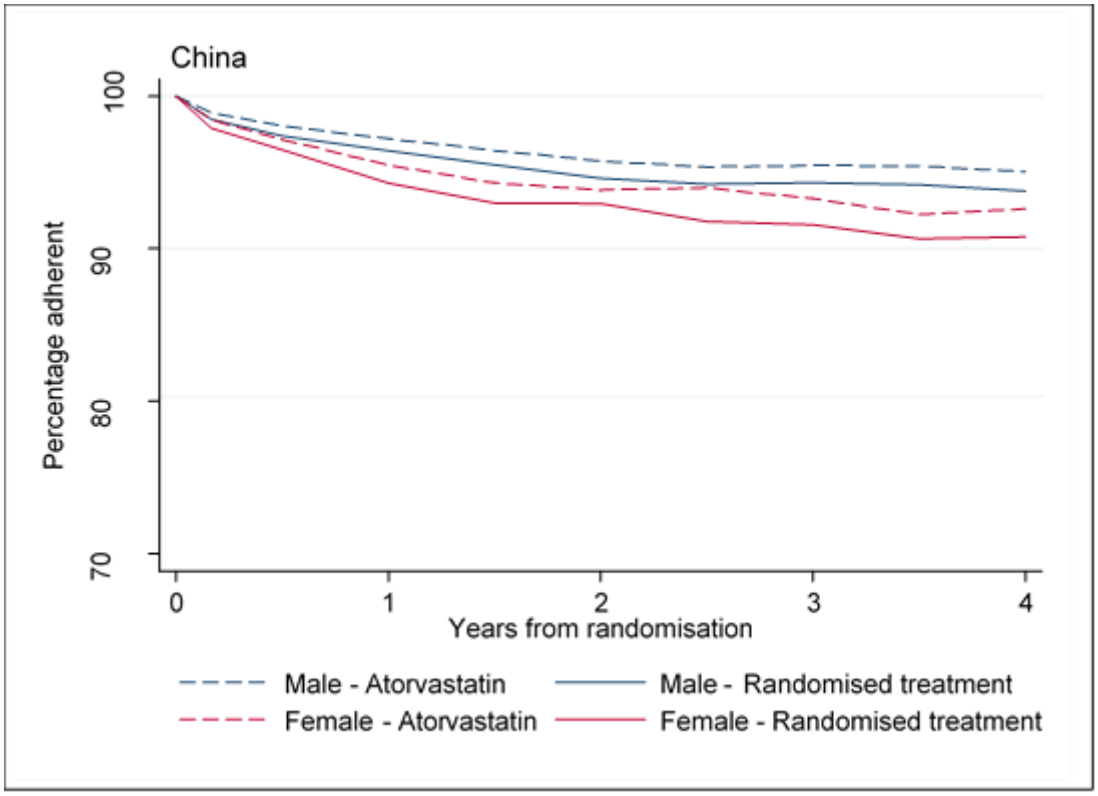
Adherent to treatment = taking ≥80% of study treatments since last visit

FIGURE 4.4: ADHERENCE TO STUDY TREATMENT BY SEX (REVEAL - EUROPE EXCLUDING UK)



Adherent to treatment = taking ≥80% of study treatments since last visit

FIGURE 4.5: ADHERENCE TO STUDY TREATMENT BY SEX (REVEAL - NORTH AMERICA)



Adherent to treatment = taking ≥80% of study treatments since last visit

FIGURE 4.6: ADHERENCE TO STUDY TREATMENT BY SEX (REVEAL - CHINA)

4.4 DISCUSSION

4.4.1 FINDINGS

Chapter 3 demonstrated that there are significant differences in sex throughout the recruitment process. In line with this, the work shown in this chapter establishes that there is a significant difference between males and females in both follow-up method and adherence to treatment. In HPS, females were less likely than males to attend a follow-up clinic in person ("direct follow-up"). As participants who switched away from clinic/hospital visits had to stop taking their study treatment for safety reasons, this means that there were fewer female participants taking their treatments. In both HPS and REVEAL, female participants were also less likely than males to be adherent to their treatments. In HPS, females were less likely to be adherent to their vitamin or

matching placebo capsules and even less likely to be adherent to their statin or matching placebo tablet. As adherence to treatment at each time point was calculated as a percentage of those taking their treatments, it is likely that the sex difference observed in HPS underestimates the true effect as fewer women were taking their treatments at all as they had switched to non-direct follow-up. In REVEAL, a similar sex difference was observed with fewer females being adherent to the study treatments. There was a smaller difference between the treatments though with similar numbers being adherent to their atorvastatin and their anacetrapib or matching placebo.

There was also a difference observed in socioeconomic status, with those from more deprived areas being less likely to attend a clinic visit in person. In both trials, all participants were offered travel expenses, so it is unlikely that this is the reason for switching to non-direct follow-up, although could have been a contributing factor. Capturing more information on reasons for switching to telephone or GP follow-up in future trials may help to clarify this.

4.4.2 LIMITATIONS

There are a few limitations to this research. Firstly, the data for type of follow-up were unavailable for REVEAL. This means that we are unable to assess whether the differences in follow-up method by sex have changed over time. Secondly, in HPS participants were not asked why they wanted to switch to a non-direct method of follow-up, and in both trials very little data was captured on the reasons for non-adherence to the study treatments. The recorded reason “patient wishes” was the only way to differentiate between those leaving for reported side effects and those leaving for other reasons. There are also issues with the reliability of self-reported adherence to treatment, and there is a potential for reporting bias. However, the levels of cholesterol

and their sub-fractions in the blood were checked against the adherence reporting, and in general in both studies the blood lipid levels were in keeping with the adherence reported in the trials. Another limitation is that there was not enough information collected to investigate whether treatment burden was a factor affecting adherence to treatment. It has been established that adherence is lower in those taking more medications and in those with more comorbidities.[96] Unfortunately the data were not available to determine if this was a factor in HPS. A further limitation is that the data collected during the trial did not capture whether patients stopped taking their treatments due to perceived muscle pain. More females than males reported muscle pain (35.3% vs. 31.6%) but this was not linked to adherence or stopping reasons. It could be hypothesised that due to a smaller body mass, females experienced a greater rate of side effects which may explain the differences in adherence. However, there were no apparent differences between the treatment groups (statin vs. placebo) in any reported side-effects or in any biochemical abnormalities[3] so this may not be the case.

Finally, the differences observed in adherence to treatment may be specific to certain disease types; a study of patients with a first-episode psychosis in the schizophrenia spectrum found that females were more likely than males to be adherent to their medications.[95] On the other hand, in keeping with these data, three studies that used pharmacy data to assess medication adherence found that females were less likely than males to take their prescribed regimen for a variety of health problems including hypertension, hyperlipidaemia and diabetes. [96,97,100]

The time involved in attending study visits, particularly over many years, can be off-putting and may be the reason that more female participants switch to non-direct follow-up or stop taking their tablets. Another reason may be that they are taking more medications, or that they experience more comorbidity. One measure of this was

captured on the screening form (Appendix I) under other medical history. Patients were asked if they had experienced an MI, ischaemic stroke, TIA or angina, whether they had been hospitalised for angina, had a revascularisation procedure, had been diagnosed with diabetes, currently had hypertension treated with antihypertensive drugs, or were currently being treated with an oral anticoagulant. A similar proportion of eligible males and females reported these events (98% females vs. 99% males) implying a similar level of reported comorbidity (Appendix III). The average number of other medical events was also similar in male and female participants (2.23 in males, 2.16 in females). A further measure of comorbidity was the proportion of participants experiencing a major vascular event (MVE). In HPS, approximately 29% of male participants experienced an MVE compared to 18% of females. This indicates that males may be more comorbid than females and so does not explain the differences in adherence. Future studies may benefit from capturing further information on the reasons for discontinuing treatment.

4.4.3 CONCLUSIONS

In summary, female participants were less likely to attend clinic follow-ups and were less likely to be adherent to their study treatments. This was the case in both HPS and REVEAL. This causes concern as Chapter 3 detailed the difficulties in recruiting female participants to cardiovascular disease trials and having fewer females taking part and even fewer being adherent to their treatments could leave studies underpowered for sex specific effects. These factors potentially limit the ability to establish the safety and efficacy of interventions in females and have the potential to have serious adverse public health consequences.

This chapter has looked at the sex differences observed in-trial, both in type of follow-up and in adherence to study treatment. Fewer females attend their visits in person, and fewer females are adherent to their treatments. The next chapter will explore the literature around using electronic health records for clinical trial follow-up. If electronic health records can be used as a method of follow-up, the number of follow-up visits may be able to be reduced or even completely removed, which may encourage more females to take part in these trials.

5 FOLLOW-UP IN CLINICAL TRIALS USING ROUTINE DATA

5.1 INTRODUCTION

This chapter includes a review of the existing literature of studies that have used electronic health records for the follow-up of participants in clinical trials. This will summarise the current knowledge and experience in this field and provide some information about the use of routine data prior to the investigation of using electronic health records for follow-up in the Heart Protection Study (HPS) in Chapter 6.

5.1.1 FOLLOW-UP IN RANDOMISED CONTROLLED TRIALS

The reliable detection of moderate treatment effects requires the follow-up of large numbers of suitable patients in randomised controlled trials (RCTs).[15] In order to determine efficacy and assess clinical outcomes, reliable follow-up information is needed.[80] In an ideal world, all participants would maintain follow-up until the end of the trial, however this is rarely the case. This can cause issues and introduce bias. Attrition bias is caused by an unequal loss of participants in a randomised controlled trial. The differences between those leaving the trial and those continuing can be the cause of an observed effect and not the intervention, particularly when these differences occur between arms of the trial.[13]

There are many reasons why participants are lost to follow-up, for example, participants may move address, they might withdraw due to ill health, or they may lose interest in participating in the research. Deciding that they are not interested in attending follow-up visits any more can be caused by many factors but is more likely to occur when

individual follow-up appointments are lengthy, frequent, or the trial continues for an extended period of time.

The use of routine datasets and large databases is not new in health research. Many epidemiological studies use routine data to capture outcomes. There are large studies that have used mortality data,[101] disease registries,[102] health records,[103] and demographic databases.[104]

An advantage of using routinely collected data is that it is often much cheaper than collecting data for each study. There are many datasets available to researchers, and these can often be combined to provide a wealth of information, often more than could be collected via surveys or other follow-up methods. Another advantage of using routine data is the size of the data available. Routine datasets tend to cover a much larger population than data collected for study; they also tend to cover a greater time period allowing for longer periods of follow-up. This is beneficial as it can be a more cost-effective way to follow-up large numbers of participants, and trials must be adequately sized to reliably detect moderate treatment effects.[15]

One disadvantage of routinely collected data is that the data are not collected for the purposes of research. This means that the data quality can be variable, and the dataset may not include the most relevant variables for specific research questions. A further disadvantage is that the data may not be up to date or available in real time. While this may be less of a problem for epidemiological studies that can wait to receive the data, it may cause issues in clinical trials where real-time safety information may be needed to monitor the trial.

There has been lots of research into improving retention in clinical trials, with many strategies suggested. These mainly relate to methods for improving questionnaire

feedback, and not into encouraging attendance at follow-up visits.[14] If routine data could be used to reduce the number of follow-up visits needed, or to remove the need for in-person follow-up altogether, then it may encourage more patients to participate in clinical trials as it would be less time-consuming for them. It also has the potential to make trials more efficient by reducing the costs for the study sponsors.

5.1.2 OBJECTIVES

The objectives of this review are to identify trials where routine data have been used for follow-up in clinical trials, to describe which sources of data have been used and in which disease areas, and to summarise the literature to provide an understanding of how these data are currently used.

5.2 METHODS

This review was undertaken using the guidelines for conducting and reporting a systematic review in the PRISMA (Preferred Reporting for Systematic Reviews and Meta-Analyses) statement.[19]

This review is looking at the methodology used for follow-up in these trials therefore there were no age, size, or disease restrictions. All types of intervention were included along with all outcomes. In order to be included in this review, the study had to be an RCT that had used a form of routine data for the follow-up of at least one outcome for the trial. The routine data had to have been used during the trial for follow-up, and not done as a retrospective review.

5.2.1 SEARCH STRATEGY

All studies published in English before January 2018 were included in this review. The search strategy used to search the electronic databases was for any of the abstracts to contain the following words and phrases: ("clinical trial" or "randomised controlled trial" or "randomised" or "randomisation" or "randomized controlled trial" or "randomized" or "randomization") and ("electronic health record*" or "EHR" or "electronic record*" or "electronic medical record*" or "EMR" or "routine data") and ("follow-up" or "follow up"). This search strategy was used to identify potentially eligible abstracts from MEDLINE, EMBASE, and Cochrane Review electronic databases. Abstracts of papers identified were reviewed to determine whether they met the criteria to be included in this review; those that did not meet the criteria were excluded. The full text of the remaining papers was then examined to determine whether the papers met the criteria for inclusion and to extract any relevant data from them. The following data were identified as relevant for this review:

- Type of routine data used
- Disease area of clinical trial
- Follow-up strategy used (just routine data or in combination with others)
- Sample size

The relevant details of the RCT needed to be specified in order to be included within this review. These data are summarised in a qualitative review about the use of routine data for follow-up in clinical trials. There are no meta-analyses performed given the methodological issues with the broadness of the results.

5.3 RESULTS

There were 74 studies identified by the search strategy. Figure 5.1 contains the flowchart listing the process of selection for this review, from the studies identified to those included in the final review. Of these 74 studies, 51 were excluded after reading the full text. The reasons for exclusion are listed in Table 5.1. There were 28 papers excluded as they were conference abstracts or letters with no full text available, seven papers that were excluded as they were not an RCT, ten that were excluded as routine data had not been used for follow-up, five excluded as they were a retrospective analysis of a current trial, and one excluded as there were no details of the RCT given.

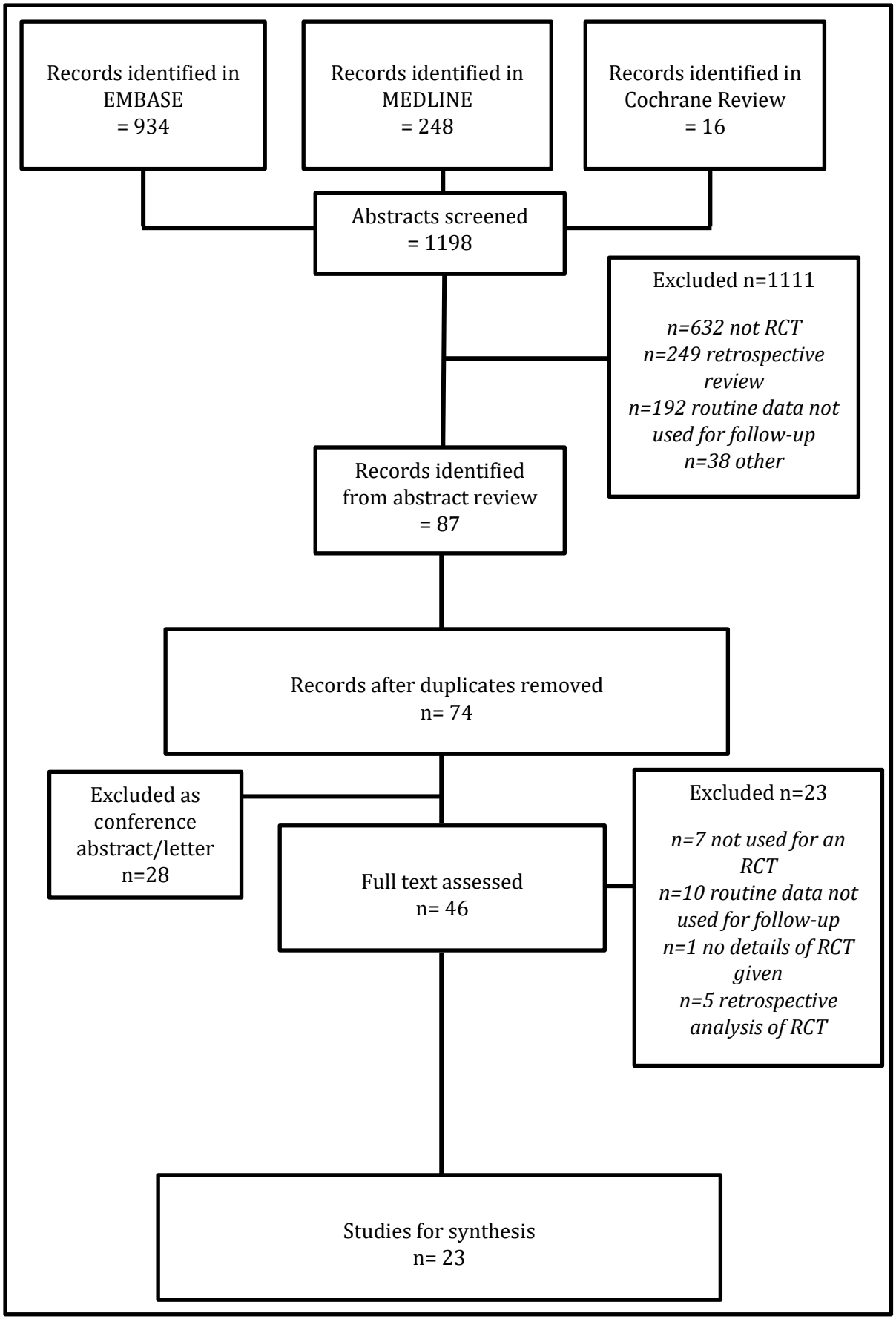


FIGURE 5.1: FLOWCHART SHOWING NUMBER OF STUDIES AT EACH STAGE OF THE REVIEW

TABLE 5.1: STUDIES EXCLUDED FROM THE REVIEW AFTER FULL TEXT REVIEW

Title	Authors	Year	Reason for exclusion
Can a 10-minute injury prevention program decrease injuries in military cadets? A jump-acl study[105]	Beutler et al.	2013	Conference abstract
A randomized controlled trial of peer mentoring and financial incentive to improve glucose control in African American Veterans[106]	Long et al.	2011	Conference abstract
Translation of diabetes interventions into community settings: 12-month results of the lifestyle intervention for the treatment of diabetes trial[107]	Bertoni et al.	2017	Conference abstract
The diabetes telephone study: Design and challenges of a pragmatic cluster randomized trial to improve diabetic peripheral neuropathy treatment[108]	Adams et al.	2015	Conference abstract
Individualized risk communication and lay outreach for the primary prevention of cardiovascular disease in community health centers: Preliminary results of a randomized controlled trial[109]	Shah et al.	2014	Conference abstract
Preliminary results of the study MEETinCY[110]	Lambrinou et al.	2015	Conference abstract
Mortality rate and associated factors in older adults following hip fracture[111]	Maher et al.	2014	Conference abstract
Can the use of computer decision support system prevent complicated ulcer among patients treated with nsaid or aspirin? A randomised controlled trial in general practice[112]	Petersen et al.	2017	Conference abstract
Randomized trial evaluating a coordinated survivorship care program for early stage breast cancer[113]	Ruddy et al.	2013	Conference abstract
The effectiveness of a tailored intervention to increase follow-up adherence in patients with glaucoma[114]	Tran et al.	2014	Conference abstract
Phase II study of cold snare polypectomy for superficial non-ampullary duodenal epithelial tumor (D-CSP trial)[115]	Takizawa et al.	2017	Conference abstract
Automated intervention with stepped increases in support to increase uptake of colorectal cancer screening: A randomized trial[116]	Green et al.	2013	Conference abstract
Effect of early palliative care on health care costs in patients with metastatic NSCLC[117]	Greer et al.	2012	Conference abstract

