Can routinely collected data be used to reliably follow-up participants in large randomised trials?

Nuffield Department of Population Health

Mr Charlie Harper

BSc, MSc

Wolfson College, University of Oxford

Thesis submitted for the degree of

Doctor of Philosophy

2023

It feels like a lifetime ago since I started my DPhil in autumn 2017, it has encompassed the very best of times (getting married and my two children Heidi and Alex being born) and the very worst of times (being diagnosed with and treated for granulomatosis with polyangiitis with subglottic stenosis). I want to dedicate this thesis to my family, in particular my wife Jen, who has been my rock throughout these past 6 years. Without her support I can't imagine ever having been able to finish my studies.

Acknowledgements

Thank you to the participants who took part in the ASCEND and SHARP randomised trials, without whom my research would not have been possible. Thank you also to all the people who contributed to conducting these trials.

Thank you to my supervisors: Associate Professor Will Herrington, Associate Professor Natalie Staplin, Professor Colin Baigent and Professor Martin Landray. I am incredibly grateful for their support and advice, particularly after returning from each period of sick leave, in which they helped me prioritise my goals and push forward with my DPhil.

Thank you to the ASCEND investigators: Jane Armitage, Louise Bowman, Sarah Parish, and Marion Mafham. Thank you also to the ASCEND Steering Committee: Amanda Adler, Theingi Aung, Colin Baigent, Jonathon Bodanksky, Rory Collins, Andrew Farmer, Richard Haynes, Andrew Neil, Nilesh Samani, Richard Peto, and David Simpson. And co-authors from the data management team on thesis publications: Will Stevens and Karl Wallendszus. It has been a privileged to have been involved with this large innovative trial. The advice from the team was a key factor facilitating the development of useful analyses which I believe have important methodological implications for researchers conducing trials.

Finally, thank you to the wonderful community in the Nuffield Department of Population Health, Health Data Research UK and the Medical Research Council Network of Hubs for Trials Methodology Research for their support and funding of this project.

ABSTRACT

Background

Large randomised controlled trials (RCTs) in cardiovascular disease have provided reliable evidence for interventions whose widespread use has contributed to the secular declines in mortality. But, increasing costs of conducting RCTs threaten our ability to generate new randomised evidence. Linkage to health data collected during a person's routine care (i.e. routine data) has the potential to substantially streamline trial follow-up, however, there remains considerable uncertainty about the reliability of such information.

The primary aims of this thesis were to investigate, whether for large cardiovascular RCTs in the UK, follow-up of their primary and secondary outcomes via routinely collected hospital admission and death registry data: (1) are ascertained as accurately and completely as adjudicated direct follow-up; and (2) can yield the same estimated treatment effects as adjudicated direct follow-up. Secondary aims were: (3) to assess whether clinical adjudication of direct trial follow-up is necessary; and (4) to use simulated data to investigate more generally the impact on randomised trials of removing true events and adding false events, as may happen if routine data are used for follow-up.

Methods

A systematic review of the existing literature was conducted, where MEDLINE and Embase were searched between 1st January 2010 and 12th August 2022 using a search strategy and selection criteria to identify studies assessing the reliability of UK routine data at ascertaining cardiovascular outcomes compared to trial follow-up data.

Analyses were then carried out using two large trials: (1) the ASCEND (A Study of Cardiovascular Events in Diabetes) primary prevention trial randomised 15480 UK people with diabetes using a 2x2 factorial design to aspirin versus matching placebo, and separately to omega-3 fatty acids versus placebo. The primary efficacy outcome was

serious vascular events (SVEs; including non-fatal myocardial infarction [MI], ischaemic stroke or transient ischaemic attack [TIA], or vascular death [excluding intracranial haemorrhage]) and primary safety outcome for the aspirin comparison was major bleeding (including intracranial haemorrhage, sight-threatening eye bleeding, serious gastrointestinal bleeding, and other major bleeding). Adjudicated direct participant mail-based follow-up was the main source of outcome data, and participants were linked to their routinely collected hospital admission and death registry data, where an algorithm was applied to distinguish between major vs minor bleeding events. (2) SHARP (Study of Heart and Renal Protection) was a placebo-controlled trial of simvastatin plus ezetimibe in 9270 people with moderate-to-advanced chronic kidney disease (CKD). The primary outcome was major atherosclerotic events (MAEs; including non-fatal MI, coronary death, non-haemorrhagic stroke, or revascularisation), with the secondary outcome being kidney replacement therapy (KRT; including maintenance dialysis and kidney transplantation). Adjudicated direct participant clinic-based follow-up was the main source of outcome data, and 1576 participants living in England were linked to their routine data. Kappa statistics were used to assess agreement between follow-up data sources (where appropriate), and randomised comparisons were re-run using the outcomes dataset based on the different data sources.

Finally, simulation methods were used to investigate the impact on randomised trials more generally of removing true events and adding false events, as may happen if routine data are used for follow-up. Initially randomised trials were generated to match those of ASCEND's (n=15480) aspirin comparisons for any SVE (risk ratio=0.88) and major bleeding (risk ratio=1.29). Scenarios using smaller (n=1000 and 5000) and even larger (n=25000) sample sizes were then considered with different treatment effects (RR 0.70, 0.80, 0.90, 1.10, 1.20, 1.30). These data were considered the ground-truth. To simulate routine data follow-up, true events were removed (0% to 40%) and/or false events added (0% to 40%), where participants in either arm had an equal probability of having an error occurring (i.e.

errors were random rather than systematic). The primary performance measure was the mean difference between the ground-truth and routine data risk ratios, with the secondary performance measure being the difference in power.

Results

The systematic review identified three studies that compared cardiovascular and bleeding outcomes ascertained via UK routine data with adjudicated direct follow-up, two of which assessed old routine data collected over two decades ago (i.e. 1989-2001), and the other followed-up participants for only 28 days. Nevertheless, for cardiovascular outcomes, agreement between data sources was strong for revascularisations, and moderate for MIs and strokes. Re-running randomised comparisons using routine data only provided relative treatment effect estimates similar to adjudicated direct follow-up. For major bleeding events, routine data showed moderate completeness but accuracy was low, because the study's routine data outcome definition included major/minor bleeding and anaemia events. No studies were identified that assessed kidney disease outcomes.

In the ASCEND trial, agreement between modern routine data (i.e. 2005-2017) and adjudicated direct follow-up was strong for SVEs (kappa 0.78, 95% confidence interval [CI] 0.76-0.80), with similar levels of agreement for all SVE components except TIA (kappa 0.43, 95% CI 0.36-0.49). For the aspirin randomised comparison, rate ratios (RRs) for SVEs were similar for adjudicated direct follow-up versus routine data alone (adjudicated data: RR 0.88, 95% CI 0.79-0.97; routine data: RR 0.91, 95% CI 0.81-1.02), and almost identical for SVEs excluding TIA (adjudicated data: RR 0.92, 95% CI 0.82-1.03; routine data: RR 0.91, 95% CI 0.80-1.02). For the major bleeding outcome, despite only moderate agreement between datasets (kappa 0.53, 95% CI 0.49-0.57), RRs and absolute treatment effect estimates were similar for adjudicated direct follow-up and routine data (adjudicated data: RR 1.29, 95% CI 1.09-1.52; absolute excess +6.3/5000 person-years [mean standard error ±2.1]; routine data: RR 1.21, 95% CI 1.03-1.41; absolute excess +5.0/5000[±2.2]). Overall, the interpretation of ASCEND's aspirin randomised comparison (i.e. the benefits of aspirin

being largely counterbalanced by the bleeding hazard) would have not changed had follow-up been solely via routinely collected data, with similar findings for pre-adjudicated mail-based follow-up.

For the SHARP trial in people with moderate-to-severe CKD, agreement between routine data and adjudicated follow-up for MAEs was moderate-to-strong (kappa 0.67, 95% CI 0.61-0.72) and very strong for KRT (kappa 0.87, 95% CI 0.84-0.90). Analysis of SHARP's pre-adjudicated data showed that about one-quarter of patient reported MAEs were refuted. Randomised analyses using SHARP's pre-adjudicated data found similar results to the adjudicated data (pre-adjudication: RR 0.80, 95% CI 0.72-0.89; adjudicated: RR 0.83, 95% CI 0.74-0.94) and also suggested refuted MAEs were likely to represent atherosclerotic disease (refuted MAEs: RR 0.80, 95% CI 0.65−1.00).

Analysing simulated trial data from a wide range of trial scenarios showed that routine data sources with small-to-moderate random errors (i.e. true event removed and false events added ≤20%, which are equally distributed between trial arms) have little impact on relative treatment effect estimates in randomised trials. In the situation where routine data sources do introduce large amounts of error into a trial (i.e. true event removed and false events added >20%), larger trials (n≥5000) are better protected against random errors biasing the treatment effect towards the null and reducing trial power.

Conclusions

In summary, data from large cardiovascular trials in people with diabetes and CKD indicate that linkage to routinely collected UK hospital admission and death registry data may be a reliable method of streamlining follow-up for non-fatal MI, hospitalised ischaemic stroke, vascular death (including its subtypes), arterial revascularisation, kidney replacement therapy, and major bleeding outcomes. More generally, analysis of real and simulated randomised trial data consistently showed that small-to-moderate random errors in trial follow-up (which can be expected whatever the method of follow-up) had little impact on

relative treatment effect estimates on account of the protection afforded by large-scale randomisation. This thesis provides a rationale to consider using UK routine data as the sole source of follow-up for future cardiovascular trials.

Statement of contribution to work

The work in this thesis is my own work, and was supervised by Associate Professor Will Herrington, Associate Professor Natalie Staplin, Professor Martin Landray, and Professor Colin Baigent. I planned and conducted data preparation, data analysis, interpretation, and writing.

I was not involved in the design or conduct of the ASCEND (A Study of Cardiovascular Events in Diabetes) and SHARP (Study of Heart and Renal Protection) randomised trials, and was provided access to the trial data once the trials were completed. For the analyses using data from the ASCEND trial, in addition to my supervisors I also received comments and advice from the ASCEND investigators (Jane Armitage, Louise Bowman, Sarah Parish, and Marion Mafham), ASCEND Steering Committee (Amanda Adler, Theingi Aung, Colin Baigent, Jonathon Bodanksky, Rory Collins, Andrew Farmer, Richard Haynes, Andrew Neil, Nilesh Samani, Richard Peto, and David Simpson), and other co-authors (Will Stevens and Karl Wallendszus). For the SHARP analyses, in addition to my supervisors I also received comments from SHARP Steering Committee members: Richard Haynes, Jonathan Emberson, Christina Reith, Lai Seong Hooi, Adeera Levin, and Christoph Wanner.