Title	Authors	Year	Reason for exclusion
Challenges in the design and conduct of controlled clinical effectiveness trials in schizophrenia[118]	Rosenheck et al.	2011	Conference abstract
Rare disease cohort using ambulatory electronic medical record (AEMR) data: Duchenne muscular dystrophy (DMD)[119]	Anderson et al.	2017	Conference abstract
Efficacy of clinical notifications to health care providers in management of heart failure patients[120]	Akhavein et al.	2015	Conference abstract
Depression treatment and diabetes risk: A 9-year follow-up study of the impact trial[121]	Khambaty & Callahan	2015	Conference abstract
Long-term impact of the mActive physical activity trial on cardiometabolic health[122]	Urrea et al.	2016	Conference abstract
A comparison of three system-based interventions to improve follow-up adherence in a primary eye care setting[123]	Kenney et al.	2014	Conference abstract
Design and methods for a pilot study and multicenter pragmatic trial comparing HCTZ and chlorthalidone for cardiovascular outcomes[124]	Margolis et al.	2016	Conference abstract
Long-term outcomes following a randomized study of oral antibiotics in primary sclerosing cholangitis[125]	Tabibian	2016	Conference abstract
Is nurse-led virtual monitoring of chronic hepatitis B patients a safe, feasible and cost-effective complement to standard care?[126]	Teo et al.	2016	Conference abstract
The effect of alteplase on health care utilisation and quality of life in Sweden, Norway and Scotland: Results from the 3rd international stroke trial (ist-3)[127]	Berge et al.	2017	Conference abstract
The effect of pharmaceutical care on three-year mortality a follow-up study[128]	Buus et al.	2013	Conference abstract
Ascertaining study participant safety using centralized electronic medical records in a clinical trial setting - lessons learned from the veterans affairs NEPHRON-D trial[129]	Huang et al.	2017	Conference abstract
Main outcomes from the relax trial of telephone delivered collaborative care for panic and generalized anxiety disorder[130]	Rollman et al.	2010	Conference abstract
Supporting clinical trials through healthcare informatics[131]	Jones et al.	2015	Conference abstract
A1C and weight outcomes at 18 months in patients with type 2 diabetes treated with exenatide in an ambulatory care setting[132]	McAdam-Marx et al.	2009	Letter

Title	Authors	Year	Reason for exclusion
The ADAPTABLE trial and PCORnet: Shining light on a new research paradigm[133]	Hernandez et al.	2015	No details of trial
Synthesizing lessons learned from get with the guidelines: The value of disease-based registries in improving quality and outcomes[134]	Ellrodt et al.	2013	Not RCT
Better population health through behavior change in adults: A call to action[135]	Spring et al.	2013	Not RCT
Introduction to a supplement on innovative approaches to studying health outcomes in rare diseases[136]	Kesselheim	2014	Not RCT
The (pre)hypertension limbo: Is it time to lower the treatment bar?[137]	Towfighi	2011	Not RCT
The research-to-practice gap in mood disorders: A role for the U.S. Department of Veterans Affairs[138]	Kilbourne& Valenstein	2007	Not RCT
Commentary on vickers[139]	McCowan	2014	Not RCT
Innovative assessment of inpatient and pulmonary drug costs for children with cystic fibrosis[140]	Levy & Rosenberg	2016	Not RCT
Healthy Aging 5 Years After a Period of Daily Supplementation With Antioxidant Nutrients: A Post Hoc Analysis of the French Randomized Trial SU.VI.MAX[141]	Assmann et al.	2015	Post hoc analysis
Long-Term Effect of Participation in an Early Exercise and Education Program on Clinical Outcomes and Cost Implications, in Patients with TIA and Minor, Non-Disabling Stroke[142]	Faulkner et al.	2017	Retrospective
Increased mortality among patients taking digoxin - Analysis from the AFFIRM study[143]	Whitbeck et al.	2013	Retrospective
A Healthy Weight for Toddlers? Two-Year Follow-up of a Randomized Controlled Trial of Group Well-Child Care[144]	Shah & Fenick	2016	Retrospective
Effects of diabetes self-management programs on time-to-hospitalization among patients with type 2 diabetes: A survival analysis model[145]	Adepoju et al.	2014	Retrospective
What is the effect of a formalised trauma tertiary survey procedure on missed injury rates in multi-trauma patients? Study protocol for a randomised controlled trial[146]	Keijzers et al.	2015	Routine data not used for follow-up

Title	Authors	Year	Reason for exclusion
Evaluation of a risk-stratification strategy to improve primary care for low back pain: the MATCH cluster randomized trial protocol[147]	Cherkin et al.	2016	Routine data not used for follow-up
Cost-effectiveness of home telemonitoring in chronic kidney disease patients at different stages by a pragmatic randomized controlled trial (eNephro): rationale and study design[148]	Thilly et al.	2017	Routine data not used for follow-up
Telephone-Delivered Stepped Collaborative Care for Treating Anxiety in Primary Care: A Randomized Controlled Trial[149]	Rollman et al.	2016	Routine data not used for follow-up
The opportunities and challenges of pragmatic point-of-care randomised trials using routinely collected electronic records: evaluations of two exemplar trials[63]	van Staa et al.	2014	Routine data not used for follow-up
Effectiveness of Computer Automation for the Diagnosis and Management of Childhood Type 2 Diabetes: A Randomized Clinical Trial.[150]	Hannon et al.	2017	Routine data not used for follow-up
Use of behavioral economics and social psychology to improve treatment of acute respiratory infections (BEARI): Rationale and design of a cluster randomized controlled trial [1RC4AG039115-01] - study protocol and baseline practice and provider characteristics[151]	Persell et al.	2013	Routine data not used for follow-up
A clinical decision-support system with interactive alerts improved CD4 cell count in HIV[152]	Robbins & Lester	2013	Routine data not used for follow-up
Outpatient blood pressure monitoring using bi-directional text messaging[153]	Anthony et al.	2015	Routine data not used for follow-up
Sustained Blood Pressure Control Following Discontinuation of a Pharmacist Intervention for Veterans[154]	Carter et al.	2015	Routine data not used for follow-up

There were 23 studies that were included in this review. The details of these are listed in Table 5.2, along with the key information about the type of routine data used, the disease area, other follow-up methods used, the sample size, and the duration of the follow-up.

Ten of the trials (37%) used routine data as the sole means of follow-up for the trials. Others used other forms of follow-up including telephone follow-up, mailed surveys, in-person follow-up (using questionnaires or surveys), pharmacy data, billing records, and mortality data. The disease area of the trials varied considerably, and the sample size ranged from 68 to 88 150. Length of follow-up was also very variable, with the shortest follow-up period being 1 week and the longest being 10 years. All of the studies are from 2008 onwards; as with the recruitment review in Chapter 2, this is not due to a restriction on the dates used for the search, but likely due to the feasibility of accessing electronic records prior to this date. This review found trials from a variety of countries. The majority (n=14) were conducted in the USA, four were conducted in the UK, two in The Netherlands, and one in Canada, one in Chile, and one in Finland.

TABLE 5.2: STUDIES INCLUDED IN THE REVIEW

Title	Country	Year	Type of routine data	Area of RCT	Other follow-up methods used?	Sample size	Length of follow-up
Effectiveness of the head CT choice decision aid in parents of children with minor head trauma: study protocol for a multicenter randomized trial.[155]	USA	2014	Medical records	Trauma	Used alongside telephone follow-up	1004	1 week
Effect of Collaborative Care for Depression on Risk of Cardiovascular Events: Data From the IMPACT Randomized Controlled Trial[156]	USA	2014	Medical records and insurance claim databases, death certificates	Cardiovascular disease & mental health	Alone	235	8 years
Are benefits from diabetes self-management education sustained?[157]	USA	2013	Electronic health records	Diabetes	With mailed surveys	623	1 year
Improving Chronic Disease Care by Adding Laypersons to the Primary Care Team: A Parallel Randomized Trial[158]	USA	2013	Electronic health records	Cardiovascular disease & diabetes	With survey	2135	1 year
Long-Term Safety and Efficacy of Lowering Low-Density Lipoprotein Cholesterol With Statin Therapy 20-Year Follow-Up of West of Scotland Coronary Prevention Study[159]	UK	2016	Electronic health records	Cardiovascular disease	Alone	6595	5 years (in-trial) 20 years (post-trial)

Title	Country	Year	Type of routine data	Area of RCT	Other follow-up methods used?	Sample size	Length of follow-up
Using the 4 pillars™ practice transformation program to increase adult influenza vaccination and reduce missed opportunities in a randomized cluster trial[160]	USA	2016	EMR database	Vaccination	Alone	70549	2 years
Challenges in the design and conduct of controlled clinical effectiveness trials in schizophrenia[118]	USA	2011	VA Healthcare system	Psychiatric	With interview	382	2 years
The Iowa Continuity of Care study: Background and methods[161]	USA	2008	Inpatient records	Pharmacy	With pharmacy data, surveys, billing records	1000	90 days
Long-term safety and efficacy of antithymocyte globulin induction: use of integrated national registry data to achieve ten-year follow-up of 10-10 Study participants[162]	USA	2015	Transplant registry	Transplantation	Alone	183	10 years
Effectiveness of the Chest Pain Choice decision aid in emergency department patients with low-risk chest pain: study protocol for a multicenter randomized trial[163]	USA	2014	Health records	Cardiovascular disease	With phone call	844	45 days
Design of a cluster-randomized trial of electronic health record-based tools to address overweight and obesity in primary care[164]	USA	2015	Health records	Obesity	Alone	88150	12 months

Title	Country	Year	Type of routine data	Area of RCT	Other follow-up methods used?	Sample size	Length of follow-up
Effects of a patient oriented decision aid for prioritising treatment goals in diabetes: pragmatic randomised controlled trial[165]	The Netherlands	2014	Health records	Diabetes	With patient surveys	344	4 months
Electronic health record-based patient identification and individualized mailed outreach for primary cardiovascular disease prevention: a cluster randomized trial.[166]	USA	2013	Electronic health records	Cardiovascular disease	Alone	435	9 months
A Randomized Controlled Effectiveness Trial for PSA Screening Decision Support Interventions in Two Primary Care Settings.[167]	USA	2015	Electronic health records	Cancer screening	With survey	2001	1 year
The Heart failure and Optimal Outcomes from Pharmacy Study (HOOPS): Rationale, design, and baseline characteristics[168]	UK	2011	NHS discharges (Scotland) with death records	Cardiovascular disease	Alone	2169	5 years
Mixed methods evaluation of targeted case finding for cardiovascular disease prevention using a stepped wedged cluster RCT[169]	UK	2012	Medical records	Cardiovascular disease	Alone	7247	1 year

Title	Country	Year	Type of routine data	Area of RCT	Other follow-up methods used?	Sample size	Length of follow-up
A Randomized Trial Comparing Ceramic-on-Ceramic Bearing vs Ceramic-on-Crossfire-Polyethylene Bearing Surfaces in Total Hip Arthroplasty[170]	Canada	2016	Electronic health records	Orthopaedic surgery	Along with questionnaires	68	10 years
Strategies for increasing mammography screening in primary care in Chile: Results of a randomized clinical trial[171]	Chile	2010	Electronic health records	Cancer screening	Along with survey	500	6 months
Patient-centered physical activity coaching in COPD (Walk On!): A study protocol for a pragmatic randomized controlled trial[172]	USA	2016	Electronic health records	COPD	With survey	1650	1 year
Effectiveness of a Proactive Primary Care Program on Preserving Daily Functioning of Older People: A Cluster Randomized Controlled Trial[173]	The Netherlands	2016	Electronic health records	Daily functioning	With questionnaires	3092	1 year
Population-based outreach versus care as usual to prevent suicide attempt: study protocol for a randomized controlled trial.[65]	USA	2016	Electronic health records & insurance claim databases	Suicide	Alone	19500	18 months

Title	Country	Year	Type of routine data	Area of RCT	Other follow-up methods used?	Sample size	Length of follow-up
Feasibility cluster randomised controlled trial of a within-consultation intervention to reduce antibiotic prescribing for children presenting to primary care with acute respiratory tract infection and cough[174]	UK	2017	Medical notes	Respiratory	With questionnaires	501	30 days
Impact of improved recording of work-relatedness in primary care visits at occupational health services on sickness absences: study protocol for a randomised controlled trial[175]	Finland	2017	Occupational health records	Occupational health	Alone	50400	1 year

5.3.1 TYPE OF CLINICAL TRIAL

The trials included in this review looked at a wide range of diseases and treatments. Seven of the trials were looking at cardiovascular disease outcomes,[156,158,159,163,166,168,169] three were looking at diabetes outcomes,[157,158,165] three were looking at psychiatric outcomes,[65,118,156] two were looking at cancer-related outcomes,[167,171] and two were looking at respiratory outcomes.[172,174] There were also studies that looked at trauma,[155] vaccinations,[160] pharmacy related outcomes,[161] transplantation,[162] obesity,[164] orthopaedics,[170] daily functioning,[173] and occupational health.[175] These are not mutually exclusive as some of the trials looked at multiple outcomes. Three of the trials were investigating the effects of new drugs.[118,159,162] However, they were all looking at the long term outcomes after the in-trial period had ended, and participants in these trials had received follow-up visits during the trial.

5.3.2 TYPE OF FOLLOW-UP

Ten of the trials (37%) used routine data as the sole means of follow-up for the trials. The majority of the trials used routine data alongside other means of follow-up. Two trials used telephone follow-up alongside routine data[155,163], ten used interviews and questionnaires[118,158,161,165,167,170–174], one used pharmacy data[161], one used billing records[161], and one used mailed questionnaires[157]. These are not mutually exclusive as some trials used multiple methods of follow-up.

5.3.3 SAMPLE SIZE

There was a large range in the sample size of the trials in included in this review. The smallest trial had 68 participants[170], whereas the largest trial recruited 88 150 participants.[164] The median sample size was 1004 with an interquartile range of

468 - 4844. The largest studies (those recruiting >10 000 participants) all used routine data as the only form of follow-up.

5.3.4 LENGTH OF FOLLOW-UP

The length of follow-up varied a lot between the trials included in this review. The smallest trials had a short follow-up period of 7 days[155], and the longest two trials followed participants up for 10 years.[162,170] The median length of follow-up was one year (interquartile range, 202.5 – 543.7 days).

5.4 DISCUSSION

5.4.1 FINDINGS

This review demonstrates that electronic health records can be used for follow-up in randomised controlled trials in a variety of countries and disease areas. This review has identified different types of routine data that have been used to follow-up participants in randomised controlled trials. The most common data source was electronic health records. These ranged from inpatient records to discharge records and included both primary and secondary care records. Accessing electronic medical records allows for more patients to be followed up, and for a longer period of time as it is a more cost-effective method and requires fewer members of staff. Using routine databases is also likely to be the most cost-effective way to follow-up large numbers of participants, and for longer periods of time.

5.4.2 LIMITATIONS

As with the recruitment review, there were difficulties finding adequate search terms. Often the MESH terms are assigned do not relate to the follow-up methodology, and

many trials do not report the information about the follow-up methods in detail. The search did not exclude trials that used registry data or mortality data, but very few of these were picked up, despite many of them existing. Many of the papers excluded from this review were excluded as the routine data was part of the intervention and not a method of follow-up, or because the study was not randomised. Unfortunately, there was no way to exclude these from the review using different search terms.

Another limitation is the inability to synthesize the results. A meta-analysis was not appropriate as this was a methodological review, and so no quantitative analyses are presented.

Some of the papers raised concerns about the accuracy of the reporting within electronic records, and whether the outcomes of interest are recorded. One study struggled to link the trial data with registries outside of the co-ordinating country[162] and others discussed the limitations when not all of the outcomes are recorded in the routine data sources.[164,170] However, one of the trials, the West of Scotland Coronary Prevention Study, has previously reported on the accuracy of electronic records capturing events compared to those reported and adjudicated during the trial.[176] They found that routinely collected data, in this case NHS hospital records for Scotland, could be used to record cardiovascular endpoints in clinical trials. They compared in trial events with events recorded in the routine data and found >80% of first events matched. Those that didn't match were either diagnosed as "non-cardiac unspecified chest pain" which could be a hospital miscoding or were due to the events occurring at a hospital outside of Scotland or a private hospital. This suggests that the events recorded in certain administrative databases may be accurate enough for clinical trials.

5.4.3 CONCLUSIONS

Routine data, particularly electronic health records can be used to follow-up participants in clinical trials, either solely or in conjunction with other follow-up methods. Electronic health records were used to follow-up patients in a variety of clinical trials across many different disease areas. They were generally used to follow-up larger clinical trials, and for longer periods of time.

These routine databases contain considerable amounts of data and are a relatively inexpensive way to follow-up large numbers of participants. In the era of 'big data', there is the potential for more clinical trials to adopt follow-up using routine data to maximise the follow-up periods in clinical trials, to make it easier for participants to take part in research, and to increase the sample sizes in trials so that moderate treatment effects may be observed.

As with the recruitment review, there is considerable room for improvement in the reporting of methodological techniques used for follow-up, and there are trials that have used electronic health records as a form of follow-up that were not picked up by the search.

This chapter has demonstrated that routine data and electronic health records can be used for follow-up in clinical trials. The next chapter will look to establish how accurate electronic health records are at capturing trial outcomes compared to in-trial reporting in HPS. It will also look to see whether there are any sex differences observed in the reporting of trial outcomes, and whether electronic health records could be used to follow participants up in cardiovascular disease trials.

6 FOLLOW-UP USING HOSPITAL EPISODE STATISTICS (HES)

6.1 INTRODUCTION

This chapter aims to investigate the use of Hospital Episode Statistics (HES) for follow-up in the Heart Protection Study (HPS). The recruitment methods used in HPS have been reported in Chapter 3, and this chapter describes how the trial outcomes were analysed and replicates this analysis using HES data. The accuracy of the HES data is discussed and an analysis by sex is presented to investigate whether the differences observed in recruitment (Chapter 3) and adherence to treatment (Chapter 4) are also observed in the reporting of outcomes during follow-up.

6.1.1 HOSPITAL EPISODE STATISTICS

HES data are available for all admissions to NHS hospitals in England and have been since 1989. It does not contain any information on hospital activity in Scotland, Wales or Northern Ireland which has to be applied for separately for each respective region. The data are collected for administrative purposes (reimbursement), not research. The HES datasets contain information on all admitted patient care (since 1989), hospital deliveries and births (since 1989), outpatient appointments (since 2003), and A&E attendances (since 2008). These attendances are linked by a HES ID (since 1997). This HES ID is a pseudonymised identifier that allows records to be linked to create a longitudinal record for an individual over many years, without disclosing any potential identifiers. The HES datasets are processed on a monthly basis, and a final annual HES

extract is produced at the end of the financial year, which cannot then be updated. A person can register with their GP practice to opt out, meaning that none of their confidential health data can leave NHS digital. This is estimated to affect 2.5% of HES inpatient episodes overall.[177]

A hospital admission is classed as any activity that takes up a hospital bed (including day-cases allocated a bed). The admitted patient care dataset contains information on diagnoses, procedures or operations, dates of admission and discharge, type of admission (elective, emergency, birth, delivery etc.), discharge destination (home, other hospital etc.), the hospital provider, geographic information, and costing information. It does not contain information on drugs provided or dispensed. Some of this information is derived from local patient administrative systems (postcodes etc.). Diagnostic and procedural codes are entered by clinical coders based on the patient care notes that were written by a clinician. Each row within the dataset represents one finished consultant episode (FCE), which is a period of care under one consultant. A stay in hospital is referred to as a spell, and a patient can have multiple FCEs within a spell.[178] Each FCE contains 20 diagnosis fields, including a primary diagnosis field. These are classified using the 10th International Classification of Disease codes (ICD-10). Each FCE can also contain up to 24 procedure codes, these are classified using Office of Population Censuses and Surveys: Classification of Interventions and Procedures version 4 (OPCS-4) codes. One operation can consist of many procedures. Not all episodes contain procedures, but each episode will contain a diagnosis code.

6.2 METHODS

HPS recruited participants from 1994-1997 and subsequently followed them up for a median of five years. Consent to access participants' hospital records was sought at the start of the study.

The original analysis dataset was used to verify the correct methods were being used. Participants from Scotland, Wales, and Northern Ireland were then removed from the dataset as records for these participants are not held in HES. Any outcome events that occurred prior to the linking of HES by HESID (pre-1997) were also removed as these would also not be captured within the HES dataset. The analyses were then performed on this dataset in order to provide a comparable set of events, in order to calculate rate ratios and 95% confidence intervals (95% CI) for the HES dataset. The agreement of these events with those reported in trial was than analysed, and these analyses were then repeated by sex.

The primary outcome in HPS was total mortality[74] and cause specific mortality was also a pre-specified outcome. The main analyses looked at Coronary Heart Disease (CHD) mortality, and non-CHD mortality. These were defined using the 9th International Classification of Disease codes (ICD9). Additionally, there were separate analyses of ten specific causes of death. The codes used for each category are presented in Table 6.1.

TABLE 6.1: ICD9 CATEGORY CODES (HPS)

Category	ICD9 Codes
CHD Mortality	410-414
Haemorrhagic Stroke	430-432
Other Stroke	433-438
Other Vascular	390-459
Neoplastic	140-239
Respiratory	460-519
Hepatic	570-576
Renal	580-593
Other Medical	000-799
Suicide	950-959
Other Non-Medical	800-999

Participants in HPS were flagged with the Office for National Statistics (ONS) and death certificates were obtained for participants who died during the study. Additionally, information was sought from GPs, and from hospital records and the cause of death was adjudicated by two clinicians, who determined the underlying cause of death ICD-9 code, and in cases where this was discrepant, a third clinician was involved for further discussion.[3] An analysis of adjudicated vs. non-adjudicated causes of death has already been presented.

The secondary outcomes in HPS included the first major vascular events (MVEs) in each arm of the trial. MVEs were defined to be coronary events (split into non-fatal myocardial infarction (MI) and coronary death), strokes, (split into fatal and non-fatal stroke), and revascularisations (split into coronary and non-coronary).[75] In order to assess whether the HES records accurately captured the trial outcomes, these categories were looked at separately and then combined with the adjudicated death records to produce an overall MVE analyses.

In HPS, the non-fatal events were recorded using Read codes whereas in HES records, events are coded using the 10th International Classification of Disease codes (ICD-10) and the Office of Population Censuses and Surveys: Classification of Interventions and

Procedures version 4 (OPCS-4) codes. The relevant codes for each of the non-fatal categories used in the analyses are presented in Table 6.2.

TABLE 6.2: NON-FATAL EVENT CODES (HPS)

Category	Read Codes	ICD10 Codes
Non-fatal MI	G30*, G3P*, G3Q*, G3R*, G3S*, G3T*	I21*, I22*
Stroke		
Haemorrhagic	G60, G60*-G61*	I60*, I61*, I62*
Non-haemorrhagic	G64	I63*
Revascularisation		
Coronary	792	K234, K4*, K50, K75
Non-coronary	7A1, 7A20, 7A22, 7A31, 7A32, 7A33, 7A41, 7A44, 7A46, 7A48, 7A4A, 7A54, 7L06, 7L07, 7L08, 7L0F, 7L0P, 7L0Q, 7L0X	L16*, L18*- L19*, L22*- L23*, L25*-L31* (not L264, L312), L37*-L39* (not L393, L394), L48*-L54* (not L543) L57-L63* (not L633, L634), L66*, L68*, X07*-X11*, X121

6.2.1 STATISTICAL METHODS

All comparisons are log-rank analyses of the first occurrence of the event after randomisation between those allocated simvastatin, and those allocated the placebo tablets (intention to treat analyses). Patients are censored at their final follow-up visit, date of death, or the end of the study (12th October 2001), whichever occurred first. A small number of participants had a final follow-up visit after the end of the study, and so were censored at that date. All analyses were performed in SAS 9.3 (SAS Institute, Cary NC). Kappa statistics are also presented and are interpreted using the interpretations provided by Altman with κ >0.81 indicating very good agreement, κ >0.61 indicating good agreement, κ >0.41 indicating moderate agreement, κ >0.21 indicating fair agreement, and κ <0.20 indicating poor agreement.[179] Results were

considered to be statistically significant when $p<0.05$. Forest plots were produced using the ONYX program in R version 2.10.1.

6.3 RESULTS

The results for the non-fatal events are presented individually and then combined with the fatal events for MVE at the end. There were 47 737 FCEs within the trial period, and 41 412 spells among 16 690 participants.

6.3.1 NON-FATAL MYOCARDIAL INFARCTION

In the main trial, allocation to simvastatin was shown to reduce the number the risk of experiencing a non-fatal myocardial infarction (MI). Using the main trial dataset, the published results were replicated with 931 events across the 20 536 randomised participants, and a reduction in the rate of first non-fatal MI of 35% (Table 6.3). Participants residing in Scotland, Wales, and Northern Ireland were removed from the trial dataset, leaving a total of 18 241 participants, and events that occurred prior to 1997 were also removed as these would not be captured in HES. Using these records, the reduction in the rate of first non-fatal MI is 38% (95% CI: 0.53-0.72). The HES records for these 18 241 participants were then used, and relevant ICD-10 codes were searched for. Two different definitions of MI were used; acute MI (ICD 21*) and any MI (ICD21*, ICD22*). These codes were searched for in the primary position and in any position. This provided estimates of the risk reduction of between 36-38%. Using the definition “Any MI” in any position within the HES record, the reduction in the rate of first non-fatal MI is 36% (95% CI: 0.56-0.74).

TABLE 6.3: NON-FATAL MI BY DATASET

	Type of event data	Simvastatin-allocated (n=10269)	Placebo-allocated (n=10267)	Event Rate Ratio	95% CI
Lancet Results:	Non-fatal MI - HPS	357 (3.5%)	574 (5.6%)	0.65	0.57-0.74
Remove Scot/Wales/N.Ire:		**Simvastatin-allocated (n=9127)**	**Placebo-allocated (n=9114)**		
Change pre-HES dates*	Non-fatal MI	247 (2.7%)	421 (4.6 %)	0.62	0.53-0.72
Using different definitions from HES◦	Acute MI+ (Primary)	192 (2.1%)	304 (3.3%)	0.62	0.52-0.74
	Acute MI+ (Any)	234 (2.6%)	362 (4.0%)	0.64	0.54-0.75
	Any MI++ (Primary)	277 (3.0%)	424 (4.6%)	0.64	0.55-0.75
	Any MI++ (Any)	313 (3.4%)	481 (5.3%)	0.64	0.56-0.74

+*Acute MI defined as ICD 21**
++*Any MI defined as ICD 21* or ICD 22**
**Change pre-HES dates – Removing events prior to 1997*
◦*Using only HES records from 1997 onwards*

Figure 6.1 illustrates the rate ratios and confidence intervals for each of the datasets. The HES records appear to provide a reasonable estimation for the overall results seen in the trial, with a similar risk reduction and only slightly wider confidence intervals.

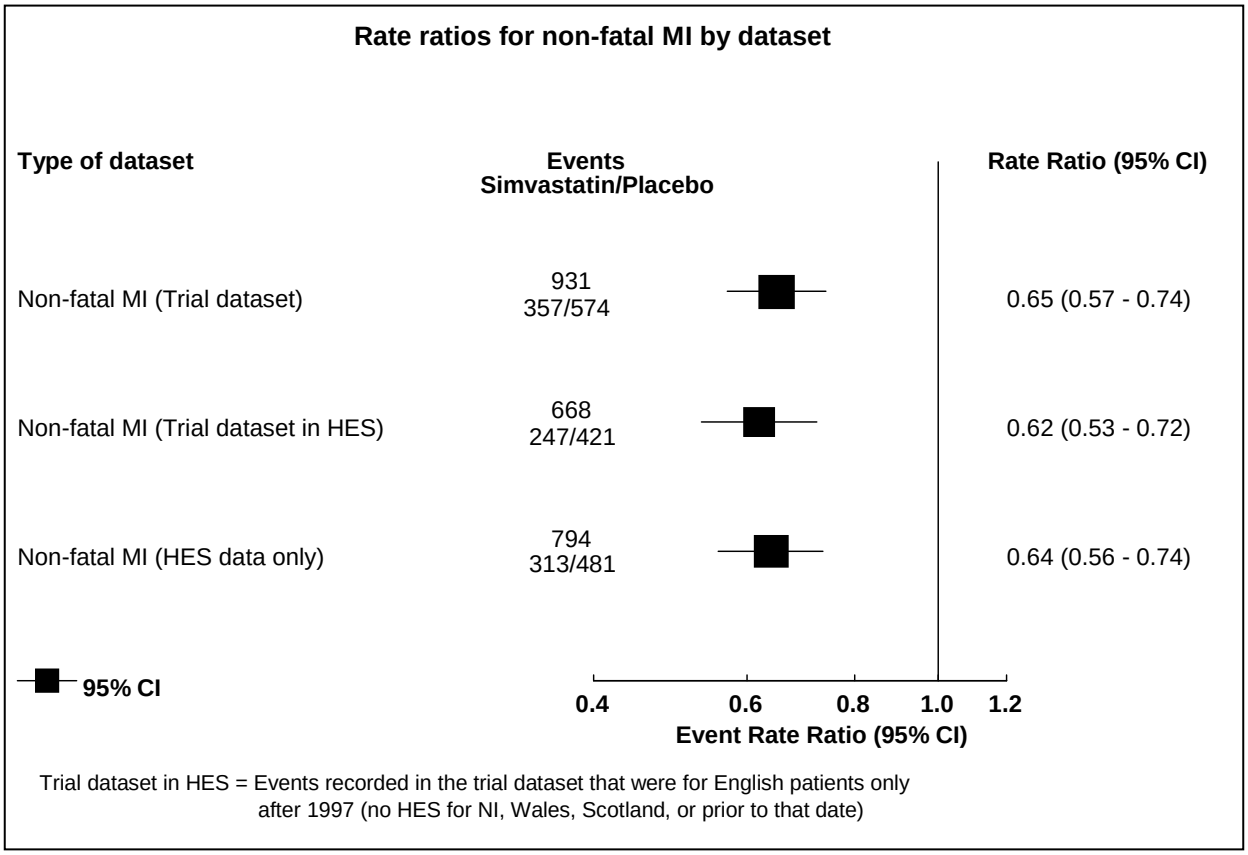


FIGURE 6.1: RATE RATIOS FOR NON-FATAL MI BY DATASET

6.3.2 NON-FATAL STROKE

In HPS, allocation to simvastatin was also shown to reduce the risk of experiencing a non-fatal stroke. Using the main trial dataset, the published results were replicated with 865 across the 20 536 randomised participants, and a reduction in the rate of first non-fatal stroke of 30% (Table 6.4). Participants residing in Scotland, Wales, and Northern Ireland were removed from the trial dataset, leaving a total of 18 241 participants and, as in the previous analyses, events that occurred prior to 1997 were also removed as these would not be captured in HES. Using these records, the reduction in the rate of first non-fatal stroke is 32% (95% CI: 0.59-0.80). The HES records for these 18 241 participants were then used, and relevant ICD-10 codes were searched for. These codes were searched for in the primary position and in any position. Using codes in the primary position, the risk of non-fatal stroke was shown to be 31% (95% CI: 0.58-0.82).

When relevant ICD-10 codes for stroke in any position within the HES record were used, the risk of non-fatal stroke was 26% (95% CI: 0.65-0.85). More non-fatal strokes were captured using the HES records than were captured during the trial.

TABLE 6.4: NON-FATAL STROKE BY DATASET

	Type of event data	Simvastatin-allocated (n=10269)	Placebo-allocated (n=10267)	Event Rate Ratio	95% CI
Lancet Results	Non-fatal Stroke	366 (3.6%)	499 (4.9%)	0.70	0.62-0.81
Remove Scot/Wales/N.Ire.		**Simvastatin-allocated (n=9127)**	**Placebo-allocated (n=9114)**		
Change pre-HES event dates*	Non-fatal stroke	273 (3.0%)	381 (4.9%)	0.68	0.59-0.80
Using different definitions from HES◦	Non-fatal stroke (Primary)	232 (2.5%)	331 (3.6%)	0.69	0.58-0.82
	Non-fatal stroke (Any)	392 (4.3%)	519 (5.7%)	0.74	0.65-0.85

**Change pre-HES dates – Removing events prior to 1997*
◦Using only HES records from 1997 onwards

The separate definitions of stroke (haemorrhagic and ischemic) were also investigated. There were 59 non-fatal haemorrhagic strokes (including sub-arachnoid haemorrhage) within the main trial and there was no apparent risk reduction with simvastatin allocation (Table 6.5). Removing participants in Scotland, Wales and Northern Ireland, and events prior to 1997, there were 47 events recorded (with no risk reduction). Using the primary position in HES there were 72 non-fatal events and using any position in HES there were 81 events recorded. Neither of these definitions produced statistically significant risk reductions with simvastatin.

TABLE 6.5: NON-FATAL HAEMORRHAGIC STROKE BY DATASET

	Type of event data	Simvastatin-allocated (n=10269)	Placebo-allocated (n=10267)	Event Rate Ratio	95% CI
Trial Results	Non-fatal haem stroke	28 (0.3%)	31 (0.3%)	0.87	0.52-1.50
Remove Scot/Wales/N.Ire.		**Simvastatin-allocated (n=9127)**	**Placebo-allocated (n=9114)**		
Change pre-HES event dates*	Non-fatal haem stroke	23 (0.2%)	24 (0.3%)	0.92	0.52-1.63
Using different definitions from HES◦	Non-fatal haem stroke (Primary)	34 (0.4%)	38 (0.4%)	0.88	0.56-1.40
	Non-fatal haem stroke (Any)	39 (0.4%)	42 (0.5%)	0.92	0.59-1.42

**Change pre-HES dates – Removing events prior to 1997*
◦Using only HES records from 1997 onwards

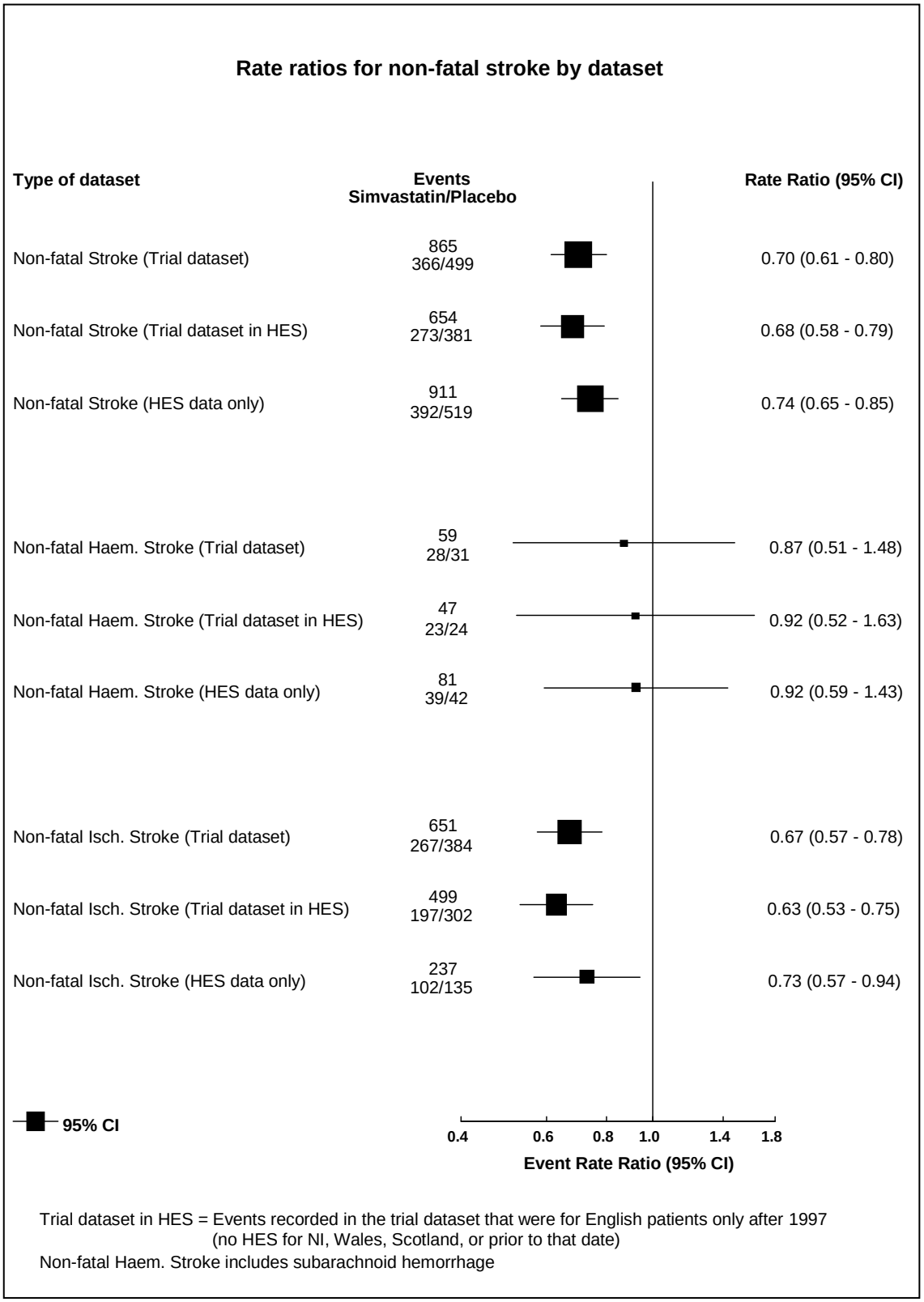
The results for ischaemic stroke are shown in Table 6.6. The number of non-fatal ischaemic strokes recorded in the trial was 651, with a risk reduction in the simvastatin arm of 33% (95% CI: 0.58-0.79). There were fewer ischaemic strokes recorded in HES records (using codes found in the primary position, or any position). Using codes for non-fatal ischaemic stroke in any position, there were 137 events with a risk reduction of 27% (95% CI: 0.57-.95). The difference in the number of events is likely to be due to the way in which ischaemic strokes are diagnosed and recorded in HES records.