I am aware of and understand the university's policy on plagiarism. I certify that this thesis is my own work, except where indicated by references. The work presented in it has not been submitted in support of another degree or qualification from this or any other university or institute of learning.

Outputs

Much of the work in this thesis has been published in peer-reviewed journals. These publications are listed below.

All analyses in **CHAPTER 3** has been published in JAMA Network Open:

> Harper C, et al. Comparison of the Accuracy and Completeness of Records of Serious Vascular Events in Routinely Collected Data vs Clinical Trial–Adjudicated Direct Follow-up Data in the UK: Secondary Analysis of the ASCEND Randomized Clinical Trial. JAMA Network Open 2021; 4(12): e2139748-e.

All analyses in **CHAPTER 4** has been published in BMJ Heart:

> Harper C, et al. Reliability of major bleeding events in UK routine data vs clinical trial adjudicated follow-up data. BMJ Heart 2023; 109: 1467-1472.

The pre-adjudicated direct follow-up analysis in **CHAPTER 5** has been published in Kidney International Reports:

> Herrington WG*, Harper C*, et al. Impact of outcome adjudication in kidney disease trials: observations from the Study of Heart and Renal Protection (SHARP). Kidney International Reports 2023; 8(8): 1489-1495. **Joint first authors.*

The results from this thesis have also contributed to the Standardising Outcomes in Healthcare Systems Data (SCORE-CVD) project run by the British Heart Foundation (BHF) Data Science Centre, which aims to develop standardised definitions for cardiovascular outcomes ascertained from routinely collected health data for UK randomised trials.

Funding and resources

This DPhil was funded by a PhD studentship award from the Medical Research Council Network of Hubs for Trials Methodology Research (HTMR) MR/L004933/2.

ASCEND (Current Controlled Trials number: ISRCTN60635500; ClinicalTrials.gov number: NCT00135226) was funded by the British Heart Foundation and by Bayer (Germany and the United States), Solvay, Abbott, and Mylan.

SHARP (Current Controlled Trials number: ISRCTN54137607; ClinicalTrials.gov number: NCT00125593) was funded by Merck & Co., Inc., Whitehouse Station, NJ USA, with additional support from the Australian National Health Medical Research Council, the British Heart Foundation, and the UK Medical Research Council.

Contents

List of abbreviations and acronyms

A&E - Accident and emergency

ASCEND - A Study of Cardiovascular Events in Diabetes

BHF – British Heart Foundation

CI - Confidence interval

CHD - Coronary heart disease

CKD - Chronic kidney disease

CV - Cardiovascular

CVD - Cardiovascular disease

FEA - False events added

HEART-PPCI - How Effective are Antithrombotic Therapies in Primary Percutaneous Coronary Intervention

HPS - Heart Protection Study

HR - Hazard ratio

GP - General Practice

ICD-10 - International Statistical Classification of Diseases and Related Health Problems (10th revision)

KRT - Kidney replacement therapy

MAE - Major atherosclerotic event

MCE - Major coronary event

MI - Myocardial infarction

MRC - Medical Research Council

MVE - Major vascular event

N - Number of participants

NHS - National Health Service

NPV - Negative predictive value

ONS - Office for National Statistics

OPCS-4 - Office of Population Censuses Surveys Classification of Surgical Operations and Procedures (4th revision)

PEDW - Patient Episode Database Wales

PPV - Positive predictive value

RCT - Randomised controlled trial

RRR – Relative risk reduction

SHARP - Study of Heart and Renal Protection

SMR - Scottish Morbidity Records

SVE - Serious vascular event

TER - True events removed

TIA - Transient ischaemic attack

UK - United Kingdom

WOSCOPS - West of Scotland Coronary Prevention Study

List of tables

List of figures

CHAPTER 1 Routinely collected health data and its potential to streamline randomised trial follow-up

Chapter summary

Randomised controlled trials (RCTs) are considered the gold-standard in assessing the efficacy and safety of medical interventions, but RCTs are becoming increasingly more expensive and complex to conduct. Current trial methods for ascertaining adverse event information via direct participant follow-up in research clinics are highly burdensome for both staff and participants alike. Verification of participant reports using collected medical notes by clinician adjudicators is another important cost. The utilisation of routinely collected health data (i.e. coded health information recorded during routine care) has the potential to substantially reduce costs of trial conduct, by reducing both the dependence on large numbers of research staff performing face-to-face visits with participants and the amount of medical documentation needing to be manually collected at local sites for the purposes of adjudication. The United Kingdom is particularly well placed to use such relatively novel methods as it has a National Health Service in which health data are collected nationally. These rich datasets are widely used as the primary source of follow-up for large observational studies, yet few UK RCTs have utilised such data. One potential reason for this lack of uptake is the uncertainty around whether RCTs can still provide reliable results if participants are followed-up via routinely collected health data sources. This thesis aims to investigate whether, for large cardiovascular RCTs in the UK, follow-up via routinely collected health data are reliable enough to be the sole method of follow-up.

1.1 Randomised controlled trials

Large-scale randomised controlled trials (RCTs) are considered the gold-standard method of evaluating the efficacy and safety of medical interventions with moderate treatment effects, as the process of randomly allocating large numbers of study participants to a particular intervention removes confounding, which cannot be definitively excluded in observational studies.[1-3] Risks of bias are further minimised by many RCTs using a double-blind approach in which both trial participants and research staff are blinded to treatment allocation.

1.1.1 Randomised controlled trial in cardiovascular disease

Cardiovascular disease (CVD; i.e. conditions affecting the heart or blood vessels) is one of the leading causes of death worldwide.[4] RCTs in CVD (i.e. cardiovascular trials) commonly assess (1) whether a particular intervention may prevent future cardiovascular events from occurring (i.e. efficacy), and (2) whether the intervention is safe (i.e. minimal increased risk of adverse events). The primary outcome for cardiovascular trials traditionally include CVD events that are considered the most serious (i.e. results in death, are life threatening, results in disability, or require hospitalisation), herein termed Major Cardiovascular Events (MVE).[3,5] MVEs are defined in this thesis as including non-fatal myocardial infarctions (MI), non-fatal strokes (including both ischaemic and haemorrhagic), deaths due to CVD as the underlying cause (including subtypes), and arterial revascularisations (i.e. commonly subcategorised into coronary and non-coronary).[6]

Depending on the particular research question a trial is aiming to investigate, its primary outcome may differ from the broad definition of MVE defined above. For example, in the ASCEND[7-9] trial (A Study of Cardiovascular Events in Diabetes; the focus of **CHAPTER 3** and **CHAPTER 4**) that assessed the effects of aspirin on preventing future serious vascular events (SVEs); due to bleeding being a known risk factor for antithrombotic treatment, haemorrhagic stroke was excluded from the SVE composite outcome. Additionally, as the

annual event rate for SVEs was lower than anticipated, to improve the trial's power the composite outcome was expanded to include transient ischaemic attack events.[7] Similarly, in the SHARP[10,11] trial (Study of Heart and Renal Protection; the focus of **CHAPTER 5**) that assessed the effects of simvastatin plus ezetimibe on preventing future major atherosclerotic events (MAEs) in people with moderate-to-severe chronic kidney disease (CKD); due to initiation of kidney replacement therapy being an important secondary outcome, revascularisations related to dialysis access were excluded from the MAE composite outcome.

1.1.2 Follow-up methods for cardiovascular trials

For RCTs in CVD, to capture information about cardiovascular and safety events the trial team follow-up participants after the date of randomisation, typically until a final follow-up visit or death. Traditionally, trial follow-up procedures performed by the Oxford Clinical Trial Service and Epidemiological Studies Unit (CTSU) involves participants attending regular appointments at their local research clinic (e.g. SHARP which is the focus of **CHAPTER 5**), in which a member of the trained local research team (usually a nurse) interviews the participant assessing whether they have experienced any new adverse events since their previous follow-up visit (**Figure 1.1**).[10,12-15] When participants are unwilling or unable to attend these face-to-face appointment or have died (**Figure 1.2**), information is obtained via telephone from them, their relative, carer, or medical professional. This information is recorded and coded in real-time directly into clinical terms on a computer-system and in this thesis is referred to as "clinic-based follow-up". These methods differ from the more resource intensive approach of "investigator-reported follow-up" used by other research organisations, where trained physicians at the research site (i.e. investigators) report events (based on clinical presentation) that they consider to be serious adverse events (including all suspected primary and secondary outcomes).[16]

For RCTs conducted by Oxford CTSU, the trial protocol pre-defines which events require verification. Medical notes (e.g. a participant's hospital discharge summary and/or death

certificate) are then collected for such events by the local research team and sent to the study central coordinating office (CCO) as source documents. A trained clinician at the CCO, blinded to treatment allocation, then assesses the accuracy of the report by reviewing collected medical notes about the event against predefined outcome criteria. The outcome of such a review is classed in one of three ways: confirmed, where the medical notes confirm that the reported event occurred; refuted, where the adjudicator finds that the reported event either did not to occur or was misclassified; and unconfirmed, where based on the available documentation the adjudicator can neither confirm nor refute the event. Typically trials place an emphasis on a combination of events that are either confirmed or unconfirmed, commonly termed unrefuted events. This is the process of adjudication, and in this thesis is referred to as "adjudicated clinic-based follow-up".

Some more recent Oxford CTSU trials (e.g. ASCEND which is the focus of **CHAPTER 3** and **CHAPTER 4**) have undertaken a more streamlined remote approach, where instead of clinic-based interviews, participants are mailed questionnaires assessing whether they have experienced a new adverse event, this is referred to as "mail-based follow-up" and "adjudicated mail-based follow-up" after clinical verification. When describing cardiovascular trial follow-up more broadly, in this thesis "adjudicated direct follow-up" refers to any form of trial follow-up where the participant or their family/carer/medical professional are contacted.

1.1.3 The rising costs of conducting cardiovascular trials

Cardiovascular trials have provided reliable evidence for interventions whose widespread use, in conjunction with increases in smoking cessation and dietary changes, have contributed to the secular declines in mortality due to CVD.[17] For example, in the United Kingdom (UK) from 1950 to 2010, risk of death from vascular disease in middle-aged men (35-69 years) has decreased from 22% to 6%.[4] But with rising levels of obesity and sedentary lifestyle behaviours, such improvements in cardiovascular health are at risk.[18] Despite the importance of RCTs in reducing the global burden of disease, the rising costs

of conducting trials threaten our ability to generate new randomised evidence.[19,20] Phase three trials in CVD are the third most costly therapeutic area, with the average per-study costs equating to about $25.2 million.[21] Clinic-based follow-up has become particularly burdensome and costly, where for a large international trial, hundreds of research staff may need to be employed in order to both perform face-to-face visits with participants and manually collect medical documentation about selected adverse events, with further costs attributed to the verification of participant-reports by clinician adjudicators.[22] In addition, due to incremental improvements in medical interventions having led to CVD events being less common, cardiovascular trials now need to be much larger to have appropriate statistical power to detect moderate treatment effects.[4] For future cardiovascular trials to be feasible, researchers are now having to consider methods to streamline trial follow-up; one such potential approach is utilising health data already being recorded during a participant's routine care, commonly referred to as routinely collected health data, herein referred to as "routine data".