TABLE 6.6: NON-FATAL ISCHAEMIC STROKE BY DATASET

	Type of event data	Simvastatin-allocated (n=10269)	Placebo-allocated (n=10267)	Event Rate Ratio	95% CI
Trial Results	Non-fatal Isch Stroke	267 (2.6%)	384 (3.8%)	0.68	0.58-0.79
Remove Scot/Wales/N.Ire:		**Simvastatin-allocated (n=9127)**	**Placebo-allocated (n=9114)**		
Change pre-HES event dates*	Non-fatal Isch stroke	197 (2.2%)	302 (3.8%)	0.67	0.58-0.79
Using different definitions from HES◦	Non-fatal Isch stroke (Primary)	90 (1.0%)	118 (1.3%)	0.77	0.59-1.01
	Non-fatal Isch stroke (Any)	102 (1.1%)	135 (1.5%)	0.73	0.57-0.95

**Change pre-HES dates – Removing events prior to 1997*
◦Using only HES records from 1997 onwards

Figure 6.2 shows the risk ratio and 95% confidence intervals for each dataset (using ICD10 codes recorded in any position in the HES record) for non-fatal stroke, and separately, non-fatal haemorrhagic stroke and non-fatal ischaemic stroke. In each case, the rate ratio appears to provide a good estimation for the rate ratio seen in the main trial, although with wider confidence intervals than seen with non-fatal MI. A large number of strokes reported in HES were classified as neither haemorrhagic or ischaemic (n=593). This is due to medical practice at the time and the lack of scans undertaken to categorise the stroke.



Unknown type of stroke recorded in HES, n=593. Rate Ratio (95% CI): 0.73(0.62-0.84)

FIGURE 6.2: RATE RATIOS FOR NON-FATAL STROKE BY DATASET

6.3.3 REVASCULARISATION

In the main trial, allocation to simvastatin was shown to reduce the number and risk of revascularisation (with a greater reduction seen in coronary revascularisations than non-coronary). Using the main trial dataset, the published results were replicated with 2144 procedures across the 20 536 randomised participants (Table 6.7) and a reduction in the rate of revascularisation procedure of 25% (95% CI: 0.69-0.82). Participants residing in Scotland, Wales, and Northern Ireland were removed from the trial dataset, leaving a total of 18 241 participants, and events that occurred prior to 1997 were also removed as these would not be captured in HES. Using these records, the reduction in the rate of first revascularisation is 26% (95% CI: 0.67-0.82). The HES records for these 18 241 participants were then used, and relevant OPCS-4 codes were searched for. This provided an estimate of the risk reduction 25% (95%CI: 0.68-0.83). There were considerably fewer revascularisation procedures reported in the HES dataset compared to the main trial dataset.

TABLE 6.7: REVASCULARISATION BY DATASET

	Type of event data	**Simvastatin-allocated (n=10269)**	**Placebo-allocated (n=10267)**	**Event Rate Ratio**	**95% CI**
Lancet Results:	All revasc	939 (9.1%)	1205 (11.7%)	0.76	0.70-0.83
Remove Scot/Wales/N.Ire:		**Simvastatin-allocated (n=9127)**	**Placebo-allocated (n=9114)**		
Change pre-HES event dates*	All revasc	676 (7.4%)	874 (9.6%)	0.74	0.67-0.82
Using HES◦	All revasc	665 (7.3%)	864 (9.5%)	0.75	0.68-0.83

**Change pre-HES dates – Removing events prior to 1997*
◦Using only HES records from 1997 onwards

The revascularisation procedures were then investigated by coronary and non-coronary procedures separately. In the main trial, there were 1238 coronary revascularisations with a risk reduction of 32% (95% CI: 0.61-0.77) in the simvastatin arm (Table 6.8). After removing participants who resided outside of England, and procedures that occurred prior to 1997, there were 894 coronary revascularisations (risk reduction: 31%, 95% CI: 0.60-0.79). Using OPCS-4 codes recorded in HES for this period, there were 827 procedures recorded, with a 28% risk reduction (95% CI: 0.63-0.82) in the simvastatin arm.

TABLE 6.8: CORONARY REVASCULARISATIONS BY DATASET

	Type of event data	**Simvastatin-allocated (n=10269)**	**Placebo-allocated (n=10267)**	**Event Rate Ratio**	**95% CI**
Lancet Results:	Coronary revasc	513 (5.0%)	725 (7.1%)	0.68	0.61-0.77
Remove Scot/Wales/N.Ire:		**Simvastatin-allocated (n=9127)**	**Placebo-allocated (n=9114)**		
Change pre-HES event dates*	Coronary revasc	373 (4.1%)	521 (5.7%)	0.69	0.60-0.79
Using HES◦	Coronary revasc	349 (3.8%)	478 (5.2%)	0.72	0.63-0.82

**Change pre-HES dates – Removing events prior to 1997*
◦Using only HES records from 1997 onwards

There were 982 non-coronary revascularisation procedures recorded in the main trial (Table 6.9), with a risk reduction of 19% (95% CI: 0.71-0.92). Using only procedures that would appear in HES, there were 721 procedures recorded (risk reduction: 22%, 95% CI: 0.67-0.90). Searching HES records for relevant OPCS-4 codes produced 751 procedures and a resulting risk reduction of 21% (95% CI: 0.68-0.91) in the simvastatin arm.

TABLE 6.9: NON-CORONARY REVASCULARISATIONS BY DATASET

	Type of event data	Simvastatin-allocated (n=10269)	Placebo-allocated (n=10267)	Event Rate Ratio	95% CI
Lancet Results:	Non-coronary revasc	450 (4.4%)	532 (5.2%)	0.85	0.75-0.96
Remove Scot/Wales/N.Ire:		**Simvastatin-allocated (n=9127)**	**Placebo-allocated (n=9114)**		
Change pre-HES dates*	Non-coronary revasc	324 (3.5%)	397 (4.4%)	0.78	0.67-0.90
Using HES◦	Non-coronary revasc	334 (3.7%)	417 (4.6%)	0.79	0.68-0.91

**Change pre-HES dates – Removing events prior to 1997*
◦Using only HES records from 1997 onwards

Figure 6.3 shows the risk ratio and 95% confidence intervals for each dataset for any revascularisation procedure, and separately, coronary revascularisations and non-coronary revascularisations. In each case, the rate ratio appears to provide a good estimation for the rate ratio seen in the main trial, although with wider confidence intervals.

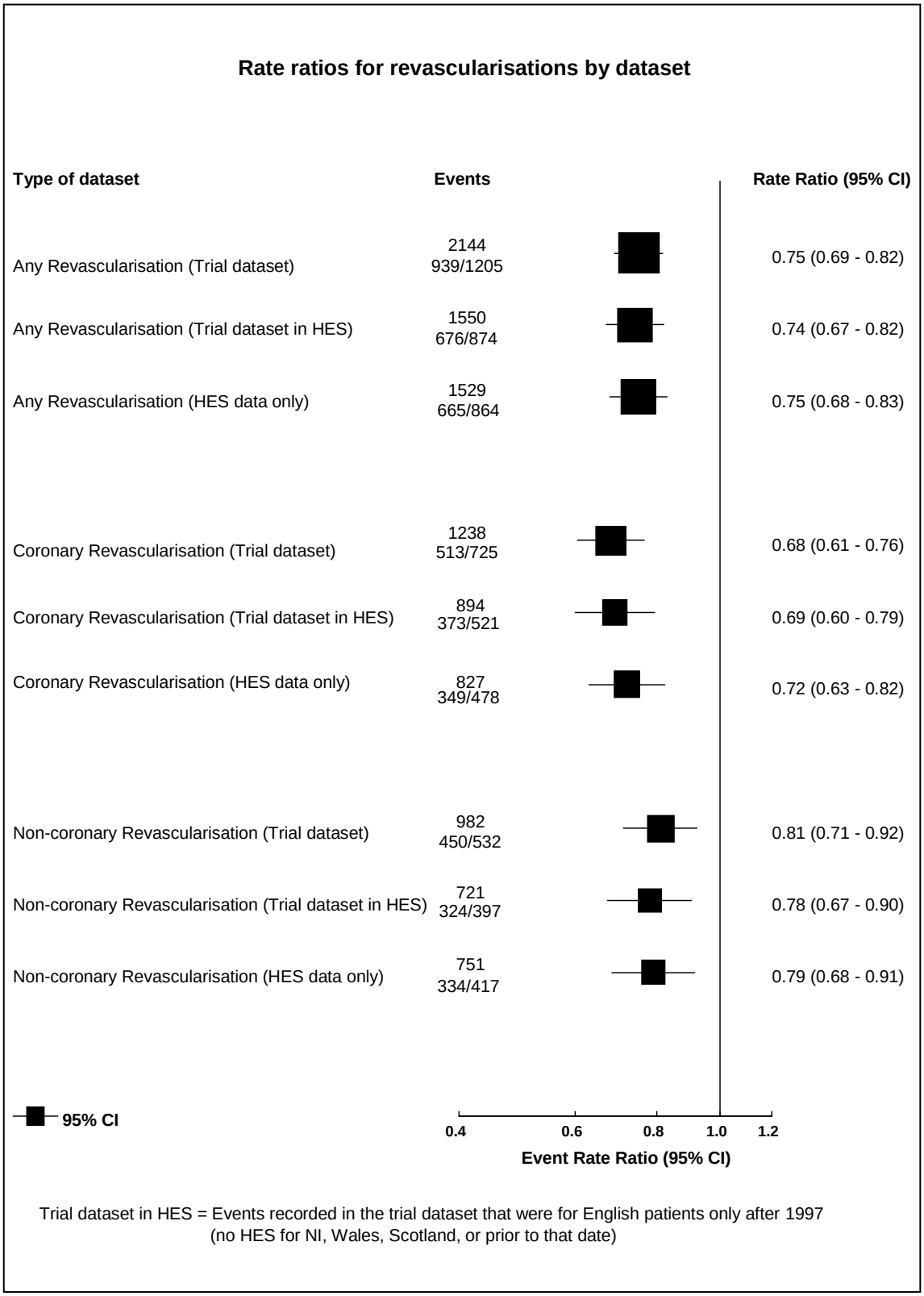


FIGURE 6.3: RATE RATIOS FOR REVASCULARISATION BY DATASET

6.3.4 MAJOR VASCULAR EVENT (MVE)

The secondary outcome in HPS was MVE. This composite outcome is made up of coronary events (non-fatal MI and coronary death), strokes (fatal and non-fatal), and revascularisations (coronary and non-coronary). These numbers of events/procedures reported in the Lancet paper[75] are shown in Table 6.10, along with the numbers of first events within the dataset. These figures differ slightly due to data surrounding these events being updated as more information became available over the post-trial follow-up period.

TABLE 6.10: MAJOR VASCULAR EVENTS BY CATEGORY

MVE	LANCET			HPS		
	Simvastatin	Placebo	Total	Simvastatin	Placebo	Total
Coronary Events	**898**	**1212**	2110	**915**	**1239**	2154
Non-fatal MI	357	574	931	357	574	931
Coronary death	587	707	1294	581	703	1284
Strokes	**444**	**585**	1029	**445**	**585**	1030
Non-fatal stroke	366	499	865	366	499	865
Fatal stroke	96	119	215	97	118	215
Revascularisations	**939**	**1205**	2144	**939**	**1205**	2144
Coronary	513	725	1238	513	725	1238
Non-coronary	450	532	982	450	532	982
Any MVE	**2033**	**2585**	4618	**2041**	**2599**	4640

Lancet results differ from HPS results due to updated event data in the post-trial follow-up period

The results for any MVE are shown in Table 6.11. There were 4640 MVEs in the HPS dataset, with a risk reduction in the simvastatin arm of 24% (95% CI: 0.71-0.80). This is the same rate ratio and confidence interval produced using the Lancet results. After removing participants who resided outside of England, and procedures that occurred prior to 1997, there were 3389 MVEs with a risk reduction of 27% (95% CI: 0.68-0.78). Using codes for MVE recorded in HES along with death records, there were 3587 events with a risk reduction of 26% (95% CI: 0.69-0.79).

TABLE 6.11: MAJOR VASCULAR EVENTS BY DATASET

		Simvastatin-allocated (n=10269)	**Placebo-allocated (n=10267)**	**Incidence Rate Ratio**	**95% CI**
Lancet Results:	Total MVE	2033 (19.8%)	2585 (25.2%)	0.76	0.71-0.80
Dataset results	Total MVE	2040 (19.9%)	2600 (25.3%)	0.76	0.71-0.80
Remove Scot/Wales/N.Ire:		**Simvastatin-allocated (n=9127)**	**Placebo-allocated (n=9114)**		
Change pre-HES dates*	Total MVE	1464 (16.0%)	1925 (21.1%)	0.73	0.68-0.78
Using HES◦	Total MVE	1549 (17.0%)	2038 (22.4%)	0.74	0.69-0.79

**Change pre-HES dates – Removing events prior to 1997*
◦Using only HES records from 1997 onwards

Figure 6.4 shows the risk ratio and 95% confidence intervals for each dataset for any MVE. The rate ratio produced from using the HES records appears to provide a good estimation for the rate ratio seen in the main trial with a very similar rate ratio estimate, and similar confidence intervals.

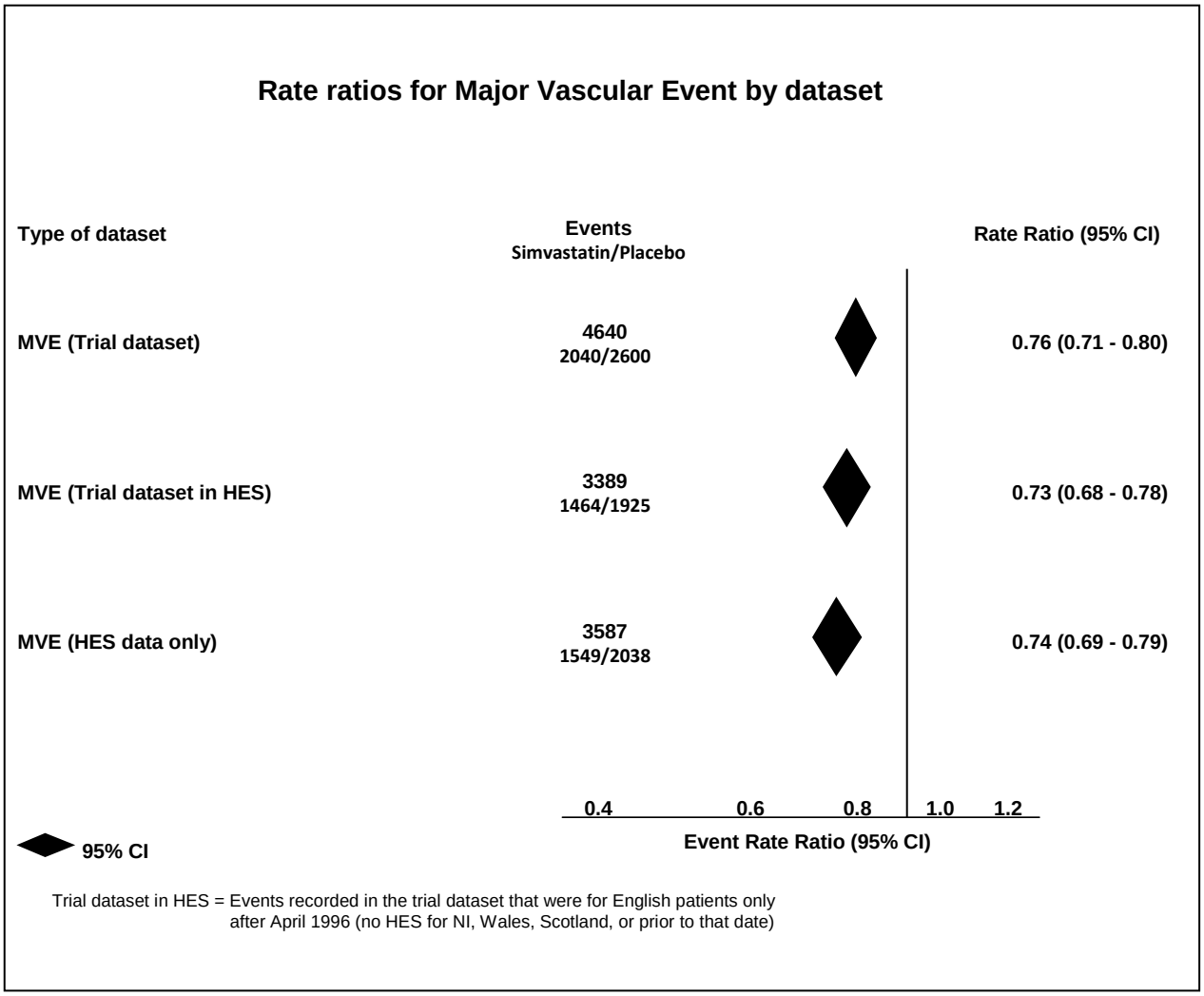


FIGURE 6.4: RATE RATIOS FOR MAJOR VASCULAR EVENTS BY DATASET

6.3.5 ACCURACY OF EVENTS IN HES RECORDS

In order to measure the accuracy of the events recorded in the Hospital Episode Statistics dataset, the Kappa statistics were calculated. These are presented in Table 6.12 for each of the events, and for the overall composite outcome of MVE. The Kappa statistics vary considerably by outcome. Non-fatal MI events had a Kappa statistic of 0.66 which indicates good agreement. The sensitivity of using HES records for ascertaining non-fatal MIs was 74% and the positive predictive value (PPV) was 62%. The Kappa statistic for non-fatal stroke was 0.51, which was the weakest agreement (classed as moderate). The sensitivity was 59% and the PPV was 48%. The agreement

for revascularisation procedures is highest at 0.81 which can be interpreted as very good agreement. Revascularisations also had high sensitivity (83%) and positive predictive value (84%). The composite outcome of MVE had good agreement between HES records and the in-trial reporting with a Kappa statistic of 0.78. The sensitivity was 85% and the PPV was 80%. For each of the four outcomes, specificity was greater than 95%, and negative predictive value (NPV) was greater than 96%.

TABLE 6.12: VALIDATION OF TRIAL OUTCOMES IN HPS USING HES RECORDS

	Kappa Statistic (95% CI)	Sensitivity	Specificity	Positive Predictive Value	Negative Predictive Value
Non-fatal MI	0.66 (0.63-0.69)	74%	98%	62%	99%
Non-fatal stroke	0.51 (0.48-0.54)	59%	98%	48%	98%
Revascularisation	0.82 (0.81-0.84)	83%	99%	84%	98%
MVE	0.78 (0.77-0.79)	85%	95%	80%	96%

6.3.6 RESULTS BY SEX

The Kappa statistic, sensitivity, specificity, positive predictive value (PPV), and negative predictive value (NPV) were also analysed by sex to investigate whether the sex difference observed in recruitment and adherence also persevered in the reporting of outcomes in trial.

The results for male participants are presented in Table 6.13 and for female participants in Table 6.14. The Kappa statistic, sensitivity, and PPV are higher for males than females non-fatal MI, revascularisation procedures, and for the composite outcome of MVE. For non-fatal stroke, female participants had a higher kappa statistic (0.56 vs 0.52) and a higher PPV (55% vs. 46%), but male participants had a higher sensitivity (60% vs. 57%).