1.2 Routinely collected health data

Routine data are commonly defined as coded health information recorded by the local health care system during a person's routine care; these data are recorded not for research, but primarily for other purposes such as patient care, reimbursement, operational data, and audits.[23] These data differ from a hospital's "Electronic Patient Records", which generally refers to the coded or free-text information recorded by medical professionals during routine care and which is accessed by local professionals using their institution's computer systems. Routine data which could be particularly valuable for trialists include: data recorded within hospitals (i.e. secondary care data); data recorded outside of hospital in community care (e.g. primary care data); disease-specific or mortality registries (e.g. cancer, deaths); and nationwide audits (e.g. National Diabetes Audit, Sentinel Stroke National Audit Programme). In some cases the recording of information may be carried out by doctors/nurses, for example during general practice consultations, while for other data

such as routinely collected hospital admission records, trained clerical staff (i.e. clinical coders) transcribe patient discharge summaries into coded information. Examples of information recorded in these routine datasets include: medical diagnoses and procedures; vital sign measurements; laboratory results; and prescribing of medications.

1.2.1 Real-world evidence

"Real-world data", a descriptive term for routine data, are increasingly being used to make inferences about the safety and efficacy of interventions. Such approaches rely on large datasets with information on medical intervention use in routine clinical practice. "Real-world evidence" is used to refer to observational associations between use of the intervention and outcomes, with inferences about the effects of the intervention extrapolated from these associations. Such non-randomised analyses cannot fully account for confounding. For example, use of interventions are not decided at random and there may be systematic and important differences between those who receive an intervention vs those who do not. Statistical approaches to control for such biases cannot guarantee residual confounding is controlled, and so such studies are much less suitable to guide clinical decision making for interventions with moderate effects than adequately-powered RCTs.[2,3]

1.2.2 Utilisation of routine data for participant follow-up in RCTs

Routine data could instead be used to follow-up participants in RCTs, and in particular, to ascertain information about important trial efficacy and safety events occurring after randomisation. For example, in cardiovascular trials, MVEs could potentially be ascertained through linkage to routinely collected hospital admission and death registry data. These methods have the potential to substantially streamline trial follow-up as routine data could substitute for clinic-based follow-up (when face-to-face appointments are not needed), thus requiring far fewer research staff costs,[24] while preserving the key methodological benefit of random allocation to interventions.

Although routine data have been collected electronically for several decades, one of the main challenges for trialists is access to such data.[24] Data are commonly stored by individual hospitals, primary care organisations, ambulance services, disease registries, and insurance providers. The time and cost of arranging data sharing agreements with each individual health organisation, even in a small local region, can be incredibly resource intensive.[25] These barriers can be somewhat alleviated in the United Kingdom (UK) as the UK's National Health Service (NHS) provides almost all healthcare services to the population, free at the point of care. Each devolved nation (excluding Northern Ireland) centrally collects much of its routine data in standardised coded format, meaning a UK-wide trial would need only a handful of data sharing agreements to enable linkage to participants living in England, Scotland and Wales.[26,27]

1.2.3 Routine datasets available in the UK

Currently in England, routine datasets available to researchers from NHS England (the national data provider) include: death registration; coded hospital data (including admissions, outpatient appointments, A&E attendances, diagnostic imaging, patient reported outcomes, and maternity records); community medication dispensing data, mental health services, and national diabetes audit data (**Figure 1.3**).[28] Primary care data from GP practices and secondary care prescribing data are available, but only to trials investigating treatments for COVID-19. As the majority of these nationally collected routine datasets have either only recently been established, do not record clinical information, or are not available in all devolved nations, this thesis focuses on the two longest established datasets: routine hospital admission data and death registration data.[29-33] Both datasets have been in operation since at least the early 1990s and have a long history of nationally standardised clinical coding practices.[34-37]

Death registry data (available from NHS England for England/Wales and National Health Scotland Central Register) includes all deaths in the UK irrespective of location of death.[32,33] Every death is certified by a medical professional using the Medical Certificate of Cause of

Death. The form includes parts I (a) to (d), with the immediate cause of death in 1(a), with any antecedent direct causes recorded in 1(b-d) and other contributory causes recorded in part II. This certificate is then sent to the national death registry where the causes of death are coded using the into International Statistical Classification of Diseases and Related Health Problems 10th revision (ICD-10) and a single Underlying Cause of Death established using (a now automated) decision making system developed by the World Health Organisation (WHO) (**Figure 1.2**).[37]

Hospital admission data can be accessed in England via NHS England (in which the dataset is referred to as Hospital Episode Statistics Admitted Patient Care), Public Health Scotland via the electronic Data Research & Innovation Service (Scottish Morbidity Records 01), and the Welsh SAIL Databank (Patient Episode Database for Wales).[29-31] These are coded data transcribed from the patient's hospital discharge summary by hospital coding teams using national clinical coding standards (**Figure 1.1**).[34-36] In England and Wales, a record is created for each period of care under an individual consultant (i.e. an episode) during a hospital admission, while in Scotland only a single record is created for the entire hospital admission. Examples of information included on the record that may be relevant to trials include: clinical diagnoses recorded using ICD-10; any medical operations/procedures (including dates) using OPCS-4; admission method (i.e. emergency or planned); admission, discharge, and episode start and end dates; and the number of nights the participant stayed in hospital.

Unlike operations and procedures (for which exact dates are recorded), any diagnosis recorded in the hospital admission data does not include an onset date or other information on whether this is a new or old event. For primary prevention cardiovascular trials (i.e. no prior CVD) this should not be an issue, but for secondary prevention trials (i.e. prior CVD) it is a challenge to reliably identify whether diagnoses recorded after the randomisation date represent new events or re-reports of a pre-existing conditions. On each hospital record up to 20 diagnoses codes can be recorded, in which the code in the first position (i.e. primary

diagnosis) is generally assumed to be the primary cause of admission (or primary reason for care during a subsequent episode within that admission), and often a new diagnosis.[34] Codes in any other diagnostic position are usually considered to be either contributing causes of admission or any other comorbidities (and could represent pre-randomisation diagnoses). In England, since 2003, diagnoses and procedures have been attributed a particular tariff which contributes to how the hospital gets reimbursed by the government; this system is referred to as Payment by Results.[38] It is generally accepted that since the introduction of these financial incentives coding accuracy has somewhat improved in England.[38,39] But there are also downsides to such tariffs. Increases/decreases in financial reimbursements for the recording of particular conditions can have a substantial impact on how frequently such diseases are recorded in the routine data (i.e. coding shift and coding drift). Thus disease definitions in the routine data may only be appropriate for particular reporting time periods.[40] ICD-10 is the 10th revision of the WHO's International Statistical Classification of Diseases and Related Health Problems.[41] This is a hierarchical medical classification system where diseases are classified first into volumes, then chapters, down to four character codes. For example: IX "Disease of the circulatory system" (chapter); I20-I25 "Ischaemic heart diseases" (blocks of categories); I21 "Acute myocardial infarction" (three-character category); I21.0 "Acute transmural myocardial infarction of anterior wall" (four-character sub-category). Each category includes an unspecified code (i.e. when further details about the disease are not specified or do not fit into one of the other sub-categories); for example, I21.9 "Acute myocardial infarction, unspecified". The OPCS-4 coding structure used for recording operations and procedures occurring during UK hospital admissions follows a similar hierarchical structure to ICD-10.[42] For example: K "Procedure relating to the heart"; K43 "Prosthetic replacement of coronary artery"; K43.1 "Prosthetic replacement of one coronary artery". Both coding systems are routinely updated but any amendments/additions are generally minor.

1.3 Current use of routine data by UK randomised trials

Routinely collected hospital admission and death registry data are often the primary method of follow-up for large UK observational studies, including UK Biobank[43] and the Million Women Study[44]. Despite the wide-spread use of routine data in observational research, the utilisation of such data in clinical trials remains to be lacking, with recent estimates of about 3% of UK RCTs accessing routine data sources.[45,46] Commentaries by various trialists have described barriers including: (1) applying for data access, this can often be a very time consuming and burdensome process, where each UK nation has very different application processes and requirements (e.g. the correct wording on patient consent forms); (2) the timeliness of accessing data, where there are considerable lags (often many months) due to data only being collected periodically with lengthy data cleaning procedures; (3) data stability, for example changes in financial reimbursement (i.e. Payment by Results) for particular diseases can change how frequently the disease is recorded; (4) the complexities of processing and interpreting such data; and (5) regulatory requirements for clinical trials to demonstrate the provenance (i.e. detailed records of how the data were recorded) and integrity (i.e. the consistency and reliability of data throughout the data lifecycle) of all data used in the trial.[47,48]

Beyond these barriers, the main concern appears to be the uncertainty about the reliability of the information recorded in routine data sources, and whether randomised trials can still provide robust results if participants are followed-up via routine data sources.[49] In the recently completed CORMORANT-UK[50] study, trialists and other stake-holders across the UK were surveyed to identify existing challenges that still needed to be addressed before UK RCTs could better utilise routine data. Importantly, one of the top seven priority questions selected by stake-holders was: “How should the trials community decide when routinely collected data for outcomes is of sufficient quality and utility to replace bespoke data collection?” The purpose of this thesis is to investigate whether, for large

cardiovascular RCTs in the UK, follow-up via routinely collected health data are reliable enough to be the sole method of follow-up.

1.4 Assessing the reliability of routine data follow-up

The primary analyses in cardiovascular trials typically assess whether (by the end of the follow-up period) fewer cardiovascular events have occurred in participants allocated to one treatment (e.g. treatment arm A) compared to those allocated to another (e.g. treatment arm B). These comparisons either compare two simple proportions (i.e. the proportion of participants with an event in treatment arm A divided by the same proportion in treatment arm B), where a risk ratio is estimated, or survival methods are used where the comparison also takes into account the time-to-event (e.g. days from the randomisation date).[51] Although analytical approaches are available which consider first and any subsequent (i.e. "recurrent") events[52], this thesis focuses on the most common approach where only the first event for each participant is considered (e.g. if during the follow-up period a person has two distinct myocardial infarction events only the first is counted). For such analyses typically either parametric methods such as the Cox proportional hazard model are used, or non-parametric methods such as log-rank methods where no distributional assumptions are made.[53,54] Both methods have strengths, with the Cox model able to adjust for additional covariates to improve precision of the treatment effect estimate, and log-rank methods perform best at minimising the type II error rate (i.e. falsely rejecting a real treatment effect).[55] This thesis will mainly present analyses using log-rank methods because this was the analytical approach taken by ASCEND and SHARP (the two main trials analysed in this thesis), where a rate ratio is estimated. Log-rank methods compare the survival distributions of two groups (e.g. active treatment arm vs placebo arm). At each time point where an event occurs the observed (O) and expected (E) number of events in each group and their variances (V) are calculated. These are summed to generate an overall statistic (i.e. O minus E divided by V) and exponentiated to calculate the rate ratio. Similarly, statistical significance differences in the survival distributions can be assessed by comparing the Z

statistic (i.e. O minus E divided by the square root of V) to the chi-squared distribution. Compared to adjudicated direct follow-up (assumed here to be equivalent to the ground-truth; i.e. all true events have been correctly identified and no false events have been identified as true events), using routine data as an alternative source of follow-up can introduce three types of errors: adding false events, where a person without an event is falsely identified by the routine data as having an event; missing true events, in which a person with an event is falsely categorised by routine data as not having an event; and an incorrect event date, where the routine data event date is not the same date reported in adjudicated direct follow-up. Thus trial participants can be categorised into four mutually exclusive groups: (1) “outcome in both datasets” (i.e. routine data correctly identified true events); (2) “outcome in routine data only” (i.e. routine data added false events); (3) “outcome in adjudicated direct follow-up alone” (i.e. true events not identified by routine data); and (4) “no such outcome in either dataset”. From such categories, agreement statistics can be calculated, measuring the participant-level reliability of routine data compared to adjudicated direct follow-up. These include: sensitivity (i.e. completeness), calculated by dividing the number of participants with an “outcome in both datasets” by the total number of participants with an outcome reported via adjudicated follow-up. Specificity (i.e. accuracy), calculated by dividing “no such outcome in either dataset” by the number of participants with no such outcome reported via adjudicated follow-up. Positive and negative predictive values (PPV and NPV) assess, for a given number of participants categorised by routine data as having or not having a particular outcome, the proportion of such participants that were correctly categorised. Overall levels of agreement can be measured using the kappa statistic[56], that takes into account agreement by chance (i.e. expect a concordance 50% of the time if no agreement), and interpreted using an established approach[57] where 0.01-0.20 is considered to represent slight agreement, 0.21-0.40 fair agreement, 0.41-0.60 moderate agreement, 0.61-0.80 strong agreement, and 0.80-0.99 very strong agreement. Whilst the kappa statistic is widely used, it has a number of limitations, including the tendency to underestimate agreement when event rates are high

and when errors are not equally distributed between false positives and negatives. A technical overview of analytical approaches used in this thesis can be found in **Appendix 1 Statistics of agreement**.

The reliability of routine data at a trial-level (i.e. whether the results of a trial would have been altered if routine data follow-up had been used) can be assessed by re-running the trial's randomised comparisons using outcomes derived from routine data follow-up alone, and comparing the estimated rate ratio with outcomes derived from adjudicated direct follow-up. Differences between the rate ratios for adjudicated direct follow-up versus routine data can then be calculated, with 95% confidence intervals derived using bootstrap methods.[58] Finally, where a trial must weigh up the benefits and hazards of a particular intervention, absolute treatment effect estimates derived from routine data only can be compared to adjudicated direct follow-up.

1.5 Thesis objectives

The primary aims of this thesis are to investigate, whether for large cardiovascular RCTs in the UK, follow-up of their primary and secondary outcomes via routinely collected hospital admission and death registry data: (1) are ascertained as accurately and completely as adjudicated direct follow-up; and (2) can yield the same estimated treatment effects as adjudicated direct follow-up (**Table 1.1**). Secondary aims are: (3) assessing whether clinical adjudication of direct trial follow-up is necessary; and (4) use simulated data to investigate the impact on randomised trials more generally of removing true events and adding false events, as may happen if routine data are used for follow-up.

1.6 Summary of randomised trials analysed

This thesis presents findings from the analysis of two large cardiovascular trials in people with diabetes and CKD conducted by Oxford CTSU. These trials are introduced below.

1.6.1 The ASCEND trial

The ASCEND (A Study of Cardiovascular Events in Diabetes)[7-9] study was a randomised double-blind trial in UK adults with diabetes but no prior vascular disease, using a 2x2 factorial design to low-dose aspirin (100 mg once per day) versus placebo, and separately to omega-3 fatty acids (1 g once per day) versus placebo. Between 2005 and 2011, 15,480 participants were randomised, with mean follow-up of 7.4 years and ended in 2017. The principal method of direct participant recruitment and follow-up was by mailed questionnaires (paper and, in later years of the follow-up period for some participants, electronic) every 6-months until the end of the trial.

The primary efficacy assessment was time to first serious vascular event (SVE; a composite of non-fatal MI, presumed ischaemic stroke or transient ischaemic attack [TIA], and vascular death [excluding intracranial haemorrhage]). A secondary assessment was time to first SVE or any arterial revascularisation. The primary safety assessment for the aspirin comparison was time to first major bleeding event (a composite of any intracranial haemorrhage, sight-threatening eye bleeding, serious gastrointestinal bleeding [including upper, lower or unspecified bleeding, and perforation], and other major bleeding [including epistaxis, haemoptysis, haematuria, vaginal, and other bleeding]). More than 90% of the SVEs and major bleeding events included in the analysis were verified by adjudicators.

The ASCEND trial's main results were published[8,9] in 2018. The trial found that compared to placebo, allocation to aspirin reduced SVEs by 12% (95% confidence interval [CI] 3% to 21%; 658 participants [8.5%] aspirin vs. 743 [9.6%] placebo), but increased the major bleeding risk by 29% (95% CI 9% to 52%; 314 [4.1%] vs 245 [3.2%]), thus the benefits of aspirin were largely counterbalanced by the bleeding hazard. For the fatty acids comparison no difference in SVEs was observed between the two groups (rate ratio [RR] 0.97; 95% CI 0.87 to 1.08).

In addition to ASCEND's adjudicated mailed-based follow-up data, written informed consent was obtained from participants to allow access to their routinely collected data. This included linkage to death registry data and routinely collected hospital admission data. Routine data was available for almost all ASCEND participants (15,436 [99.7%]), with the exception of 44 participants living in Northern Ireland where linkage was not possible due to a lack of national registry data.

1.6.2 The SHARP trial

The SHARP (Study of Heart and Renal Protection)[10,11] study was a randomised double-blind trial in people with moderate-to-severe chronic kidney disease (CKD) with no history of MI or coronary revascularisation. Between 2003 and 2006, 9,270 participants were randomly assigned to simvastatin 20 mg plus ezetimibe 10 mg daily versus matching placebo. Median follow-up was 4.9 years and ended in 2010. SHARP used clinic-based follow-up, in which participants were seen at 2, 6, and 12 months after randomisation, and then every 6 months.

The primary efficacy assessment was time to first major atherosclerotic event (MAE), defined as non-fatal MI or coronary death (together referred to as major coronary events), non-haemorrhagic stroke (ischaemic and unknown subtypes combined), or arterial revascularisation (coronary or non-coronary, excluding dialysis access). SHARP's renal outcome included start of maintenance kidney replacement therapy (KRT), which includes initiation of maintenance dialysis or receipt of a kidney transplant.

The SHARP trial's main results were published[11] in 2011. They found that compared to placebo, allocation to simvastatin and ezetimibe reduced LDL cholesterol by an average of 0.85 mmol/L and reduced risk of MAE by 17% (95% CI 6% to 26%; 526 [11.3%] simvastatin/ezetimibe vs 619 [13.4%] placebo). In the subset of 6,247 participants not on dialysis at baseline, study treatment had no effect on risk of needing to start KRT (1057 [33.9%] vs 1084 [34.6%]; RR 0.97, 95% CI 0.89 to 1.05).

In addition to SHARP's adjudicated clinic-based follow-up data, written informed consent was obtained from participants to allow access to their routinely collected data. This included linkage to death registry data and routinely collected hospital admission data. Routine data was available for 1,574 (17.0% of the total randomised cohort) SHARP participants living in England.

1.7 Overview of the thesis

This thesis starts with a systematic review (**CHAPTER 2**) to summarise the current literature and identify knowledge gaps in the understanding of how reliable UK routine data are at ascertaining cardiovascular outcomes compared to adjudicated direct follow-up in RCTs.

CHAPTER 3 and **CHAPTER 4** present the results from analyses using data from the ASCEND primary prevention trial (i.e. people with diabetes and no prior vascular disease at baseline), where due to the study having obtained linkage to routine hospital admission and death registry data, the reliability of routine data at ascertaining serious vascular events (**CHAPTER 3**) and major bleeding events (**CHAPTER 4**) can be assessed by comparing these data to adjudicated mail-based follow-up. As the majority of the trial's participants were successfully linked to their routine data, the ASCEND trial's main randomised comparisons were re-run based on the hypothetical scenario that the trial had relied solely on routine data follow-up. The secondary objective of these two chapters was to assess whether clinical adjudication of streamlined mail-based follow-up approaches is necessary; this is a particularly important question for trials conducted in countries or clinical settings where routine data are not easily accessible.

CHAPTER 5 investigates the reliability of routine data at ascertaining major atherosclerotic events and initiation of kidney replacement therapy in a CKD population using a subset of the SHARP cohort living in England, with the secondary objective to assess whether adjudication of clinic-based follow-up is necessary. This is an important question because atherosclerotic events can present very differently in people with CKD, particularly those on

maintenance kidney replacement therapy[59,60], thus reliable ascertainment of events can be more challenging.

Beyond the particular examples in ASCEND and SHARP, in **CHAPTER 6** simulation methods are used to investigate more generally the effects on RCTs of removing true events and adding false events, as may happen if routine data are used for follow-up.

Finally **CHAPTER 7** summarises the key findings from the thesis, interprets the results, provides recommendations for researchers conducing streamlined RCTs, and sets out direction for future research.