TABLE 6.13: VALIDATION OF TRIAL OUTCOMES IN HPS USING HES RECORDS FOR MALES

	Kappa Statistic (95% CI)	Sensitivity	Specificity	Positive Predictive Value	Negative Predictive Value
Non-fatal MI	0.67 (0.64-0.70)	74%	99%	63%	99%
Non-fatal Stroke	0.52 (0.48-0.56)	60%	97%	46%	99%
Revascularisation	0.84 (0.83-0.85)	84%	98%	84%	98%
MVE	0.85 (0.84-0.86)	85%	95%	81%	96%

TABLE 6.14: VALIDATION OF TRIAL OUTCOMES IN HPS USING HES RECORDS FOR FEMALES

	Kappa Statistic (95% CI)	Sensitivity	Specificity	Positive Predictive Value	Negative Predictive Value
Non-fatal MI	0.61 (0.54-0.68)	70%	99%	56%	99%
Non-fatal Stroke	0.56 (0.50-0.63)	57%	98%	55%	98%
Revascularisation	0.84 (0.80-0.87)	83%	99%	83%	99%
MVE	0.82 (0.80-0.85)	81%	97%	80%	97%

6.4 DISCUSSION

6.4.1 FINDINGS

These analyses have demonstrated that there is a good agreement between the outcomes reported in the trial, and those seen in the HES dataset. This was the case for non-fatal myocardial infarction, non-fatal stroke, revascularisation procedures, and the composite end-point of MVE. If the trial analyses had been conducted solely using those events observed in HES, the trial results would be very similar when compared to the results

for participants that we would expect to see in HES (not including events prior to 1997, and not including participants who do not reside in England).

The kappa statistics, along with the sensitivity, specificity, positive predictive value, and negative predictive value, highlight that there were events captured in the trial that were not captured in the HES records, and events recorded in HES that were not reported in-trial. In the case of non-fatal MI, there were participants reporting an MI in-trial, but the closest event within their HES record was angina or atherosclerotic heart disease. Similarly, the ICD-10 diagnosis code I21.9 (relating to an acute MI, unspecified) was recorded for participants in the HES records, but no MI was reported in the trial with some participants having Read codes relating to chest pain but others reporting nothing around that time. Non-fatal strokes were reported more often with the HES dataset than in the trial. This is likely due to the lack of diagnostic criteria at the time with many of the strokes in the HES records being reported as unspecified strokes (ICD-10 code I64). However, there were strokes reported in the trial that did not occur in HES. These may have been old strokes that were recorded as part of an admission due to something else. After investigation, there was no obvious pattern to any codes reported in HES around that time. Revascularisations were the most well captured event in the HES record when compared to the events reported in-trial. This is not surprising as HES data are used for reimbursement and so procedures will have been well reported, even in the earlier years of HES. There were participants who reported a revascularisation procedure in the trial, but with no corresponding event in HES. This could be because the operations were conducted in hospitals just across the border (for example, in Wales or Scotland), or because the operations took place in a private hospital.

When analysing these statistics by sex, there was a small difference in the outcomes reported in-trial and those captured in HES. There was greater agreement between these two datasets for male participants than for females. However, these differences were generally quite small, and may be caused by a smaller proportion of events in female participants.

6.4.2 LIMITATIONS

One of the main limitations with this analysis is that the HES dataset does not cover all participants, or the full time period of the trial. There are no hospital records for participants not residing in England, or for those treated in countries other than England. The dataset also doesn't capture operations or procedures performed privately. However, by limiting the comparisons to only those reported in trial that should also be captured in the HES records (events occurring in 1997 or later, and only in participants residing in England); the comparison of results should be reasonably fair.

Another limitation is that at the start of the trial, linked HES records did not exist, and even once these records were linked (from 1997 onwards), the reporting of events was variable. Coding standards have improved over the last 20 years, and events are reported more accurately as diagnostic medicine improves. Whilst the early years of HES records may leave room for improvement, they will have improved over time. This furthers the case for using routine data as a method of follow-up as they showed reasonable agreement with in-trial event capturing even in the infancy of HES records.

An additional limitation is the difficulties faced when working with older datasets. The datasets that hold the in-trial information for HPS are historic, and in some cases have been overwritten by information that was received in the post-trial follow-up period. Additionally, the historic nature of the dataset means that diagnostic codes used at the

time are no longer the diagnostic codes used now, and so care should be taken when translating between these codes.

6.4.3 CONCLUSIONS

Electronic health records, even in 1997 provided very similar results to those reported in the main HPS paper. Whilst some events were not captured in the HES data, there were other events captured that were potentially missed in the trial. These analyses have demonstrated that there is good agreement between the trial data, and routinely collected data from HES, for cardiovascular disease related outcomes. There is the potential to link HES records with hospital data from other devolved nations which would provide complete coverage of the UK.

This chapter has described how HES data can be used to supplement or even replace the need for follow-up in cardiovascular disease trials. Chapter 3 demonstrated the difficulty in recruiting participants to two cardiovascular disease trials, particularly female participants. This chapter has demonstrated that follow-up appointments can be reduced in cardiovascular disease trials, both in time and quantity, as major vascular events are captured accurately in routine data. By removing the need for lengthy and frequent follow-up visits, recruitment to cardiovascular clinical trials may prove to be easier. Chapter 5 demonstrated that there are other benefits to using routine data, namely being able to follow-up larger numbers of participants for a longer period of time in a cost-efficient manner. This, along with the results in this chapter, provide evidence that large randomised trials can be conducted using routine data as the primary form of follow-up to improve trial efficiency, and by increasing the sample size and length of follow-up, and improve the accuracy of the results.

7 DISCUSSION

7.1 SUMMARY

This chapter summarises the main findings from this research; the use of routine data for recruitment in randomised controlled trials, the differences observed between males and females for adherence to treatment, and the use of routine data for follow-up in clinical trials. This chapter also discusses the implications of the sex effects for clinical care, the evidence for using routine data in clinical trials, and future work that is required in this area.

7.1.1 RECRUITMENT TO CLINICAL TRIALS USING ROUTINE DATA

Chapters 2 and 3 looked at the use of routine data for recruitment to clinical trials. Chapter 2 summarised the literature and provided details on the types of routine data that have previously been used for recruitment to clinical trials. Electronic health records were the most commonly used data source, and the review demonstrated that they have successfully been used for recruitment to both large and small trials in a variety of disease areas. Chapter 3 then looked in detail at the recruitment to two large cardiovascular disease trials. Both of the trials, the Heart Protection Study (HPS), and the Randomized EValuation of the Effects of Anacetrapib through Lipid-modification trial (REVEAL), used electronic health records to recruit participants. These trials, which took place 17 years apart, both invited large numbers of potentially eligible patients by identifying relevant codes in electronic records derived from medical notes. The demographics of those progressing through the recruitment pathway were investigated.

Sex was the main demographic factor that determined successful recruitment in these cardiovascular disease trials, with significantly fewer females agreeing to participate/progress to the next stage than males. Although the overall proportion of invited patients who agreed to attend a screening appointment was lower in the more recent REVEAL trial, similar sex differences were seen in both trials. Additionally, the proportion of patients identified from electronic health records that respond to postal invitations has declined substantially over this time period. This suggests the main issue is encouraging patients to attend a screening appointment in the first place. However, the differences in sex are not driven solely by one stage of the recruitment process as at every stage, females were less likely to continue. Fewer females responded to the invitation, and those that did respond were less likely to accept the invitation than males. At the screening visit, females were less likely to be eligible, and less likely to consent to enter the run-in period. A smaller proportion of females attended the randomisation visit than males, and those that did were less likely to be randomised. In HPS, the proportion of invited females that went on to be randomised was 10.8% compared to 18.8% of males. In REVEAL, the proportion of invited females that went on to be randomised was 0.9% compared to 3.4% of males.

The sex related differences observed in those progressing through the recruitment pathway were still significant after adjusting for age, ethnicity, and socioeconomic status. Unfortunately, there was very little information available about why patients decided not to take part in the trials which makes it difficult to identify areas for improvement. Additionally, the differences in sex occurred at every stage of the recruitment pathway and there was no single point during recruitment where changes could resolve this problem.

7.1.2 ADHERENCE TO TREATMENT

The differences observed by sex in recruitment in Chapter 3 were then investigated further in Chapter 4 where adherence to treatment was explored. The analyses demonstrated that there was a significant difference between males and females in both type of follow-up visit and adherence to treatment. In HPS, females were less likely than males to attend a follow-up clinic in person ("direct follow-up"). As participants who switched away from clinic/hospital visits had to stop taking their study treatment for safety reasons, this meant that there were fewer female participants taking their treatments. Females were significantly less likely than males to have an in-person follow-up by the first visit after randomisation (OR: 0.73, 95% CI: 0.61-0.87). This was still the case at the end of the follow-up period (five years after randomisation) with the odds of a female having an in-person follow-up of 0.68, 95% CI: 0.61-0.76.

In both of the trials, female participants were also less likely than males to be adherent to their treatments. In HPS, females were less likely to be adherent to their vitamin or matching placebo capsules and even less likely to be adherent to their statin or matching placebo tablet. These differences in sex were observed from the first study visit four months after randomisation where females were significantly less likely than males to report being adherent to their study treatments (statin/placebo OR: 0.81, 95% CI: 0.71-0.92, vitamin/placebo OR: 0.78, 95% CI: 0.68-0.89) and continued throughout the study period until the five-year post randomisation visit (OR: 0.74, 95% CI: 0.68-0.81).

As adherence to treatment at each time point was calculated as a percentage of those eligible to be taking their treatments i.e. those still attending in-person follow-up visits, it is likely that the sex difference observed in HPS underestimates the true effect as fewer women were taking their treatments at all as they had switched to non-direct

follow-up. In REVEAL, a similar sex difference was observed with fewer females being adherent to the study treatments. There was a smaller difference between the two treatments though with similar proportions of participants being adherent to their atorvastatin and their anacetrapib or matching placebo. Whilst there were no significant sex differences in adherence to either treatment at the first follow-up visit two months after randomisation, there were significant differences at the midpoint of the trial (two years after randomisation). Females were significantly less likely than males to report being adherent to their study treatments (atorvastatin OR: 0.73, 95% CI: 0.60-0.88, anacetrapib/placebo OR: 0.75, 95% CI: 0.62-0.91). This was still the case for both treatments at the end of the in-trial follow-up period. Females were significantly less likely than males to still be adherent to both their atorvastatin treatment (OR: 0.76, 95% CI: 0.60-0.94) and their anacetrapib/placebo treatment (OR: 0.67, 95% CI: 0.55-0.84).

This causes concern as there are already difficulties in recruiting female participants to cardiovascular disease trials and having fewer females taking part and even fewer being adherent to their treatments could leave studies underpowered for sex specific effects. A meta-analysis looking at statin therapy for primary prevention of coronary heart disease (CHD) found that the use of statins reduced the risk of CHD events in males, but did not find evidence of this protection in females.[180] A previously published meta-analysis looking at the impact of gender on statin efficacy found that women receiving statin therapy did not have reduction in their risk of stroke or death, and whilst there was a reduction in risk of myocardial infarction, it was not statistically significant.[181] They concluded that the lack of statistical significance was most likely due to the small female population within the trials, rather than due to there being no effect. This further highlights the need to include adequate numbers of female participants in cardiovascular disease clinical trials. These factors potentially limit the ability to

establish the safety and efficacy of interventions in females which could lead to serious adverse public health consequences. One reason for this may be the time involved in attending study visits, particularly over many years, which could cause participants to switch to non-direct follow-up or stop taking their tablets. This might be particularly relevant to female participation in trials as females are more likely to be care-givers and therefore may have other commitments that affect their ability to come in for regular study visits.

7.1.3 FOLLOW-UP IN CLINICAL TRIALS USING ROUTINE DATA

Chapters 5 and 6 looked at the use of routine data for follow-up in clinical trials. Chapter 5 reviewed the literature and identified different types of routine data that have been used previously for follow-up. As with the review of randomised trials that used routine data for recruitment, electronic health records were the most commonly used data source for follow-up. Other sources of routine data that were used included registry data, insurance databases, and mortality records. The trials that used routine data were relatively large and generally had longer follow-up periods. The use of electronic health records for follow-up is a cost-effective way to follow-up large numbers of participants, and there is potential for more clinical trials to adopt follow-up using routine data to maximise the follow-up periods in clinical trials, to make it easier for participants to take part in research, and to increase the sample sizes in trials so that moderate treatment effects can be detected reliably.

Chapter 6 looked at whether routine data, specifically Hospital Episode Statistics (HES) data, could be used instead of clinic visits to follow-up participants in clinical trials as a potentially more cost-effective method. Using data from HPS, the trial efficacy outcomes were analysed and then replicated using only events that were recorded in the

HES datasets. These analyses demonstrated very good agreements between the trial outcomes and those recorded in the HES data. This was the case for non-fatal myocardial infarction, non-fatal stroke, revascularisation procedures, and the composite end-point of major vascular event (MVE). Using a comparable subset of the trial dataset (removing events prior to 1997 and participants that resided outside of England), the relative reduction in the rate of first non-fatal myocardial infarction (MI) was 38% (95% CI: 0.53-0.72) in the simvastatin arm compared to the placebo arm. Using the events recorded in HES, the relative reduction in the rate of first non-fatal MI was 36% (95% CI: 0.56-0.74). For the composite outcome, MVE, there was a relative risk reduction of 27% (95% CI: 0.68-0.78) using the subset of the trial dataset. Using the events recorded in HES, the relative reduction in the rate of first MVE was 26% (95% CI: 0.69-0.79).

If the trial analyses had been conducted solely using those events observed in HES, the trial results would have been very similar to the trial results among participants that we would expect to see in HES (not including events prior to 1997, and not including participants who do not reside in England). As expected, there were some outcomes recorded in HES that were not recorded in the trial, and some outcomes reported in the trial that were not captured in HES, but overall using electronic health records provided very similar results to those reported in the trial. These analyses demonstrate that HES data can be used to supplement follow-up appointments in cardiovascular disease trials, reducing both the time and quantity of follow-up appointments. Unfortunately, these analyses could not be replicated using the REVEAL trial data as REVEAL is still in post-trial follow-up. This would have provided a more recent view of using HES records for follow-up but there is potential for this research to be completed after the post-trial follow-up period has finished.

By reducing the need for lengthy and frequent follow-up visits, recruitment to cardiovascular clinical trials may prove to be easier. Both Chapters 5 and 6 provide evidence that large randomised controlled trials can be conducted using routine data as the primary form of follow-up. This improves trial efficiency and should allow larger numbers of participants to be followed up over a greater period of time, which will improve the accuracy of the results.

7.2 STRENGTHS AND LIMITATIONS

The work presented in this thesis has demonstrated that routine data can be used for both recruitment and follow-up in large cardiovascular disease trials in the UK. One strength of this research is that both of the trials used for these analyses contained information on the whole recruitment process from the identification of potentially eligible patients through to randomisation. This ensures that data on response to invitation are based on a true known denominator and the use of a common method across very large numbers of patients.

Another strength of this research is the age of the trials used for these analyses. HPS started recruiting in 1994 and REVEAL started recruiting in 2007. The differences observed in recruitment and adherence to treatment between males and females were seen in both trials, which indicates that this is an ongoing problem. In fact, the differences were greater in the more recent REVEAL trial, which may mean that the problem with sex disparity in trials is getting worse. Furthermore, the follow-up analyses in HPS used HES data from the beginning of linked HES records. Even using data from the early years of HES, the trial outcomes were captured accurately. As the quality of coding and recording of events in HES has improved over time[182–184], this implies that outcomes will be recorded more accurately within the HES datasets in

more recent years. This means the use of HES data for follow-up in clinical trials should be more accurate for trials going forwards.

One limitation of this research is that it has been conducted on two similar trials that were run by the same unit, HPS and REVEAL. Although these trials were nearly 20 years apart, they recruited from similar patient populations and both trials were investigating the effects of lipid-modifying treatment among people with cardiovascular disease. Both trials recruited patients who were at an increased risk of cardiovascular disease. The differences observed may be specific to cardiovascular disease trials and may not be generalisable to other areas of research.

Another limitation of this research is its retrospective nature. Working with older datasets presents challenges. The datasets that hold the in-trial information for HPS are older, and unfortunately the database was maintained according to different standards in that era. This means that some of the information held has been overwritten during the post-trial follow-up period.

In addition to this, HPS was a paper-based trial that was dual-entered into the database. Working with the raw data proved challenging as the dataset required a large amount of data cleaning which was a lengthy process. There was very little documentation around the data cleaning procedures that had been used in the main trial to recreate an analysis dataset. Additionally, accessing older datasets can be difficult due to differences in operating systems and protocols and care must be taken to ensure that the relevant approvals are in place to access the data as retrospective consent can be difficult to obtain. Furthermore, working retrospectively meant that there was no way to capture any additional information about participants' reasons for discontinuing with the studies. In HPS, the only information captured around not wishing to continue was

recorded as participant choice. While these data do not allow further exploration of the underlying causes of these sex differences, recognition by investigators and consideration of these issues when designing recruitment materials, in particular to ensure that there are no hidden sex biases which might be causing female patients to not attend or take part.

One final limitation of this research is the access to routine datasets. Access to the HES records for HPS had already been granted, prior to commencing this research. However, it is a time-consuming process to gain access to data from NHS Digital, and data are not provided in real time. This raises issues in clinical trials where safety information may be needed, including blood work, to monitor how a new therapy is acting and to allow data monitoring committees to oversee the safety and efficacy of the trial. These analyses looked at well-defined trial outcomes (non-fatal MI, non-fatal stroke, revascularisation procedures), and not at the safety outcomes which included liver and muscle enzymes. These would not necessarily be recorded in HES data, and so the safety of the experimental treatments could not be monitored routinely. However, there is the potential to access routinely taken blood samples in trial participants which would allow blood safety to be monitored remotely. However, this still does not resolve the problems caused by the delays in accessing this data.

7.3 FUTURE WORK

There is a need for better understanding from qualitative studies about the reasons for differences in trial related behaviour between males and females. By including questions in future research that would capture the reasons for not wanting to continue through the recruitment pathway, some of these reasons could be addressed. The inclusion of more females in focus groups prior to a trial starting could also help ensure

that the recruitment process and trial literature do not contain any hidden biases that may dissuade females from taking part.

Theses analyses demonstrate the value of routine data and its use for follow-up in clinical trials. There are many other large data sources that could be used for clinical trials. Health registries contain a large amount of detailed information. Examples of these include the National Cancer Registration and Analysis Service (NCRAS) and the UK Renal Registry (UKRR). Linking data from these registries with clinical trial data can allow more detailed analyses of specific diseases. In HPS, data were collected on mortality from the Office for National Statistics (ONS), and on cancer from national cancer registries.[75] These were used to provide further information for adjudication of events, but there is the potential to use these as the main data sources for these outcomes in future research. There are also examples of clinical trials where participants are identified, randomised, and followed up using the registry data.[185,186] Future research may look at combining data from registries with HES datasets to allow more clinical information to be captured or to allow the randomisation and follow-up of patients to occur without the need for extra study visits.

There is more work that could be done investigating the clinical implications for the lack of power for effects in women. Cardiovascular disease mortality is higher in females than in males (35% of deaths worldwide in females compared to 32% of deaths in males).[187] However, cardiovascular disease is still perceived to be a male illness. The under-representation of females in cardiovascular disease clinical trials has been acknowledged in two meta-analyses, where the proportion of females across the included trials was between 25-38%.[180,181] These meta-analyses both concluded that there was not sufficient evidence to support the reduction in risk of death or stroke for

women on statins. Further research and future clinical trials should aim to include a greater proportion of female participants so that the effects of treatments in cardiovascular disease can be suitably measured. Taking potentially ineffective treatments is wasteful and without suitable numbers of females, clinical trials lose their ability to establish the safety and efficacy of interventions in a female population.