Chapter 1 Tables and figures

Table 1.1 Thesis aims

PRIMARY AIMS:
Aim 1: To measure how accurate and complete routine data follow-up are at ascertaining outcomes compared to adjudicated direct follow-up in large cardiovascular trials
1.1 Serious vascular events and major bleeding events in people with diabetes and no prior vascular disease
1.2 Major atherosclerotic events and kidney replacement therapy in people with CKD and no prior myocardial infarction or coronary revascularisation
Aim 2: To assess whether, for large UK primary-prevention cardiovascular trials in people with diabetes, outcomes ascertained via routine data follow-up can yield the same estimated treatment effects as adjudicated direct follow-up
2.1 Serious vascular events
2.2 Major bleeding events
SECONDARY AIMS:
Aim 3: To assess whether clinical adjudication of direct trial follow-up is necessary in large cardiovascular trials
2.1 Mail-based follow-up of serious vascular events and major bleeding events in people with diabetes
2.2 Clinic-based follow-up of major atherosclerotic events and kidney replacement therapyin people with CKD
Aim 4: Use simulated data to investigate the impact on randomised trials more generally of removing true events and adding false events
Trial components to investigate:
3.1 Treatment effect magnitude
3.2 Sample size

CKD = Chronic kidney disease. UK = United Kingdom.

Figure 1.1 Methods of adjudicated direct follow-up and routine data follow-up in randomised trials for participant hospitalisations

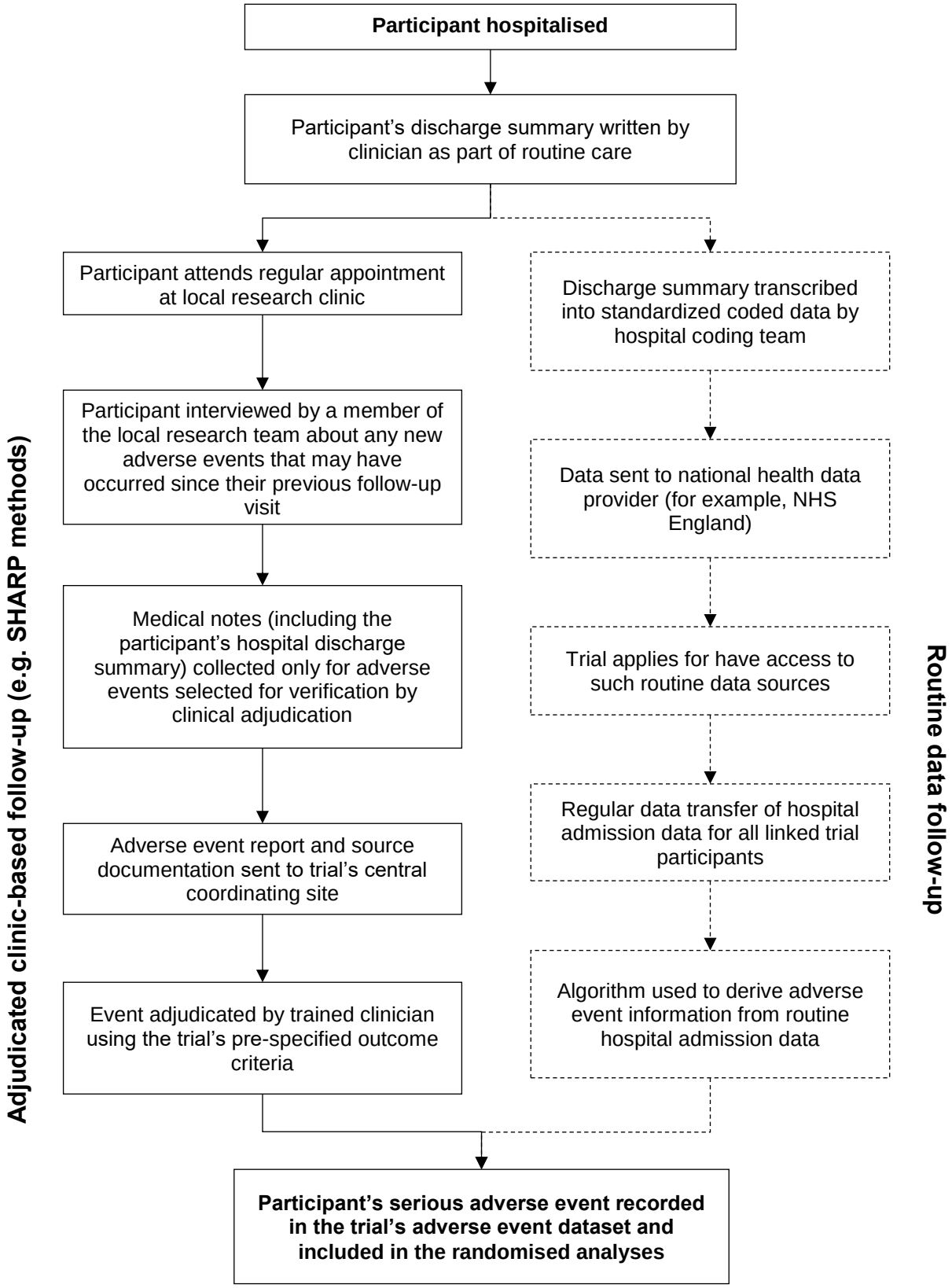


Figure 1.2 Methods of adjudicated direct follow-up and routine data follow-up in randomised trials for participant deaths

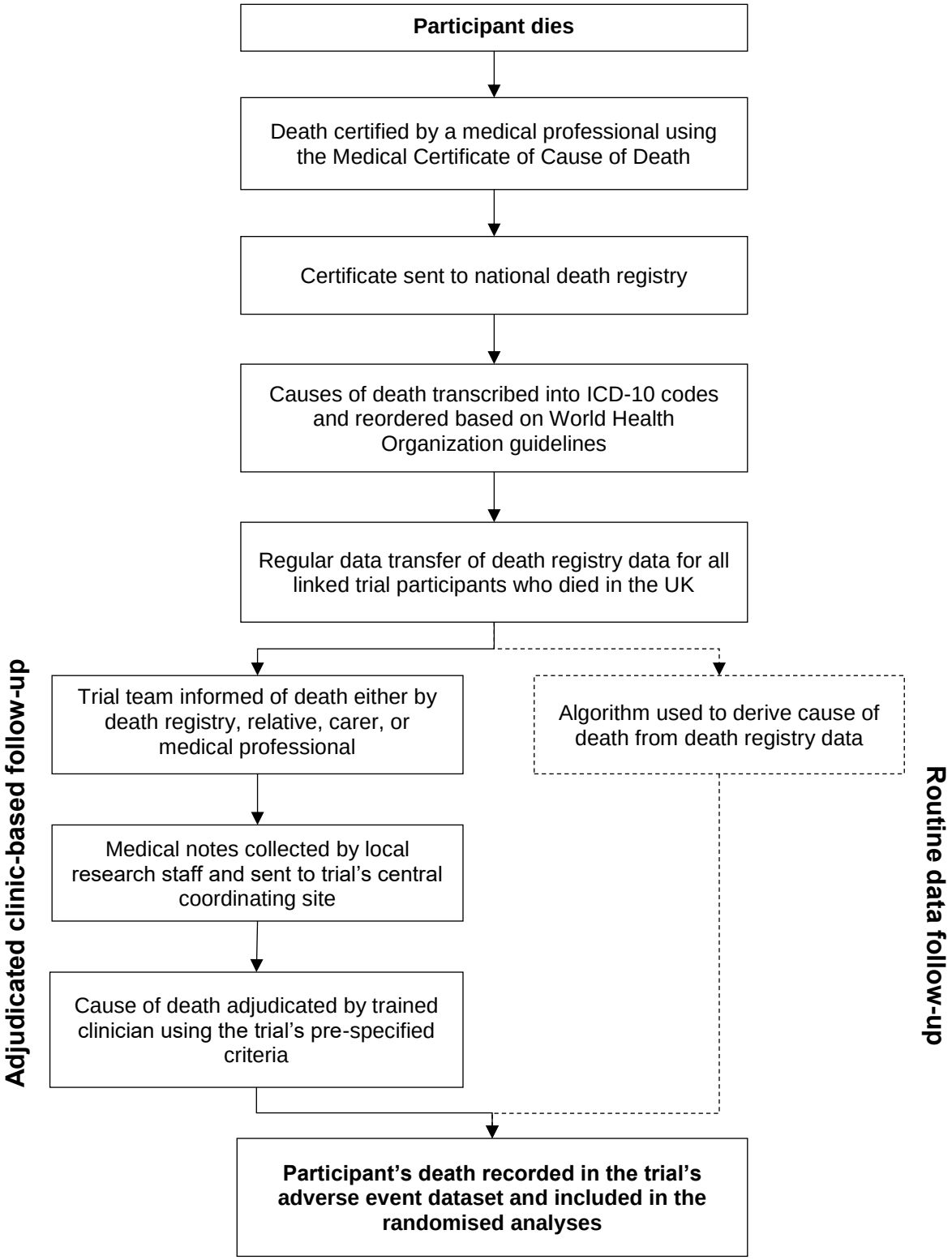
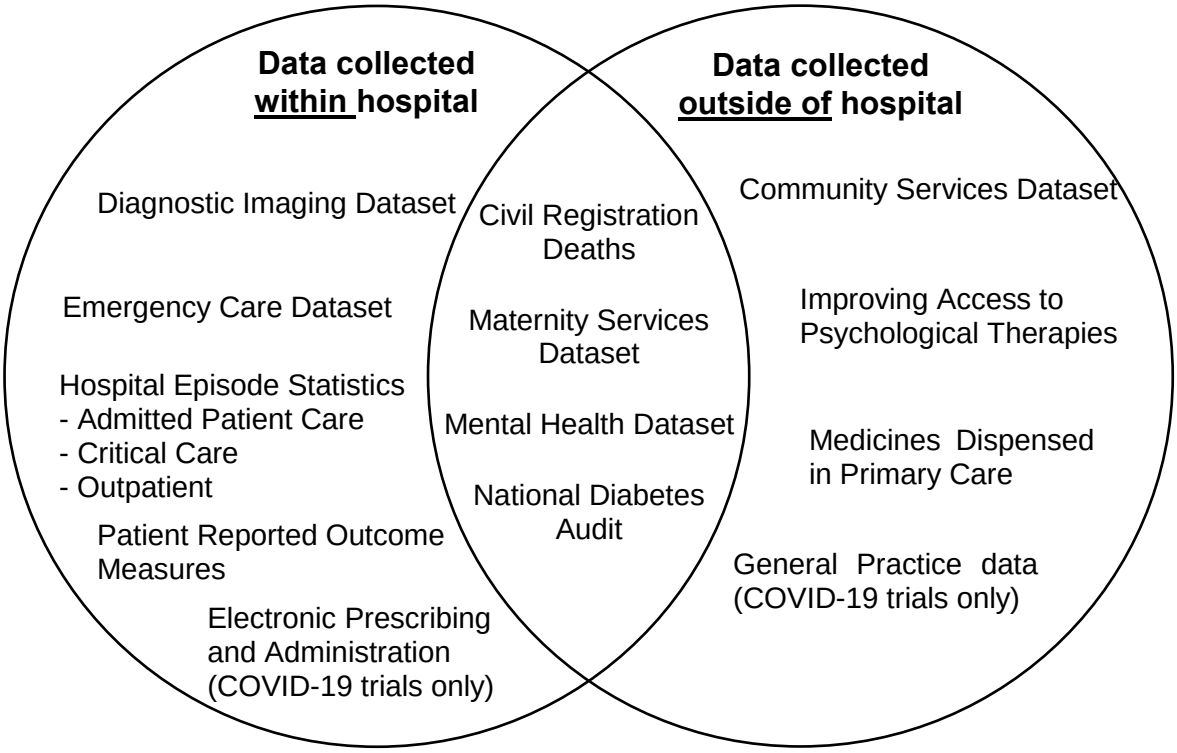


Figure 1.3 Routinely collected health data accessible to researchers from NHS England in 2023



CHAPTER 2 Systematic review of studies assessing the reliability of UK routinely collected data at ascertaining cardiovascular and bleeding outcomes compared to adjudicated direct follow-up in randomised trials

Chapter summary

Aims: This systematic review aims to summarise the current literature and identify knowledge gaps on the reliability of UK routine data to ascertain cardiovascular outcomes compared to adjudicated direct follow-up in RCTs.

Methods: MEDLINE and Embase were searched between 1st January 2010 and 12th August 2022 using a search strategy and selection criteria to identify studies assessing the reliability of UK routine data at ascertaining outcomes compared to trial follow-up data. Information was extracted on the trial's design, routine data linkage, agreement statistics, and re-run randomised comparisons using routine data follow-up.

Results: The systematic search identified 753 articles, after screening, eleven studies were found to satisfy the selection criteria, three of which investigated cardiovascular outcomes. These included: (1) the WOSCOPS (West of Scotland Coronary Prevention Study) double-blind trial placebo controlled trial of pravastatin in 6595 men, living in Scotland, with no history of MI, conducted between 1989 and 1995; (2) the HPS (Heart Protection Study) double-blind placebo controlled trial of simvastatin in 20536 adults with prior coronary disease, other occlusive arterial disease, or diabetes, conducted between 1994 and 2001, with routine data linkage available for 18241 participants living in England only; (3) the HEART-PPCI (The How Effective are Antithrombotic Therapies in Primary Percutaneous Coronary Intervention) open-label intervention trial of herparin vs bivalirudin in 1812 adults, living in England, with suspected ST-elevation myocardial infarction, conducted between

2012 and 2013 (with 28-days participant follow-up). No studies were identified that assessed kidney disease outcomes.

Compared to adjudicated direct follow-up, there was strong agreement for routine data ascertained revascularisations (sensitivity 83%, positive predictive value [PPV] 84%). While for non-fatal MI and stroke outcomes, agreement was moderate (sensitivity ranged from 59% to 74%, PPV 48%-62%), and improved in populations without a prior MI or stroke (sensitivity 72%-81%, PPV 68%-94%). For the major bleeding outcome, routine data identified events (using any code in any hospital admission) compared to adjudication defined major bleeding using BARC (Bleeding Academic Research Consortium) score 3-5 was shown to have moderate completeness (sensitivity 65%) but poor accuracy (PPV 45%). In which routine data follow-up overestimated the number of major bleeding events that occurred in the trial by 43% (adjudicated data 222 vs routine data 318).

In the Heart Protection Study, when randomised comparisons were re-run using routine data only, for major vascular events (including non-fatal myocardial infarction, coronary death, stroke, or any revascularisation), there were no important differences in relative treatment effect estimates compared to adjudicated direct follow-up (adjudicated follow-up: simvastatin 1464 [16.0%] vs placebo 1925 [21.1%]; risk ratio (RR) 0.73, 95% confidence interval [CI] 0.68-0.77; versus routine data: 1549 [17.0%] vs 2038 [22.4%]; RR 0.74, 95% CI 0.69-0.79). Similarly, in the WOSCOPS trial, for major coronary events (including non-fatal MI and coronary heart disease death), comparted to adjudicated direct follow-up, relative treatment effect estimated using routine data follow-up were similar (adjudicated follow-up: pravastatin 215 events [6.5%] vs placebo 295 [8.9%]; hazard ratio [HR] 0.71, 95% CI 0.56-0.85; versus routine data: 121 [3.7] vs 195 [5.9%]; HR 0.61, 95% CI 0.49-0.76). However, it was notable that 38% (316/510) of adjudicated events were not identified in the Scottish routine data.

These studies had a number of limitations: (1) WOSCOPS and HPS assessed old routine data recorded over two decades ago (i.e. 1989-2001) and HEART-PPCI followed-up

participants for only 28 days; (2) these studies did not investigate whether the reliability of such data were consistent across UK nations, and in populations where atherosclerotic events present atypically (e.g. people with moderate-to-severe chronic kidney disease); and (3) no differentiation was made between major and minor bleeding events recorded in the routine data, thus limiting relevance to anti-thrombotic trials where major bleeding if often the primary safety outcome.

Conclusions:

In summary, this systematic review identified three studies that compared cardiovascular and bleeding outcomes ascertained via UK routine data with adjudicated direct follow-up, two of which assessed old routine data collected over two decades ago (i.e. 1989-2001), and the other followed-up participants for only 28 days. The previous evidence is therefore limited in terms of the numbers of studies and the reliance on data which is two decades old. The available published data raise the hypothesis that routine data could be used for trial follow-up for cardiovascular outcomes. Agreement between data sources was strong for revascularisations, and moderate for MIs and strokes. Re-running randomised comparisons using routine data only in two studies provided relative treatment effect estimates similar to adjudicated direct follow-up. For major bleeding events, routine data showed moderate completeness but accuracy was low, as the study made no differentiation between major and minor bleeding events recorded in the routine data. No studies were identified that assessed kidney disease outcomes. This thesis will aim to address some of these gaps in the evidence.

2.1 Introduction

The increasing complexity and costs of conducting trials threaten our ability to generate new randomised evidence.[19] Routinely collected data could be used to streamline trial follow-up as it could potentially substitute for traditional trial follow-up methods of face-to-face visits at local research clinics.[22] However, there remains considerable uncertainty about whether information derived from routine data follow-up are reliable enough to be used in randomised trials.[24] A systematic review of studies (including those conducted outside of the UK) assessing the reliability of routinely collected data at ascertaining cardiovascular outcomes compared to trial follow-up has already been conducted by Rodrigues et al[61]. However, the review may have missed important UK studies, as: (1) it only included published studies up to August 2020; and (2) did not use search terms specifically relating to UK routine data sources. The systematic review presented in this thesis therefore aims to provide a rigorous, up-to-date, summary of the current literature on the reliability of UK routinely collected data at ascertaining cardiovascular outcomes compared to adjudicated direct follow-up in randomised trials, and identify existing knowledge gaps.

2.2 Methods

2.2.1 Search strategy and selection criteria

A systematic search was conducted in MEDLINE and Embase including articles published between 1st January 2010 and 12th August 2022, and followed PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) guidelines (see the completed checklist in **Supplemental Table 2.1;** databases were selected based on established approaches used by previous systematic reviews in this field).[61,62] The search terms and strategy are outlined in full within the protocol (see **Appendix 2 Systematic review protocol**); briefly, for an article to have satisfied the search strategy it must have included a randomised trial term, routine data source term (either a specific dataset, or non-specific

routine data term with reference to data linkage), and a term referencing a method of comparison between data sources. After duplicate articles were removed, each article's title and abstract were screened by a single reviewer in order to exclude any study where there was no comparison between randomised trial direct follow-up and routine data for the reliable ascertainment of trial outcomes. For the remaining articles full text screening was then conducted to exclude studies investigating non-UK routine data sources. Finally, the studies selected for review were categorised by whether they investigated cardiovascular or non-cardiovascular outcomes.

2.2.2 Data extraction

Information from each full-text article was extracted including: the original trial's design; trial participant characteristics, particularly prior diseases at baseline; routine data linkage (i.e. number of participants linked, time-period where linkage was available, and routine datasets used); algorithms used to define the outcomes (e.g. primary or secondary diagnostic position on the hospital record); criteria for an event to be considered a "match" between the two data sources (e.g. <14 days difference); trial outcomes investigated; and method used to assess the reliability of routine data follow-up.

For studies investigating cardiovascular outcomes, agreement statistics (i.e. statistics to assess how accurate and complete routine data are at ascertaining trial outcomes compared to trial follow-up) were extracted, including: sensitivity, specificity, PPV, NPV and kappa statistics (see **Appendix 1 Statistics of agreement**).[56] Where studies re-ran randomised comparisons using routine data follow-up as the sole source of outcome data, data was extracted on: the number of participants by trial arm, the number of events by trial arm, and the relative treatment effect estimates (e.g. rate ratio or hazard ratio) including its 95% confidence interval.

2.3 Results

2.3.1 Study selection

Of the 753 studies identified from the systematic search (MEDLINE: 383; Embase: 370), 352 were duplicates, 332 were not randomised trials, and 47 trials were either not linked to a routine data source (n=24), linked but no comparison against the original trial data (n=10), and 13 trials investigated the validity of routine data for purposes other than ascertaining trial outcomes (for example, recruitment processes and participant baseline characteristics; **Figure 2.1**). Among the 22 studies that underwent full text screening, 11 trials were excluded as they were linked to non-UK routine data sources, and the article documenting the analysis of serious vascular events[63] in the ASCEND trial presented in **CHAPTER 3** of this thesis was also excluded. In addition to the 10 studies identified from the systematic search, the DPhil candidate's supervisors were aware of an additional study included in another recently submitted DPhil thesis within the department. After a full-text review it was decided that the study satisfied the criteria to be included in the systematic review. Therefore, in total, eleven articles were included in the systematic review. Three investigated cardiovascular and bleeding outcomes, including: the West of Scotland Coronary Prevention Study[64,65] (WOSCOPS), Heart Protection Study[62,66] (HPS), and How Effective are Antithrombotic Therapies in Primary Percutaneous Coronary Intervention[67,68] (HEART-PPCI); and eight studies[69-76] investigated non-cardiovascular outcomes. No studies were identified that assessed kidney disease outcomes.