7.4 CONCLUSIONS

The research presented in this thesis has demonstrated how routine data could potentially be used for recruitment and follow-up in large randomised trials, particularly in cardiovascular disease trials in the UK. There do appear to be systematic differences in those progressing through the recruitment pathway to randomisation with fewer females progressing at each stage, and these differences persist during the trial with fewer females having direct follow-up, and less females being adherent to their study treatments. However, the trial outcome events of non-fatal MI, non-fatal stroke, and revascularisation procedures are captured well within HES datasets, and therefore could replace follow-up appointments in clinical trials if data can be made available within a suitable time frame for safety monitoring. If electronic health records can be used as a method of follow-up, the number of study visits may be able to be reduced or even completely removed, which may encourage more females to take part in these trials.

Routine health records, for example HES data, and national disease registries contain a wealth of information about patients that can be used to optimise both recruitment to trials and the follow-up of patients. The use of electronic health records (EHR) has the potential to provide more cost-effective and efficient methods for research.

REFERENCES

1. Hariton E, Locascio JJ. Randomised controlled trials - the gold standard for effectiveness research: Study design: randomised controlled trials. *BJOG*. 2018;125(13):1716. doi:10.1111/1471-0528.15199

2. Spencer SA, Davies MP. Hospital episode statistics: improving the quality and value of hospital data: a national internet e-survey of hospital consultants. *BMJ Open*. 2012;2(6):e001651. doi:10.1136/bmjopen-2012-001651

3. MRC/BHF Heart Protection Study Collaborative Group. MRC / BHF Heart Protection Study of cholesterol-lowering therapy and of antioxidant vitamin supplementation in a wide range of patients at increased risk of coronary heart disease death : early safety and efficacy experience. *Eur Heart J*. 1999;20:725-741.

4. Heart Protection Study Collaborative Group. MRC/BHF heart protection study of antioxidant vitamin supplementation in 20 536 high risk individuals: a randomised placebo controlled trial. *Lancet*. 2002;360:23-33. doi:10.1016/S0140-6736(02)09328-5

5. Bowman L, Chen F, Sammons E, et al. Randomized Evaluation of the Effects of Anacetrapib through Lipid-modification (REVEAL)—A large-scale, randomized, placebo-controlled trial of the clinical effects of anacetrapib among people with established vascular disease: Trial design, recruitment, a. *Am Heart J*. 2017;187:182-190. doi:10.1016/j.ahj.2017.02.021

6. Ulmer M, Robinaugh D, Friedberg JP, Lipsitz SR, Natarajan S. Usefulness of a run-in period to reduce drop-outs in a randomized controlled trial of a behavioral intervention. *Contemp Clin Trials*. 2008;29(5):705-710. doi:10.1016/j.cct.2008.04.005

7. Watson JM, Torgerson DJ. Increasing recruitment to randomised trials: a review of randomised controlled trials. *BMC Med Res Methodol*. 2006;6:34. doi:10.1186/1471-2288-6-34

8. Patel MX. Challenges in recruitment of research participants. *Adv Psychiatr Treat*. 2003;9(3):229-238. doi:10.1192/apt.9.3.229

9. HPS2/THRIVE Collaborative Group. Effects of Extended-Release Niacin with Laropiprant in High-Risk Patients. *N Engl J Med*. 2014;371:203-212. doi:10.1056/NEJMoa1300955

10. Campbell MK, Snowdon C, Francis D, et al. Recruitment to randomised trials: strategies for trial enrollment and participation study. The STEPS study. *Health Technol Assess*. 2007;11(48). doi:10.3310/hta11480

11. Cook JA, Hislop JM, Altman DG, et al. Use of methods for specifying the target difference in randomised controlled trial sample size calculations: Two surveys of trialists' practice. *Clin Trials J Soc Clin Trials*. 2014;11(3):300-308. doi:10.1177/1740774514521907

12. Faber J, Fonseca LM. How sample size influences research outcomes. *Dental Press J Orthod*. 2014;19(4):27-29. doi:10.1590/2176-9451.19.4.027-029.EBO

13. Nunan D, Aronson J, Bankhead C. Catalogue of bias: attrition bias. *BMJ evidence-based Med*. 2018;23(1):21-22. doi:10.1136/ebmed-2017-110883

14. Brueton VC, Tierney JF, Stenning S, et al. Strategies to improve retention in randomised trials: a Cochrane systematic review and meta-analysis. *BMJ Open*. 2014;4(2):e003821. doi:10.1136/bmjopen-2013-003821

15. Collins R, MacMahon S. Reliable assessment of the effects of treatment on mortality and major morbidity, I: clinical trials. *Lancet*. 2001;357:373-380. doi:10.1016/S0140-6736(00)03651-5

16. McDonald AM, Knight RC, Campbell MK, et al. What influences recruitment to randomised controlled trials? A review of trials funded by two UK funding agencies. *Trials*. 2006;7:9. doi:10.1186/1745-6215-7-9

17. Puffer S, Torgerson DJ. Recruitment difficulties in randomised controlled trials. *Control Clin Trials*. 2003;24:214S--215S.

18. Ioannidis JPA, Greenland S, Hlatky MA, et al. Increasing value and reducing waste in research design, conduct, and analysis. *Lancet*. 2014;383:166-175. doi:10.1016/S0140-6736(13)62227-8

19. Moher D, Liberati A, Tetzlaff J, Altman DG. Guidelines and Guidance Preferred Reporting Items for Systematic Reviews and Meta-Analyses: The PRISMA Statement. doi:10.1371/journal.pmed.1000097

20. Nichols G, Granowitz C, Reynolds K, Philip S. The potential of electronic health record data to optimize recruitment efficiency in cardiovascular outcome trials. *J Clin Lipidol*. 2017;11(3):796-797.

21. Spencer J, Wilson A, Bailey N, Longson MS. Utilizing electronic medical record networks for identifying patients for clinical trial recruitment. *Value Heal*. 2015;18(3):A18.

22. Barrett J. Matching the right patient to the right clinical study: The opportunities for more efficient clinical studies using selective-recruitment designs. *Trials*. 2017;18(Supplement 1). doi:http://dx.doi.org/10.1186/s13063-017-1902-y

23. Kamauu AG, Agbor SOT. Conformity between protocol eligibility criteria for electronic patient identification: A comparison of clinical trials. *Pharmacoepidemiol Drug Saf*. 2013;22(SUPPL. 1):222-223. doi:http://dx.doi.org/10.1002/pds.3512

24. Norton BL, Gover M, McMurry C, Deluca J, Cunningham C. Feasibility and acceptability of a group medical visit intervention to improve hcv treatment uptake among persons who inject drugs (PWID) in a primary care setting. *J Gen Intern Med*. 2017;32(2 Supplement 1):S192.

25. Keosaian J, Dresner D, Cerrada C, et al. Recruitment strategies for community-based yoga research in a predominant minority population. *BMC Complement Altern Med*. 2012;12(SUPPL. 1).

26. VanNostrand K, Fisher K. Smoking cessation after surviving critical illness-is it feasible?-Authors experience with implementation and recruitment. *Am J Respir Crit Care Med*. 2014;189(MeetingAbstracts).

27. Aroda VR, Vickery E, Brodsky I, et al. Optimizing use of the electronic health record (EHR) to identify persons at high risk of diabetes in the D2D study. *Diabetes*. 2016;65(Supplement 1):A377. doi:http://dx.doi.org/10.2337/db16-1375-1656

28. Parikh PS, Ratanamaneechat S, George D, et al. Recruitment and enrollment of asthmatics in a phase II clinical trial. *Am J Respir Crit Care Med*. 2010;181(1 MeetingAbstracts).

29. Wu L-TT, Spratt S, Heidenfelder B, Tai B. Using electronic health records data for clinical research. *Drug Alcohol Depend*. 2017;171:e219. doi:http://dx.doi.org/10.1016/j.drugalcdep.2016.08.599

30. De Luise C, March JS, Gersing KR, Cohen G. Using electronic medical records to study the safety of psychiatric medications used in pediatric patients. *Pharmacoepidemiol Drug Saf*. 2009;18(S1):S240. doi:http://dx.doi.org/10.1002/pds.1806

31. Justin S, Wilson A, Bailey N, Longson M. Utilizing electronic medical record networks and advanced analytics for identifying patients for clinical trial recruitment. *Pharmacoepidemiol Drug Saf*. 2015;24(SUPPL. 1):587. doi:http://dx.doi.org/10.1002/pds.3838

32. Margolis KL, Kerby TJ, Asche SE, et al. Home blood pressure telemonitoring and case management to control hypertension: Hyperlink design, baseline characteristics, and intervention adherence. *J Clin Hypertens*. 2010;12(SUPPL. 1):A81-A82. doi:http://dx.doi.org/10.1111/j.1751-7176.2010.00282.x

33. Thacker T, Wegele AR. Utility of electronic medical record databases using ICD-9 criteria for recruitment in clinical research: From rare to common disease. *Mov Disord*. 2015;30(SUPPL. 1):S446. doi:http://dx.doi.org/10.1002/mds.26295

34. Khan Y, Roth C, Payne P, et al. EHR-based Clinical Trial Alert Effects on Recruitment to a Neurology Trial across Institutions: Interim Analysis of a Randomized Controlled Study. *AMIA Jt Summits Transl Sci proceedings AMIA Jt Summits Transl Sci*. 2013;2013:117.

35. Cameron K, Kho A, Hauber S, et al. On target: Identifying and reaching patients for public health interventions or research. *J Gen Intern Med*. 2010;25(SUPPL. 3):S339. doi:http://dx.doi.org/10.1007/s11606-010-1338-5

36. Thacker T, Wegele AR, Pirio Richardson S. Utility of electronic medical record for recruitment in clinical research: from rare to common disease. *Mov Disord Clin Pract*. 2016;3(5):507-509.

37. Patrao DF, Oleynik M, Massicano F. Recruit--An Ontology Based Information Retrieval System for Clinical Trials Recruitment. *Stud Health Technol Inform*. 2015;216:534-538.

38. Park ER, Quinn VP, Chang Y, et al. Recruiting pregnant smokers into a clinical trial: Using a network-model managed care organization versus community-based practices. *Prev Med (Baltim)*. 2007;44(3):223-229. doi:http://dx.doi.org/10.1016/j.ypmed.2006.10.008

39. Butt F, Khan K, Chaudhry S. Electronic Patient Records to Identify Patients in the United Kingdom with Diabetic Macular Oedema Suitable for ILUVIEN (Fluocinolone Acetonide). *Ophthalmol Ther*. 2016;5(1):81-94. doi:http://dx.doi.org/10.1007/s40123-016-0053-7

40. Schreiweis B, Trinczek B, Kopcke F, et al. Comparison of Electronic Health Record System Functionalities to support the patient recruitment process in clinical trials. *Int J Med Inform*. 2014;83(11):860-868. doi:http://dx.doi.org/10.1016/j.ijmedinf.2014.08.005

41. De Moor G, Sundgren M, Kalra D, et al. Using electronic health records for clinical research: The case of the EHR4CR project. *J Biomed Inform*. 2015;53:162-173. doi:10.1016/J.JBI.2014.10.006

42. Cowie MR, Blomster JI, Curtis LH, et al. Electronic health records to facilitate clinical research. *Clin Res Cardiol*. 2017;106(1):1-9. doi:https://dx.doi.org/10.1007/s00392-016-1025-6

43. Doods J, Holzapfel K, Dugas M, Fritz F. Development of best practice principles for simplifying eligibility criteria. *Stud Health Technol Inform*. 2013;192:1153.

44. Preston NJ, Farquhar MC, Walshe CE, et al. Strategies designed to help healthcare professionals to recruit participants to research studies. *Cochrane Database Syst Rev*. 2016;(2). doi:10.1002/14651858.MR000036.pub2

45. Schreiweis B, Gruber G. First experiences in implementing a software-based support for patient recruitment at Heidelberg university hospital. *Stud Health Technol Inform*. 2012;180:1147-1149.

46. Mosis G, Koes B, Dieleman J, Stricker BC, van der Lei J. Randomised studies in general practice: How to integrate the electronic patient record. *Inform Prim Care*. 2005;13(3):209-213.

47. Stuardi T, Cox H. Database recruitment: A solution to poor recruitment in randomized trials? *Fam Pract*. 2011;28(3):329-333. doi:http://dx.doi.org/10.1093/fampra/cmq108

48. Miller JL. The EHR solution to clinical trial recruitment in physician groups. *Health Manag Technol*. 2006;27(12):22-25.

49. Zimmerman LP, Goel S, Sathar S, et al. A Novel Patient Recruitment Strategy: Patient Selection Directly from the Community through Linkage to Clinical Data. *Appl Clin Inform*. 2018;9(1):114-121. doi:https://dx.doi.org/10.1055/s-0038-1625964

50. Embi PJ, Jain A, Clark J, et al. Effect of a clinical trial alert system on physician participation in trial recruitment. *Arch Intern Med*. 2005;165(19):2272-2277. doi:http://dx.doi.org/10.1001/archinte.165.19.2272

51. Ko NY, Fu JL, Lane SC, et al. Cancer clinical trial enrollment of diverse and underserved patients within an urban safety net hospital. *J community Support Oncol*. 2015;13(12):429-435. doi:https://dx.doi.org/10.12788/jcso.0181

52. Sumi E, Teramukai S, Yamamoto K, Satoh M, Yamanaka K, Yokode M. The correlation between the number of eligible patients in routine clinical practice and the low recruitment level in clinical trials: a retrospective study using electronic medical records. *Trials*. 2013;14(1):426. doi:https://dx.doi.org/10.1186/1745-6215-14-426

53. Gulliford MC, van Staa TP, McDermott L, et al. Cluster randomized trials utilizing primary care electronic health records: Methodological issues in design, conduct, and analysis (eCRT Study). *Trials*. 2014;15(1):220. doi:http://dx.doi.org/10.1186/1745-6215-15-220

54. Smith KS, Eubanks D, Petrik A. Using web-based screening to enhance efficiency of HMO clinical trial recruitment in women aged forty and older. *Clin Trials*. 2007;4(1):102-105. doi:http://dx.doi.org/10.1177/1740774506075863

55. Schmickl CN, Li M, Li G, Wetzstein MM, Herasevich V, Gajic O. The accuracy and efficiency of electronic screening for recruitment into a clinical trial on COPD. *Respir Med*. 2011;105(10):1501-1506. doi:http://dx.doi.org/10.1016/j.rmed.2011.04.012

56. Ng KP, Jain P, Gill PS, et al. Results and lessons from the spironolactone to prevent cardiovascular events in early stage chronic kidney disease (STOP-CKD) randomised controlled trial. *BMJ Open*. 2016;6(2):e010519. doi:http://dx.doi.org/10.1136/bmjopen-2015-010519

57. Treweek S, Pearson E, Smith N, et al. Desktop software to identify patients eligible for recruitment into a clinical trial: using SARMA to recruit to the ROAD feasibility trial. *Inform Prim Care*. 2010;18(1):51-58.

58. Mackenzie IS, Ford I, Walker A, et al. Multicentre, prospective, randomised, open-label, blinded end point trial of the efficacy of allopurinol therapy in improving cardiovascular outcomes in patients with ischaemic heart disease: protocol of the ALL-HEART study. *BMJ Open*. 2016;6(9):e013774. doi:https://dx.doi.org/10.1136/bmjopen-2016-013774

59. Ruffin MT 4th, Nease DEJ. Using patient monetary incentives and electronically derived patient lists to recruit patients to a clinical trial. *J Am Board Fam Med*. 2011;24(5):569-575. doi:https://dx.doi.org/10.3122/jabfm.2011.05.100169

60. Williams NH, Amoakwa E, Belcher J, et al. Activity increase despite arthritis (AIDA): Phase II randomised controlled trial of an active management booklet for hip and knee osteoarthritis in primary care. *Br J Gen Pract*. 2011;61(589):e452-e458. doi:http://dx.doi.org/10.3399/bjgp11X588411

61. Eakin EG, Bull SS, Riley K, Reeves MM, Gutierrez S. Recruitment and retention of Latinos in a primary care-based physical activity and diet trial: The Resources for Health study. *Health Educ Res*. 2007;22(3):361-371. doi:http://dx.doi.org/10.1093/her/cyl095

62. Newhouse N, Martin A, Jawad S, et al. Randomised feasibility study of a novel experience-based internet intervention to support self-management in chronic asthma. *BMJ Open*. 2016;6(12):e013401. doi:https://dx.doi.org/10.1136/bmjopen-2016-013401

63. van Staa T-P, Dyson L, McCann G, et al. The opportunities and challenges of pragmatic point-of-care randomised trials using routinely collected electronic records: evaluations of two exemplar trials. *Health Technol Assess*. 2014;18(43):1-146. doi:https://dx.doi.org/10.3310/hta18430

64. Kernan WN, Viscoli CM, Demarco D, et al. Boosting enrollment in neurology trials with Local Identification and Outreach Networks (LIONs). *Neurology*. 2009;72(15):1345-1351. doi:http://dx.doi.org/10.1212/WNL.0b013e3181a0fda3

65. Simon GE, Beck A, Rossom R, et al. Population-based outreach versus care as usual to prevent suicide attempt: study protocol for a randomized controlled trial. *Trials*. 2016;17(1):452. doi:http://dx.doi.org/10.1186/s13063-016-1566-z

66. Effoe VS, Katula JA, Kirk JK, et al. The use of electronic medical records for recruitment in clinical trials: Findings from the Lifestyle Intervention for Treatment of Diabetes trial. *Trials*. 2016;17(1):496. doi:http://dx.doi.org/10.1186/s13063-016-1631-7

67. Johnson EJ, Niles BL, Mori DL. Targeted recruitment of adults with type 2 diabetes for a physical activity intervention. *Diabetes Spectr*. 2015;28(2):99-105. doi:https://dx.doi.org/10.2337/diaspect.28.2.99

68. Ramsey TM, Snyder JK, Lovato LC, et al. Recruitment strategies and challenges in a large intervention trial: Systolic Blood Pressure Intervention Trial. *Clin Trials*. 2015;13(3):319-330. doi:https://dx.doi.org/10.1177/1740774516631735

69. McCurry SM, Von Korff M, Vitiello M V, et al. Frequency of comorbid insomnia, pain, and depression in older adults with osteoarthritis: predictors of enrollment in a randomized treatment trial. *J Psychosom Res*. 2011;71(5):296-299. doi:http://dx.doi.org/10.1016/j.jpsychores.2011.05.012

70. Wu L-T, Brady KT, Spratt SE, et al. Using electronic health record data for substance use Screening, Brief Intervention, and Referral to Treatment among adults with type 2 diabetes: Design of a National Drug Abuse Treatment Clinical Trials Network study. *Contemp Clin Trials*. 2016;46:30-38. doi:http://dx.doi.org/10.1016/j.cct.2015.11.009

71. Treweek S, Lockhart P, Pitkethly M, et al. Methods to improve recruitment to randomised controlled trials: Cochrane systematic review and meta-analysis. *BMJ Open*. 2013;3(2):e002360. doi:10.1136/bmjopen-2012-002360

72. Aung T, Haynes R, Barton J, et al. Cost-effective recruitment methods for a large randomised trial in people with diabetes: A Study of Cardiovascular Events iN Diabetes (ASCEND). *Trials*. 2016;17(1):286. doi:10.1186/s13063-016-1354-9

73. MRC/BHF Heart Protection Study Webpage. http://www.hpsinfo.org/. Accessed January 14, 2016.

74. MRC/BHF Heart Protection Study Protocol. http://www.hpsinfo.org/protocol/section1.shtml. Accessed January 20, 2016.

75. Heart Protection Study Collaborative Group. MRC/BHF Heart Protection Study of cholesterol lowering with simvastatin in 20 536 high-risk individuals: a randomised placebo controlled trial. *Lancet*. 2002;360:23-33.