2.3.2 Study characteristics

The WOSCOPS[64,65] trial (conducted between 1989 and 1995 in Scotland) was a double-blind placebo controlled randomised trial of pravastatin in 6595 men aged 45-64 years with high low-density lipoprotein cholesterol and no history of MI (i.e. a primary prevention trial; **Table 2.1**). The trial's primary outcome was any major coronary event (MCE), a composite of non-fatal myocardial infarction and coronary heart disease (CHD) death. The trial used

clinic-based follow-up and primary/secondary outcomes were verified by clinical adjudication. All trial participants were successfully linked to their routine hospital admission data (Scottish Morbidity Records 01[30]) and death registry data[33] for the full duration of the trial's follow-up period. Non-fatal myocardial infarctions and strokes were identified using a list of ICD-9 codes (**Supplemental Table 2.2**) using any diagnostic position on the hospital record, while coronary death was identified using the underlying cause on the death record.

The Heart Protection Study[62,66] study (conducted between 1994 and 2001 in the UK) was a double-blind placebo controlled randomised trial of simvastatin in 20536 adults aged 40 to 80 years with prior coronary disease, other occlusive arterial disease, or diabetes (i.e. a secondary prevention trial; **Table 2.1**). The trial's primary outcome was all-cause mortality and one of the trial's secondary outcomes was any major vascular event (MVE; including non-fatal MI, coronary death, stroke, or any revascularisation). The trial used clinic-based follow-up and primary/secondary outcomes were verified by clinical adjudication. Trial participants living in England (n=18241) were linked to their routine hospital admission data (Hospital Episode Statistics Admitted Patient Care[31]), however, data were only available from 1997 to 2001 as national collection of routine hospital data in England only started in 1997. For coronary deaths, the HPS study[62] used adjudicated death data both in its adjudicated direct follow-up and routine data follow-up. Non-fatal MIs and strokes were identified using a list of ICD-10 codes (**Supplemental Table 2.2**) in any diagnostic position on the hospital record. Revascularisations were identified using a list of OPCS-4 codes in any position.

The HEART-PPCI[67,68] study (conducted between 2012 and 2013 in England) was an open-label intervention trial of herparin vs bivalirudin in 1812 adults with suspected ST-elevation myocardial infarction (STEMI) scheduled for an angiography (i.e. an acute secondary prevention trial; **Table 2.1**). The trial's primary outcome was major adverse ischaemic events at 28 days, a composite of all-cause mortality, cerebrovascular accident, reinfarction, or unplanned target lesion revascularisation. The primary safety outcome was

major bleeding, defined as types 3 to 5 as per the Bleeding Academic Research Consortium[77] (i.e. requiring a blood transfusion, drop in haemoglobin, surgical intervention, intravenous vasoactive treatment, intracranial haemorrhage, or intraocular bleeding compromising vision). The trial tracked participants during their index admission using hospital medical notes and contacted participants via telephone after discharge. The trial's primary/secondary outcomes were verified by clinical adjudication. Trial participants (n=1754) with complete personal identifiers were linked to their Hospital Episode Statistics Admitted Patient Care[31] data for the 28 day follow-up period. Recurrent myocardial infarctions, strokes and bleeding events (which included codes for anaemia) were identified using a list of ICD-10 codes (**Supplemental Table 2.2**) using any diagnostic position on the routine hospital record. A detailed overview of the 8 non-cardiovascular studies can be found in **Supplemental Table 2.3**.

2.3.3 Agreement between routine data and adjudicated direct follow-up for cardiovascular and bleeding outcomes

For MI, agreement between routine data and adjudicated direct follow-up appeared to vary substantially between studies. In the HEART-PPCI trial where participants were admitted with a suspected MI, routine data identified recurrent MI (during the same admission as the first MI) had a sensitivity of 3% and 6% PPV. In the HPS trial, where 41% of participants had a previous MI, sensitivity was 74% and 62% PPV. For WOSCOPS, where no participants had a prior MI, compared to adjudicated data restricted to hospitalised MIs only, routine data were found to have 81% sensitivity and 94% PPV.

Similarly, for stroke, routine data showed better agreement in trials where no participants had a prior stroke and adjudicated data were restricted to hospitalised event only. For example, in the HPS trial, where 19% of participants had a prior stroke/TIA and adjudicated data included strokes occurring both within and outside of hospital, routine data had a sensitivity of 59% and 48% PPV. For HEART-PPCI and WOSCOPS, where no participants had a prior stroke and adjudicated data were restricted to hospitalised strokes only, routine

data had moderate-to-high sensitivity (WOSCOPS 72%; HEAR-PPCI 73%) and PPV (WOSCOPS 68%; HEART-PPCI 94%). For revascularisations, routine data had both high sensitivity (83%) and PPV (84%).

For the major bleeding outcome, routine data identified events (using any bleeding/anaemia code in any hospital admission) compared to adjudication defined major bleeding using BARC[77] (Bleeding Academic Research Consortium) score 3-5 was shown to have moderate completeness (sensitivity 65%) but low accuracy (PPV 45%). In which routine data follow-up overestimated the number of major bleeding events that occurred in the trial by 43% (adjudicated data: major bleeding events 222, vs routine data: 318). Results for the eight non-cardiovascular studies can be found in **Supplemental Table 2.3**.

2.3.4 Randomised comparisons for major vascular and major coronary events using routine data follow-up

In the HPS trial, for the composite outcome of any MVE, compared to adjudicated direct follow-up (simvastatin 1549 [17.0%] vs placebo 2038 [22.4%]; RR 0.73, 95% CI CI 0.68-0.78), randomised comparisons re-run using routine data only provided similar relative treatment effect estimates (0.74, 95% CI 0.69-0.79; 1464 [16.0%] vs 1925 [21.1%]; **Figure 2.2**). These findings were consistent for all components of the MVE outcome. Similarly, in the WOSCOPS trial, for the any major coronary event outcome, relative treatment effect estimates using routine data only (pravastatin 121 [3.7] vs placebo 195 [5.9%]; hazard ratio [HR] 0.61, 95% CI 0.49-0.76) were consistent with adjudicated direct follow-up (215 events [6.5%] vs 295 [8.9%]; HR 0.71, 95% CI 0.56-0.85; **Figure 2.3**). However, overall event numbers were found to be 38% lower (316/510) in the routine data compared to adjudicated direct follow-up.

2.4 Discussion

This systematic review identified three studies that investigated the reliability of UK routine data in ascertaining cardiovascular and bleeding outcomes compared to adjudicated trial

follow-up, and none that assessed kidney disease outcomes. In total, these studies included: 965 adjudicated MIs; 737 strokes; 1529 revascularisations; and 222 major bleeding events. For any revascularisations agreement was found to be strong. For MI and stroke events, agreement between routine data and adjudicated follow-up varied substantially, with routine data performing best in trials studying populations with no prior vascular disease (i.e. primary prevention). Presumably because the diagnostic codes recorded in routinely collected hospital admission data makes it challenging to differentiate between “old” and “recurrent” events. Nevertheless, relative treatment effects estimated using routine data alone were generally similar to adjudicated direct follow-up. However, these analyses were based on old routine data recorded over two-decades ago (i.e. 1989-2001), so its findings may be of limited relevance to trials currently in development. For major bleeding events, routine data showed moderate completeness but accuracy was low, as the study made no differentiation between major and minor bleeding events recorded in the routine data. Therefore, although the findings from these studies indicate that there is potential for UK routine data to be utilised by cardiovascular trials, further research, such as that presented in this thesis, are necessary to develop a strong evidence base to inform the design of future cardiovascular trials.

This systematic review builds on work conducted by Rodrigues et al[61], who conducted a review of studies (including in non-UK countries) assessing the reliability of routine data compared to adjudicated trial follow-up for cardiovascular outcomes (published prior to August 2020). Rodrigues et al[61] found that relative treatment effects estimated using routine data were similar to adjudicated direct follow-up, with these findings being consistent across countries (e.g. US, Denmark, UK) and outcomes (e.g. cardiovascular mortality, MI, stroke). The systematic review undertaken by the DPhil candidate identified two important additional studies (HEART-PPCI[67] published in the later part of 2020, and un-published HPS data[62]) not included in the previous review. These additional UK studies provided important data showing how routine data may be more reliable at identifying first (rather than recurrent)

cardiovascular events, and HEART-PPCI was the first study to assess the reliability of major bleeding events ascertained via UK routine data follow-up.

For MI and stroke outcomes, this systematic review has highlighted how routine data agreement can vary substantially depending on whether trial participants have prior vascular disease. As outlined in **CHAPTER 1**, routine hospital admission data can include up to 20 ICD-10 diagnosis codes, in which the primary diagnosis tends to indicate the primary reason for admission, yet there is limited information that indicates the temporality of the diagnosis. In the UK studies included in this review the accuracy of the routine data (i.e. how many participants were falsely identified as having an MI) varied substantially depending on how many trial participants had a prior MI. Using PPV as a measure of accuracy, in WOSCOPS[65] where no participants had a prior MI this was estimated at 94%, 62% for HPS[62] where just under half (i.e. 41%) of participants had a prior MI, and 6% in HEART-PPCI[67] where all participants had a suspected MI. Nonetheless, despite the apparent over counting of non-fatal MIs by routine data follow-up in the HPS trial, there appeared to be little impact on relative treatment effect estimates. Thus indicating that the very high threshold set by trialists when assessing the accuracy and completeness of routine data may be unnecessarily strict and that small-to-moderate levels of random error may have little effect on a trial's results.[3]

The HEART-PPCI[67] trial was the only study to investigate major bleeding events, a safety outcome commonly assessed in anti-thrombotic cardiovascular trials. Major bleeding events ascertained using routine data follow-up, had good sensitivity but low accuracy, in which routine data follow-up appeared to overestimate the number of major bleeding events that occurred in the trial by almost 50%. However, the study had substantial limitations, as definitions of major bleeding were very different between adjudicated direct follow-up that used a BARC[77] score 3-5 (i.e. requiring a blood transfusion, drop in haemoglobin, surgical intervention, intravenous vasoactive treatment, intracranial haemorrhage, or intraocular bleeding compromising vision) and routine data, where any bleeding or anaemia code in

any diagnostic position was used. Importantly, the study did not investigate whether an algorithm could be developed that may allow for the categorisation of routine data identified bleeding into major and minor events, or re-run randomised comparisons.

This systematic review had a number of limitations, most notably the inability to assess for publication bias. Trial methodological research is often not a priority to publish for trialists or journals (as it is a means to an end), therefore although routine data comparisons may have been carried out, very often they may not have been published or disseminated outside of research groups and not been subject to peer review. We also recommend that the Cochrane risk of bias tool be expanded to also asses methodological studies within trials, this has the potential to encourage better and more consistent reporting standards in this area where current this is lacking.[78] This is particularly the case when the results may indicate routine data being a poor substitute for adjudicated direct follow-up. Furthermore, few UK trials have yet linked to routine data sources, thus this area of research is still at its early stages.