76. HPS3/TIMI55-REVEAL. www.revealtrial.org. Accessed May 17, 2016.

77. Townsend Deprivation Index. http://www.restore.ac.uk/geo-refer/36229dtuks00y19810000.php.

78. SEARCH Study Collaborative Group. Study of the effectiveness of additional reductions in cholesterol and homocysteine (SEARCH): Characteristics of a randomized trial among 12 064 myocardial infarction survivors. *Am Heart J*. 2007;154:815-823. doi:10.1016/J.AHJ.2007.06.034

79. Fry A, Littlejohns TJ, Sudlow C, et al. Comparison of Sociodemographic and Health-Related Characteristics of UK Biobank Participants With Those of the General Population. *Am J Epidemiol*. 2017;186(9):1026-1034. doi:10.1093/aje/kwx246

80. Clark TG, Altman DG, De Stavola BL. Quantification of the completeness of follow-up. *Lancet (London, England)*. 2002;359(9314):1309-1310.

81. Baigent C, Landray MJ, Reith C, et al. The effects of lowering LDL cholesterol with simvastatin plus ezetimibe in patients with chronic kidney disease (Study of Heart and Renal Protection): a randomised placebo-controlled trial. *Lancet (London, England)*. 2011;377(9784):2181-2192. doi:10.1016/S0140-6736(11)60739-3

82. Rogers CA, Welbourn R, Byrne J, et al. The By-Band study: gastric bypass or adjustable gastric band surgery to treat morbid obesity: study protocol for a multi-centre randomised controlled trial with an internal pilot phase. *Trials*. 2014;15(1):53. doi:10.1186/1745-6215-15-53

83. Moher D, Schulz KF, Altman DG. The CONSORT statement: revised recommendations for improving the quality of reports of parallel-group randomised trials. *Lancet (London, England)*. 2001;357(9263):1191-1194.

84. Edwards PJ, Roberts I, Clarke MJ, et al. Methods to increase response to postal and electronic questionnaires. *Cochrane Database Syst Rev*. 2009;(3). doi:10.1002/14651858.MR000008.pub4

85. Edwards L, Salisbury C, Horspool K, Foster A, Garner K, Montgomery AA. Increasing follow-up questionnaire response rates in a randomized controlled trial of telehealth for depression: three embedded controlled studies. *Trials*. 2016;17(1):107. doi:10.1186/s13063-016-1234-3

86. Morgan AJ, Rapee RM, Bayer JK. Increasing response rates to follow-up questionnaires in health intervention research: Randomized controlled trial of a gift card prize incentive. *Clin Trials*. 2017;14(4):381-386. doi:10.1177/1740774517703320

87. Pieper D, Kotte N, Ober P. The effect of a voucher incentive on a survey response rate in the clinical setting: a quasi-randomized controlled trial. *BMC Med Res Methodol*. 2018;18(1):86. doi:10.1186/s12874-018-0544-4

88. Cramer JA, Roy A, Burrell A, et al. Medication Compliance and Persistence: Terminology and Definitions. *Value Heal*. 2008;11(1):44-47. doi:10.1111/J.1524-4733.2007.00213.X

89. Farabee D, Shen H, Sanchez S. Program-Level Predictors of Antipsychotic Medication Adherence among Parolees. *Int J Offender Ther Comp Criminol*. 2004;48(5):561-571. doi:10.1177/0306624X04263884

90. Lam WY, Fresco P. Medication Adherence Measures: An Overview. *Biomed Res Int*. 2015;2015:1-12. doi:10.1155/2015/217047

91. Prado JC, Kupek E, Mion D. Validity of four indirect methods to measure adherence in primary care hypertensives. *J Hum Hypertens*. 2007;21(7):579-584. doi:10.1038/sj.jhh.1002196

92. Farmer KC. Methods for measuring and monitoring medication regimen adherence in clinical trials and clinical practice. *Clin Ther*. 1999;21(6):1074-1090. doi:10.1016/S0149-2918(99)80026-5

93. Stirratt MJ, Dunbar-Jacob J, Crane HM, et al. Self-report measures of medication adherence behavior: recommendations on optimal use. *Transl Behav Med*. 2015;5(4):470-482. doi:10.1007/s13142-015-0315-2

94. Morris LS, Schulz RM. Patient compliance-an overview. *J Clin Pharm Ther*. 1992;17(5):283-295. doi:10.1111/j.1365-2710.1992.tb01306.x

95. Nordentoft M, Thorup A, Albert N, et al. Gender differences in first-episode psychosis at 5-year follow-up – two different courses of disease? Results from the OPUS study at 5-year follow-up. *Eur Psychiatry*. 2014;29:44-51. doi:10.1016/J.EURPSY.2012.11.005

96. Rolnick SJ, Pawloski PA, Hedblom BD, Asche SE, Bruzek RJ. Patient Characteristics Associated with Medication Adherence. *Clin Med Res*. 2013;11(2):54-65. doi:10.3121/cmr.2013.1113

97. Eindhoven DC, Hilt AD, Zwaan TC, Schalij MJ, Borleffs CJW. Age and gender differences in medical adherence after myocardial infarction: Women do not receive optimal treatment – The Netherlands claims database. doi:10.1177/2047487317744363

98. Bulbulia R, Bowman L, Wallendszus K, et al. Effects on 11-year mortality and morbidity of lowering LDL cholesterol with simvastatin for about 5 years in 20 536 high-risk individuals: A randomised controlled trial. *Lancet*. 2011;378(9808):2013-2020. doi:10.1016/S0140-6736(11)61125-2

99. REVEAL trial protocol. doi:10.1056/NEJMoa1706444

100. Manteuffel M, Williams S, Chen W, Verbrugge RR, Pittman DG, Steinkellner A. Influence of Patient Sex and Gender on Medication Use, Adherence, and Prescribing Alignment with Guidelines. *J Women's Heal*. 2014;23(2):112-119. doi:10.1089/jwh.2012.3972

101. Taylor-Robinson SD, Toledano MB, Arora S, et al. Increase in mortality rates from intrahepatic cholangiocarcinoma in England and Wales 1968-1998. *Gut*. 2001;48(6):816-820. doi:10.1136/GUT.48.6.816

102. Nagel G, Unal H, Rosenbohm A, Ludolph AC, Rothenbacher D, ALS Registry Study Group. Implementation of a population-based epidemiological rare disease registry: study protocol of the amyotrophic lateral sclerosis (ALS)--registry Swabia. *BMC Neurol*. 2013;13:22. doi:10.1186/1471-2377-13-22

103. Francis NA, Hood K, Lyons R, Butler CC. Understanding flucloxacillin prescribing trends and treatment non-response in UK primary care: a Clinical Practice Research Datalink (CPRD) study. *J Antimicrob Chemother*. 2016;71(7):2037-2046. doi:10.1093/jac/dkw084

104. Doebler S, Ryan A, Shortall S, Maguire A. Informal care-giving and mental ill-health - differential relationships by workload, gender, age and area-remoteness in a UK region. *Health Soc Care Community*. 2017;25(3):987-999. doi:10.1111/hsc.12395

105. Beutler A, De La Motte SJ, DiStefano L, Cameron KL, Padua D, Gerber JP. Can a 10-minute injury prevention program decrease injuries in military cadets? A jump-acl study. *Clin J Sport Med*. 2013;23(2):123.

106. Long JA, Jahnle E, Richardson D. A randomized controlled trial of peer mentoring and financial incentive to improve glucose control in African American Veterans. *J Gen Intern Med*. 2011;26(SUPPL. 1):S165. doi:http://dx.doi.org/10.1007/s11606-011-1730-9

107. Bertoni AG, Effoe VS, Bollhalter L, et al. Translation of diabetes interventions into community settings: 12-month results of the lifestyle intervention for the treatment of diabetes trial. *Circulation*. 2017;135(Supplement 1).

108. Adams AS, Bayliss E, Schmittdiel JA, et al. The diabetes telephone study: Design and challenges of a pragmatic cluster randomized trial to improve diabetic peripheral neuropathy treatment. *Value Heal*. 2015;18(7):A723-A724.

109. Shah S, Brown T, Lee JY, Jean-Jacques M, Kandula NR. Individualized risk communication and lay outreach for the primary prevention of cardiovascular disease in community health centers: Preliminary results of a randomized controlled trial. *J Gen Intern Med*. 2014;29(SUPPL. 1):S126.

110. Lambrinou EE, Christoforou N, Kalogirou F, et al. Preliminary results of the study MEETinCY. *Eur J Heart Fail*. 2015;17(SUPPL. 1):398. doi:http://dx.doi.org/10.1002/ejhf.277

111. Maher N, Steen G, Fallon N, et al. Mortality rate and associated factors in older adults following hip fracture. *Ir J Med Sci*. 2014;183(7 SUPPL. 1):S351-S352. doi:http://dx.doi.org/10.1007/s11845-014-1177-1

112. Petersen JM, Jarbol DE, Hallas J, Storsveen MM, De Muckadell OBS. Can the use of computer decision support system prevent complicated ulcer among patients treated with nsaid or aspirin? A randomised controlled trial in general practice. *Gastroenterology*. 2017;152(5 Supplement 1):S222.

113. Ruddy KJ, Baker EL, Guo H, Goldstein MJ, Winer EP, Shulman LN. Randomized trial evaluating a coordinated survivorship care program for early stage breast cancer. *Cancer Res*. 2013;73(24 SUPPL. 1). doi:http://dx.doi.org/10.1158/0008-5472.SABCS13-OT2-5-02

114. Tran J, Waisbourd M, Weiss D, et al. The effectiveness of a tailored intervention to increase follow-up adherence in patients with glaucoma. *Investig Ophthalmol Vis Sci*. 2014;55(13):6095.

115. Takizawa K, Kakushima N, Tanaka M, et al. Phase II study of cold snare polypectomy for superficial non-ampullary duodenal epithelial tumor (D-CSP trial). *United Eur Gastroenterol J*. 2017;5(5 Supplement 1):A15. doi:http://dx.doi.org/10.1177/2050640617725668

116. Green BB, Wang C-Y, Anderson ML, Chubak J, Meenan RT, Vernon SW. Automated intervention with stepped increases in support to increase uptake of colorectal cancer screening: A randomized trial. *Ann Intern Med*. 2013;158(5 PART 1):301-311.

117. Greer JA, McMahon PM, Tramontano A, Gallagher ER, Pirl WF, Jackson VA. Effect of early palliative care on health care costs in patients with metastatic NSCLC. *J Clin Oncol*. 2012;30(15 SUPPL. 1).

118. Rosenheck R, Krystal J, Lew R, et al. Challenges in the design and conduct of controlled clinical effectiveness trials in schizophrenia. *Clin Trials*. 2011;8(4):529. doi:http://dx.doi.org/10.1177/1740774511413037

119. Anderson A, Casso D, Oliveria SA. Rare disease cohort using ambulatory electronic medical record (AEMR) data: Duchenne muscular dystrophy (DMD). *Pharmacoepidemiol Drug Saf*. 2017;26(Supplement 2):291. doi:http://dx.doi.org/10.1002/pds.4275

120. Akhavein RM, Sklenar J, Minnier J. Efficacy of clinical notifications to health care providers in management of heart failure patients. *Circulation*. 2015;132(SUPPL. 3).

121. Khambaty T, Callahan CM. Depression treatment and diabetes risk: A 9-year follow-up study of the impact trial. *Psychosom Med*. 2015;77(3):A31.

122. Urrea B, Kianoush S, Feldman DI, Blaha MJ. Long-term impact of the mActive physical activity trial on cardiometabolic health. *Eur Heart J*. 2016;37(Supplement 1):335. doi:http://dx.doi.org/10.1093/eurheartj/ehw432

123. Kenney B, Hark LA, Murchison AP, Bailey R, Haller JA. A comparison of three system-based interventions to improve follow-up adherence in a primary eye care setting. *Investig Ophthalmol Vis Sci*. 2014;55(13):6096.

124. Margolis KL, Fortmann SP, Bergdall AR, Smith DH, Pawloski PA. Design and methods for a pilot study and multicenter pragmatic trial comparing HCTZ and chlorthalidone for cardiovascular outcomes. *J Am Soc Hypertens*. 2016;10(4 Supplement):e49.

125. Tabibian JH. Long-term outcomes following a randomized study of oral antibiotics in primary sclerosing cholangitis. *Gastroenterology*. 2016;150(4 SUPPL. 1):S1069-S1070.

126. Teo EP, Tan WB, Chia PY, et al. Is nurse-led virtual monitoring of chronic hepatitis B patients a safe, feasible and cost-effective complement to standard care? *Hepatology*. 2016;64(1 Supplement 1):55A.

127. Berge E, Smith J, Sandercock P, et al. The effect of alteplase on health care utilisation and quality of life in Sweden, Norway and Scotland: Results from the 3rd international stroke trial (ist-3). *Eur Stroke J*. 2017;2(1 Supplement 1):477-478. doi:http://dx.doi.org/10.1177/2396987317706897

128. Buus K, Harbigg P, Olesen C. The effect of pharmaceutical care on three-year mortality a follow-up study. *Eur Geriatr Med*. 2013;4(SUPPL. 1):S182. doi:http://dx.doi.org/10.1016/j.eurger.2013.07.610

129. Huang Y, Johnson G, Kyriakides T. Ascertaining study participant safety using centralized electronic medical records in a clinical trial setting - lessons learned from the veterans affairs NEPHRON-D trial. *Trials*. 2017;18(Supplement 1). doi:http://dx.doi.org/10.1186/s13063-017-1902-y

130. Rollman B, Mazumdar S, Belnap BH, Houck P, Lenze E. Main outcomes from the relax trial of telephone delivered collaborative care for panic and generalized anxiety disorder. *J Gen Intern Med*. 2010;25(SUPPL. 3):S326. doi:http://dx.doi.org/10.1007/s11606-010-1338-5

131. Jones C, Jefferson E, Hogarth F, Littleford R. Supporting clinical trials through healthcare informatics. *Trials*. 2015;16(Supplement 2).

132. McAdam-Marx C, Ye X, Brixner DI, et al. A1C and weight outcomes at 18 months in patients with type 2 diabetes treated with exenatide in an ambulatory care setting. *Diabetes, Obes Metab*. 2009;11(12):1173-1174. doi:http://dx.doi.org/10.1111/j.1463-1326.2009.01163.x

133. Hernandez AF, Fleurence RL. The ADAPTABLE trial and PCORnet: Shining light on a new research paradigm. *Ann Intern Med*. 2015;163(8):635-636. doi:http://dx.doi.org/10.7326/M15-1460

134. Ellrodt AG, Fonarow GC, Schwamm LH, et al. Synthesizing lessons learned from get with the guidelines: The value of disease-based registries in improving quality and outcomes. *Circulation*. 2013;128(22):2447-2460. doi:http://dx.doi.org/10.1161/01.cir.0000435779.48007.5c

135. Spring B, Ockene JK, Gidding SS, et al. Better population health through behavior change in adults: A call to action. *Circulation*. 2013;128(19):2169-2176. doi:http://dx.doi.org/10.1161/01.cir.0000435173.25936.e1

136. Kesselheim AS. Introduction to a supplement on innovative approaches to studying health outcomes in rare diseases. *J Gen Intern Med*. 2014;29(SUPPL. 3):S709-S711. doi:http://dx.doi.org/10.1007/s11606-014-2921-y

137. Towfighi A. The (pre)hypertension limbo: Is it time to lower the treatment bar? *Neurology*. 2011;77(14):1322-1323. doi:http://dx.doi.org/10.1212/WNL.0b013e31823152e9

138. Kilbourne AM, Valenstein M. The research-to-practice gap in mood disorders: A role for the U.S. Department of Veterans Affairs. *J Clin Psychiatry*. 2007;68(4):502-504.

139. McCowan C. Commentary on vickers. *Clin Trials*. 2014;11(6):624-625. doi:http://dx.doi.org/10.1177/1740774514553483

140. Levy JF, Rosenberg MA. Innovative assessment of inpatient and pulmonary drug costs for children with cystic fibrosis. *Pediatr Pulmonol*. 2016;51(12):1295-1303. doi:http://dx.doi.org/10.1002/ppul.23554

141. Assmann KE, Andreeva VA, Jeandel C, Hercberg S, Galan P, Kesse-Guyot E. Healthy Aging 5 Years After a Period of Daily Supplementation With Antioxidant Nutrients: A Post Hoc Analysis of the French Randomized Trial SU.VI.MAX. *Am J Epidemiol*. 2015;182(8):694-704. doi:https://dx.doi.org/10.1093/aje/kwv105

142. Faulkner J, Stoner L, Lanford J, Jolliffe E, Mitchelmore A. Long-Term Effect of Participation in an Early Exercise and Education Program on Clinical Outcomes and Cost Implications, in Patients with TIA and Minor, Non-Disabling Stroke. *Transl Stroke Res*. 2017;8(3):220-227. doi:http://dx.doi.org/10.1007/s12975-016-0510-6

143. Whitbeck MG, Charnigo RJ, Khairy P, et al. Increased mortality among patients taking digoxin--analysis from the AFFIRM study. *Eur Heart J*. 2013;34(20):1481-1488. doi:https://dx.doi.org/10.1093/eurheartj/ehs348

144. Shah NB, Fenick AM. A Healthy Weight for Toddlers? Two-Year Follow-up of a Randomized Controlled Trial of Group Well-Child Care. *Clin Pediatr (Phila)*. 2016;55(14):1354-1357. doi:http://dx.doi.org/10.1177/0009922815623230

145. Adepoju OE, Bolin JN, Phillips CD, et al. Effects of diabetes self-management programs on time-to-hospitalization among patients with type 2 diabetes: A survival analysis model. *Patient Educ Couns*. 2014;95(1):111-117. doi:http://dx.doi.org/10.1016/j.pec.2014.01.001

146. Keijzers GB, Del Mar C, Geeraedts LM, Byrnes J. What is the effect of a formalised trauma tertiary survey procedure on missed injury rates in multi-trauma patients? Study protocol for a randomised controlled trial. *Trials*. 2015;16(1):215. doi:http://dx.doi.org/10.1186/s13063-015-0733-y

147. Cherkin D, Balderson B, Brewer G, et al. Evaluation of a risk-stratification strategy to improve primary care for low back pain: the MATCH cluster randomized trial protocol. *BMC Musculoskelet Disord*. 2016;17(1):361. doi:https://dx.doi.org/10.1186/s12891-016-1219-0

148. Thilly N, Chanliau J, Frimat L, et al. Cost-effectiveness of home telemonitoring in chronic kidney disease patients at different stages by a pragmatic randomized controlled trial (eNephro): rationale and study design. *BMC Nephrol*. 2017;18(1):126. doi:https://dx.doi.org/10.1186/s12882-017-0529-2

149. Rollman BL, Belnap BH, Mazumdar S, Abebe KZ, Karp J., Lenze EJ. Telephone-Delivered Stepped Collaborative Care for Treating Anxiety in Primary Care: A Randomized Controlled Trial. *J Gen Intern Med*. 2017;32(3):245-255. doi:http://dx.doi.org/10.1007/s11606-016-3873-1

150. Hannon TS, Dugan TM, Saha CK, McKee SJ, Downs SM, Carroll AE. Effectiveness of Computer Automation for the Diagnosis and Management of Childhood Type 2 Diabetes: A Randomized Clinical Trial. *JAMA Pediatr*. 2017;171(4):327-334. doi:https://dx.doi.org/10.1001/jamapediatrics.2016.4207

151. Persell SD, Friedberg MW, Meeker D, et al. Use of behavioral economics and social psychology to improve treatment of acute respiratory infections (BEARI): rationale and design of a cluster randomized controlled trial [1RC4AG039115-01]--study protocol and baseline practice and provider characterist. *BMC Infect Dis*. 2013;13:290. doi:https://dx.doi.org/10.1186/1471-2334-13-290

152. Robbins GK, Lester W. A clinical decision-support system with interactive alerts improved CD4 cell count in HIV. *Ann Intern Med*. 2013;158(8):JC11. doi:http://dx.doi.org/10.7326/0003-4819-158-8-201304160-02011

153. Anthony CA, Polgreen LA, Chounramany J, et al. Outpatient blood pressure monitoring using bi-directional text messaging. *J Am Soc Hypertens*. 2015;9(5):375-381. doi:http://dx.doi.org/10.1016/j.jash.2015.01.008

154. Carter BL, Vander Weg MW, Parker CP, Goedken CC, Richardson KK. Sustained Blood Pressure Control Following Discontinuation of a Pharmacist Intervention for Veterans. *J Clin Hypertens*. 2015;17(9):701-708. doi:http://dx.doi.org/10.1111/jch.12577

155. Hess EP, Wyatt KD, Kharbanda AB, et al. Effectiveness of the head CT choice decision aid in parents of children with minor head trauma: study protocol for a multicenter randomized trial. *Trials*. 2014;15(1):253. doi:http://dx.doi.org/10.1186/1745-6215-15-253

156. Stewart JC, Perkins AJ, Callahan CM. Effect of collaborative care for depression on risk of cardiovascular events: data from the IMPACT randomized controlled trial. *Psychosom Med*. 2014;76(1):29-37. doi:https://dx.doi.org/10.1097/PSY.0000000000000022

157. Sperl-Hillen J, Beaton S, Fernandes O, et al. Are benefits from diabetes self-management education sustained?. *Am J Manag Care*. 2013;19(2):104-112.

158. Adair R, Wholey DR, Christianson J, White KM, Britt H, Lee S. Improving chronic disease care by adding laypersons to the primary care team: a parallel randomized trial. *Ann Intern Med*. 2013;159(3):176-184. doi:https://dx.doi.org/10.7326/0003-4819-159-3-201308060-00007

159. Ford I, Murray H, McCowan C, Packard CJ. Long-term safety and efficacy of lowering low-density lipoprotein cholesterol with statin therapy 20-year follow-up of west of Scotland coronary prevention study. *Circulation*. 2016;133(11):1073-1080. doi:10.1161/CIRCULATIONAHA.115.019014

160. Lin CJ, Nowalk MP, Pavlik VN, et al. Using the 4 pillarsTM practice transformation program to increase adult influenza vaccination and reduce missed opportunities in a randomized cluster trial. *BMC Infect Dis*. 2016;16(1):623.

161. Carter BL, Farris KB, Abramowitz PW, et al. The Iowa Continuity of Care study: Background and methods. *Am J Health Syst Pharm*. 2008;65(17):1631-1642. doi:https://dx.doi.org/10.2146/ajhp070600

162. Lentine KL, Schnitzler MA, Xiao H, Brennan DC. Long-term safety and efficacy of antithymocyte globulin induction: use of integrated national registry data to achieve ten-year follow-up of 10-10 Study participants. *Trials*. 2015;16:365. doi:https://dx.doi.org/10.1186/s13063-015-0891-y

163. Anderson RT, Montori VM, Shah ND, et al. Effectiveness of the chest pain choice decision aid in emergency department patients with low-risk chest pain: Study protocol for a multicenter randomized trial. *Trials*. 2014;15(1):166. doi:https://dx.doi.org/10.1186/1745-6215-15-166

164. Baer HJ, Wee CC, DeVito K, et al. Design of a cluster-randomized trial of electronic health record-based tools to address overweight and obesity in primary care. *Clin Trials*. 2015;12(4):374-383. doi:https://dx.doi.org/10.1177/1a740774515578132

165. Denig P, Schuling J, Haaijer-Ruskamp F, Voorham J. Effects of a patient oriented decision aid for prioritising treatment goals in diabetes: pragmatic randomised controlled trial. *BMJ*. 2014;349:g5651. doi:https://dx.doi.org/10.1136/bmj.g5651

166. Persell SD, Lloyd-Jones DM, Friesema EM, Cooper AJ, Baker DW. Electronic health record-based patient identification and individualized mailed outreach for primary cardiovascular disease prevention: a cluster randomized trial. *J Gen Intern Med*. 2013;28(4):554-560. doi:https://dx.doi.org/10.1007/s11606-012-2268-1

167. Lewis CL, Adams J, Tai-Seale M, et al. A Randomized Controlled Effectiveness Trial for PSA Screening Decision Support Interventions in Two Primary Care Settings. *J Gen Intern Med*. 2015;30(6):810-816. doi:https://dx.doi.org/10.1007/s11606-015-3214-9

168. Lowrie R, Mair FS, Greenlaw N, et al. The Heart failure and Optimal Outcomes from Pharmacy Study (HOOPS): Rationale, design, and baseline characteristics. *Eur J Heart Fail*. 2011;13(8):917-924. doi:http://dx.doi.org/10.1093/eurjhf/hfr083

169. Marshall T, Caley M, Hemming K, Gill P, Gale N, Jolly K. Mixed methods evaluation of targeted case finding for cardiovascular disease prevention using a stepped wedged cluster RCT. *BMC Public Health*. 2012;12:908. doi:https://dx.doi.org/10.1186/1471-2458-12-908

170. Beaupre LA, Al-Houkail A. A Randomized Trial Comparing Ceramic-on-Ceramic Bearing vs Ceramic-on-Crossfire-Polyethylene Bearing Surfaces in Total Hip Arthroplasty. *J Arthroplasty*. 2016;31(6):1240-1245. doi:http://dx.doi.org/10.1016/j.arth.2015.11.043

171. Puschel K, Coronado G, Soto G, Gonzalez K, Martinez J, Holte S. Strategies for increasing mammography screening in primary care in Chile: Results of a randomized clinical trial. *Cancer Epidemiol Biomarkers Prev*. 2010;19(9):2254-2261. doi:http://dx.doi.org/10.1158/1055-9965.EPI-10-0313

172. Nguyen HQ, Bailey A, Coleman KJ, et al. Patient-centered physical activity coaching in COPD (Walk On!): A study protocol for a pragmatic randomized controlled trial. *Contemp Clin Trials*. 2016;46:18-29. doi:http://dx.doi.org/10.1016/j.cct.2015.10.010

173. Bleijenberg N, Drubbel I, Schuurmans MJ, Dam HT, Zuithoff NPA, Numans ME. Effectiveness of a Proactive Primary Care Program on Preserving Daily Functioning of Older People: A Cluster Randomized Controlled Trial. *J Am Geriatr Soc*. 2016;64(9):1779-1788. doi:http://dx.doi.org/10.1111/jgs.14325

174. Blair PS, Turnbull S, Ingram J, et al. Feasibility cluster randomised controlled trial of a within-consultation intervention to reduce antibiotic prescribing for children presenting to primary care with acute respiratory tract infection and cough. *BMJ Open*. 2017;7(5):e014506. doi:http://dx.doi.org/10.1136/bmjopen-2016-014506

175. Atkins S, Ojajarvi U, Talola N, Viljamaa M, Nevalainen J, Uitti J. Impact of improved recording of work-relatedness in primary care visits at occupational health services on sickness absences: study protocol for a randomised controlled trial. *Trials*. 2017;18(1):352. doi:http://dx.doi.org/10.1186/s13063-017-2076-3

176. Barry SJE, Dinnett E, Kean S, Gaw A, Ford I. Are Routinely Collected NHS Administrative Records Suitable for Endpoint Identification in Clinical Trials? Evidence from the West of Scotland Coronary Prevention Study. Pajewski NM, ed. *PLoS One*. 2013;8(9):e75379. doi:10.1371/journal.pone.0075379

177. Herbert A, Wijlaars L, Zylbersztejn A, Cromwell D, Hardelid P. Data Resource Profile: Hospital Episode Statistics Admitted Patient Care (HES APC). *Int J Epidemiol*. 2017;46(4):1093-1093i. doi:10.1093/ije/dyx015

178. Hscic. *Hospital Episode Statistics (HES) Analysis Guide*.; 2013.

179. Altman DG. *Practical Statistics for Medical Research*. Chapman and Hall; 1991.

180. Petretta M, Costanzo P, Perrone-Filardi P, Chiariello M. Impact of gender in primary prevention of coronary heart disease with statin therapy: A meta-analysis. *Int J Cardiol*. 2010;138(1):25-31. doi:10.1016/J.IJCARD.2008.08.001

181. Dale KM, Coleman CI, Shah SA, Patel AA, Kluger J, Michael White C. Impact of gender on statin efficacy. *Curr Med Res Opin*. 2007;23(3):565-574. doi:10.1185/030079906X167516

182. Thorn JC, Turner E, Hounsome L, et al. Validation of the Hospital Episode Statistics Outpatient Dataset in England. *Pharmacoeconomics*. 2016;34(2):161-168. doi:10.1007/s40273-015-0326-3

183. Campbell SE, Campbell MK, Grimshaw JM, Walker AE. A systematic review of discharge coding accuracy. *J Public Health (Bangkok)*. 2001;23(3):205-211. doi:10.1093/pubmed/23.3.205

184. Burns EM, Rigby E, Mamidanna R, et al. Systematic review of discharge coding accuracy. *J Public Health (Bangkok)*. 2012;34(1):138-148. doi:10.1093/pubmed/fdr054

185. CF START - A Cystic Fibrosis Randoised Registry Trial. http://www.cfstart.org.uk/. Accessed March 30, 2019.

186. Fröbert O, Lagerqvist B, Olivecrona GK, et al. Thrombus Aspiration during ST-Segment Elevation Myocardial Infarction. *N Engl J Med*. 2013;369(17):1587-1597. doi:10.1056/NEJMoa1308789

187. GBD 2013 Mortality and Causes of Death Collaborators. Global, regional, and national age-sex specific all-cause and cause-specific mortality for 240 causes of death, 1990-2013: a systematic analysis for the Global Burden of Disease Study 2013. *Lancet (London, England)*. 2015;385(9963):117-171. doi:10.1016/S0140-6736(14)61682-2

APPENDIX I – HPS SCREENING FORM

MRC **Heart Protection Study: SCREENING FORM**

ALTERATIONS & ADDITIONS (use CAPITALS)

IDENTITY LABEL (N.B. Remember to label ALL copies)

Title: Forename:
Surname:
Address:
Postcode:
Home telephone (inc. STD code): Sex:
Date of birth: Day: Month: Year:
GP's Name:
GP's Address:

INCLUSION CRITERIA: One or more **MUST** be **Yes** to be eligible

Yes No/unsure
- Heart attack (myocardial infarction: MI)
- Stable angina at any time
- Unstable angina (i.e. sudden increase in frequency and duration of angina, new angina at rest, or hospitalisation for angina)
- Ischaemic stroke (i.e. any stroke not known to be due to bleed)
- Transient cerebral ischaemic attack (TIA: symptoms <24 hrs)
- Coronary artery bypass surgery (CABG)
- Coronary angioplasty (PTCA)
- Carotid endarterectomy or carotid angioplasty
- Other arterial bypass surgery or angioplasty
- Amputation for vascular disease
- Other evidence of peripheral vascular disease (such as intermittent claudication: cramp-like pain in calves on walking)
- Diabetes mellitus
- Treated hypertension in male aged 65 or over

EXCLUSION CRITERIA: All **MUST** be **No** to be eligible

Yes No/unsure
- Any MI, stroke or hospitalisation for angina **in past 6 months**
- Severe heart failure (i.e. causing shortness of breath at rest)
- Severely disabling stroke
- Chronic liver disease (i.e. cirrhosis or hepatitis)
- Severe kidney (renal) disease
- Inflammatory muscle disease (i.e. dermatomyositis or polymyositis)
- Condition likely to result in organ transplant and/or the need for cyclosporin
- Regular use of contraindicated drugs (see back of form)
- Child-bearing potential: tick **No** for male, or for female who is post-menopausal, sterilised or using reliable form of contraception
- Other **life-threatening** non-vascular disease (i.e. thought likely by study nurse to have life-expectancy of less than 5 years)

OTHER MEDICAL HISTORY

Yes No/unsure (Put dash if unknown) Month Year
- Myocardial infarction — If **Yes**, date of **most recent** MI: /
- Hospitalisation for angina — If **Yes**, date of **most recent** admission: /
- Ischaemic stroke or TIA — If **Yes**, date of **most recent** event: /
- CABG, other arterial surgery or angioplasty — If **Yes**, date of **most recent** procedure: /
- Angina in the past 6 months — If **Yes**, current severity of angina (see over):
- Diabetes mellitus. If **Yes**,
  - (i) approximate age at diagnosis: years
  - (ii) insulin started within 1 year of diagnosis: Yes No
  - (iii) currently treated with insulin: Yes No
- Hypertension currently treated with antihypertensive drugs
- Currently treated with oral anticoagulant (see back of form)

OTHER SCREENING INFORMATION

Yes No/unsure
- Ever smoked cigarettes regularly (i.e. on most days for at least a year) — If **Yes**, age started cigarettes regularly: years of age
- Ex-cigarette smoker — If **Yes**, age last smoked cigarettes regularly: years of age
- Current cigarette smoker — If **Yes**, number of cigarettes smoked: /day
- Current pipe or cigar smoker
- Currently drink alcohol once a month or more — If **Yes**, amount (see over): Spirits/liqueurs measures/week; Wine glasses/week; Beer/cider pints/week

Which of the following best describes ethnic origin (tick **one** box only):
White; West Indian or Guyanese; Indian, Pakistani or Bangladeshi; African; Japanese; Chinese; Arab; Mixed; Other

CLINICAL ASSESSMENT

Systolic BP: mm Hg Weight: kg Waist: cm Cholesterol: mmol/l
Diastolic BP: mm Hg Height: cm Hip: cm or Low High

ELIGIBILITY AND CONSENT

Yes No/unsure
- Likely to be any serious problem with regular clinics?
- Is patient eligible? If **No**, **STOP** here and **SIGN** form

Yes Refused Pending
- Has informed consent to study been given?

If **Refused, STOP** here and **SIGN** form; if **Pending**, wait until

AT CONSENT PENDING VISIT (if applicable)

Yes No
- Has informed consent to study **now** been given? If **No**, **STOP** here and **SIGN** form

RUN-IN TREATMENT ALLOCATION

Yes No
- Has a Run-in treatment pack been given to eligible and consenting patient?

Signature of clinic nurse:
& **PRINTED** name:
Today's date: / / (Day / Month / Year)

For any patient started on Run-in: Continue in study? Yes No
Signature of clinician:
& **PRINTED** name:

SEND **TOP** COPY (and BLOOD SAMPLES if on Run-in or Pending) TO COORDINATING CENTRE.
Retain middle copy for patient's clinic file (and, if required, bottom copy for pharmacy)

HPS/S/2/0696

APPENDIX II – REVEAL ELIGIBILITY CRITERA

Inclusion criteria

Patients must be aged at least 50 at the time of initial invitation, and at least one of the following inclusion criteria must be satisfied:

- History of myocardial infarction (MI);
- Cerebrovascular atherosclerotic disease (history of presumed ischaemic stroke or carotid revascularization);
- Peripheral arterial disease (history of non-coronary revascularization, including aortic aneurysm repair or graft); or
- Diabetes mellitus with other evidence of symptomatic coronary heart disease (treatment or hospitalization for angina, or a history of coronary revascularization or acute coronary syndrome).

Exclusion criteria

- Acute MI, acute coronary syndrome or stroke within 4 weeks prior to Screening Visit or during run-in;
- Planned coronary revascularization procedure within the next 6 months;
- Definite history of chronic liver disease, or abnormal liver function (i.e. ALT >2x ULN)*;
- Severe renal insufficiency (i.e. creatinine >2.3 mg/dL [200 µmol/L], dialysis or functioning renal transplant);
- Evidence of active inflammatory muscle disease, or CK >3x ULN;
- Previous significant adverse reaction to a statin or anacetrapib;
- Current treatment with any of the following lipid-lowering treatments:
  - (i) a regimen considered to produce substantially greater LDL cholesterol reduction than atorvastatin 20 mg daily for individuals in China or 80 mg daily for those in other countries;
  - (ii) fibric acid derivative ("fibrate", including gemfibrozil); or
  - (iii) niacin (nicotinic acid) at doses above 100 mg daily;
- Concurrent treatment with a medication that is contraindicated with anacetrapib or atorvastatin†;
- Known to be poorly compliant with clinic visits or prescribed medication;
- Medical history that might limit the individual's ability to take trial treatments for the duration of the study;
- Women of child-bearing potential (unless using adequate contraception); or
- Current participation in a clinical trial with an unlicensed drug or device.

Individuals were also to be excluded at the Screening visit if it was considered unlikely that they would achieve total cholesterol <135 mg/dL (3.5 mmol/L) on the highest atorvastatin dose available in their region‡.

Additional exclusion criteria at the randomization visit§

In addition to the above, individuals must be excluded at the randomization visit if any of the following are true:

- Total cholesterol above 155 mg/dL (4 mmol/L)
- Non-compliant with run-in treatment (<90% scheduled run-in medication taken)‖
- Individual is no longer willing to be randomized into the 4-5 year trial
- The individual's doctor is of the view that their patient should not be randomized

ALT – alanine transaminase; CK – creatine kinase; LDL – low density lipoprotein; ULN – upper limit of normal
* Individuals with a history of acute hepatitis are eligible provided this ALT limit is not exceeded
† Individuals who were taking such drugs temporarily could be re-screened when they discontinue them, if considered appropriate.
‡ atorvastatin 20 mg daily in China or 80 mg daily in other countries
§ The randomization visit occurred after an 8-12 weeks run-in period during which participants received study atorvastatin plus placebo anacetrapib
‖ Tablet counts were not required

APPENDIX III – HPS INCLUSION CRITERIA BY SEX

Inclusion criteria	Male n(%)	Female n(%)
Myocardial Infarction (MI)	13315 (44%)	4325 (34%)
Stable angina at any time	14601 (49%)	6280 (49%)
Unstable angina	6349 (21%)	2972 (23%)
Ischaemic stroke	3028 (10%)	1222 (9%)
Transient cerebral ischaemic attack (TIA)	2099 (7%)	928 (7%)
Coronary artery bypass surgery (CABG)	5423 (18%)	954 (7%)
Coronary angioplasty (PTCA)	2109 (7%)	598 (5%)
Carotid endarterectomy or carotid angioplasty	489 (2%)	206 (2%)
Other arterial bypass surgery or angioplasty	3466 (12%)	1036 (8%)
Amputation for vascular disease	320 (1%)	72 (1%)
Other evidence of peripheral vascular disease	8439 (28%)	4042 (31%)
Diabetes	8402 (28%)	4597 (36%)
Treated hypertension in male aged 65 or over	2085 (29%)	n/a
Total	29482 (99%)	12437 (98%)

Calculated as proportion of eligible patients (N=42 471, male n=29 779, female=12 692)