The findings from this review were also limited because two studies assessed old routine data recorded over two decades ago (WOSCOPS[64] 1989-1995; HPS[66] 1997-2001). In the 1990s nationally collected healthcare data were only in its infancy[79], with national clinical coding standards and payment by results not yet in force.[34-36] Over the past two decades the accuracy and completeness of diagnoses and procedures recorded in routinely collected hospital admission data are likely to have substantially improved.[38] An example of such improvements are that stroke subtypes (i.e. haemorrhagic and ischaemic) are now more accurately recorded whilst in the late 1990s (as seen in the HPS study) strokes were most commonly coded as “stroke unspecified”.[80,81] Additionally, in the HEART-PPCI study[67] no differentiation was made between major and minor bleeding events recorded in the routine data, thus limiting its relevance for antithrombotic trials where major bleeding if often the primary safety outcome. Finally, these studies were not able to investigate whether the reliability of such data were consistent across UK nations, and in populations where

atherosclerotic events present atypically (e.g. people with moderate-to-severe chronic kidney disease). These are important questions for large UK-wide trials wanting to test interventions in diverse populations.

This thesis therefore aims to address these limitations by:

- Investigating the reliability of cardiovascular outcomes ascertained using modern up-to-date routine data sources, across UK nations.
- Developing a routine data algorithm to accurately categorise bleeding events as major or minor.
- Assessing the reliability of routine data ascertained cardiovascular outcomes in populations where presentation of such events are atypical, for example in people with moderate-to-severe chronic kidney disease.
- Investigate whether the kidney replacement therapy outcome can be reliability ascertained from UK routine data.

2.5 Summary

In summary, this systematic review identified three studies that compared cardiovascular and bleeding outcomes ascertained via UK routine data with adjudicated direct follow-up, two of which assessed old routine data collected over two decades ago (i.e. 1989-2001), and the other followed-up participants for only 28 days. Nevertheless, for cardiovascular outcomes, agreement between data sources was strong for revascularisations, and moderate for MIs and strokes. Re-running randomised comparisons using routine data only provided relative treatment effect estimates similar to adjudicated direct follow-up. For major bleeding events, routine data showed moderate completeness but accuracy was low, as the study made no differentiation between major and minor bleeding events recorded in the routine data. No studies were identified that assessed kidney disease outcomes. This thesis will aim to address some of these gaps in the evidence.

Chapter 2 Tables and figures

Table 2.1 Summary of studies selected for systematic review

Trial	Trial design	Participant characteristics	Routine data linkage	Algorithm	Criteria for event match	Data available
WOSCOPS (West of Scotland Coronary Prevention Study)	**Country:** Scotland **Participants:** 6,595 **Trial arms:** Pravastatin vs placebo (double-blind) **Primary outcome:** Major coronary events **Follow-up:** Clinic-based follow-up **Adjudication:** Yes **Mean follow-up:** 4.9 years **Years trial active:** 1989-1995	Men aged 45-64 years with high low-density lipoprotein cholesterol and no history of myocardial infarction	**Participants linked:** All 6,595 trial participants **Time-period linked:** 1989-1995 **Datasets:** Scottish morbidity records (hospitalisations) and nationally collected Scottish death records	Pre-defined code list (see Supplemental table 2.1). Diagnosis code in any position on hospital record. Underlying cause on the linked death record.	Routine data event occurring within 2 weeks of adjudicated follow-up event	**Outcomes:** Non-fatal MI, CHD death, major coronary event, stroke, transient ischaemic attack **Comparisons:** - Event-level and participant-level agreement between routine data and adjudicated direct follow-up - Randomised comparisons re-run using routine data only
HPS (Heart Protection Study)	**Country:** United Kingdom **Participants:** 20,536 **Trial arms:** Simvastatin vs placebo (double-blind) **Primary outcome:** All-cause mortality **Follow-up:** Clinic-based follow-up **Adjudication:** Yes **Mean follow-up:** 5.0 years **Years trial active:** 1994-2001	Adults aged 40-80 years with coronary disease, other occlusive arterial disease, or diabetes	**Participants linked:** 18,241 participants living in England **Time-period linked:** 1997-2001 **Datasets:** Hospital Episode Statistics Admitted patient Care	Pre-defined code list (see Supplemental table 2.1). Diagnosis and procedure codes in primary or any position on hospital record. HPS adjudicated follow-up for deaths.	Routine data event and adjudicated event both occurring within the linkage follow-up period	**Outcomes:** Non-fatal MI, non-fatal stroke, revascularisation, major vascular events (including non-fatal MI, coronary death, stroke, or revascularisation) **Comparisons:** - Participant-level agreement between routine data and adjudicated direct follow-up - Randomised comparisons re-run using routine data only
HEART-PPCI (How Effective are Antithrombotic Therapies in Primary Percutaneous Coronary Intervention)	**Country:** England **Participants:** 1,812 **Trial arms:** Heparin vs bivalirudin (open-label) **Primary outcome:** Major adverse ischaemic events **Follow-up:** Tracked during index admission and mail/telephone follow-up after discharge **Adjudication:** Yes **Mean follow-up:** 28 days **Years trial active:** 2012-2013	Adults with suspected ST-elevation myocardial infarction (STEMI) scheduled for angiography	**Participants linked:** 1,754 participants with no missing identifiers **Time-period linked:** 2012-2013 **Datasets:** Hospital Episode Statistics Admitted patient Care	Pre-defined code list (see Supplemental table 2.1). Diagnosis code in any position on hospital record.	Routine data event and adjudicated event both occurring during any time within the linkage follow-up period	**Outcomes:** Recurrent MI, stroke, major bleeding **Comparisons:** - Participant-level agreement between routine data and adjudicated direct follow-up (before and after discharge)

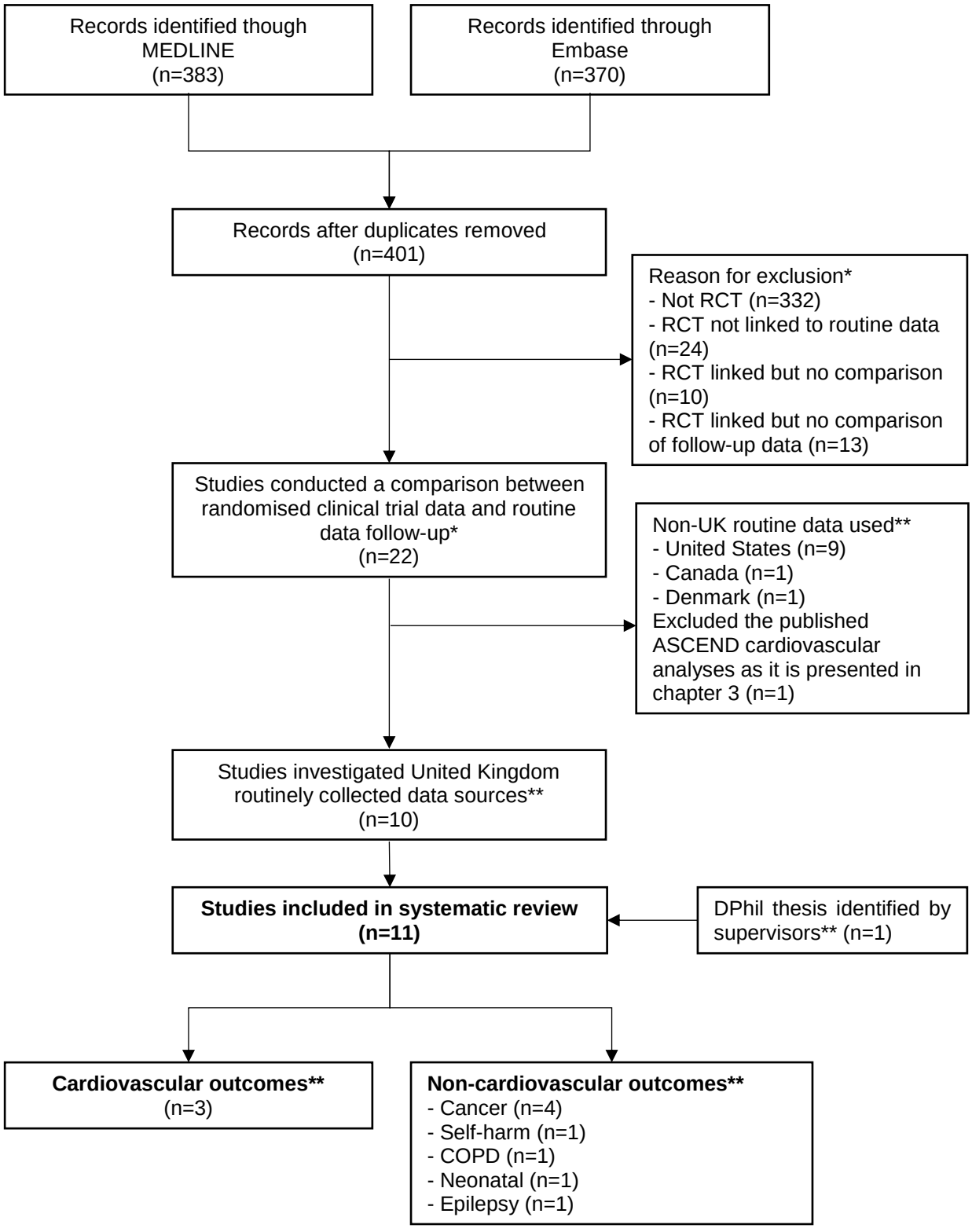
CHD = Coronary heart disease. MI = Myocardial infarction.

Table 2.2 Agreement between routine data and adjudicated direct follow-up in three studies investigating cardiovascular and bleeding outcomes

Outcome	Total events in adjudicated follow-up	Total events in routine data follow-up	Sensitivity	Specificity	Positive predictive power	Negative predictive power	Kappa (95% CI)
Non-fatal MI							
HEART-PPCI (index admission)*	29	16	3%	99%	6%	98%	-
HPS	668	794	74%	98%	62%	99%	0.66 (0.63-0.69)
WOSCOPS (hospitalised only) ⁑	268	230	81%	99%	94%	99%	-
Non-fatal stroke							
HEART-PPCI (index admission)*	22	17	73%	99%	94%	99%	-
HPS	654	911	59%	98%	48%	98%	0.51 (0.48-0.54)
WOSCOPS (hospitalised only) ⁑	61	65	72%	99%	68%	99%	-
Any revascularisation							
HPS	1529	1550	83%	99%	84%	98%	0.82 (0.81-0.84)
Any major vascular event							
HPS	3389	3587	85%	95%	80%	96%	0.78 (0.77-0.79)
Any major bleeding event**							
HEART-PPCI*	222	318	65%	89%	45%	94%	-

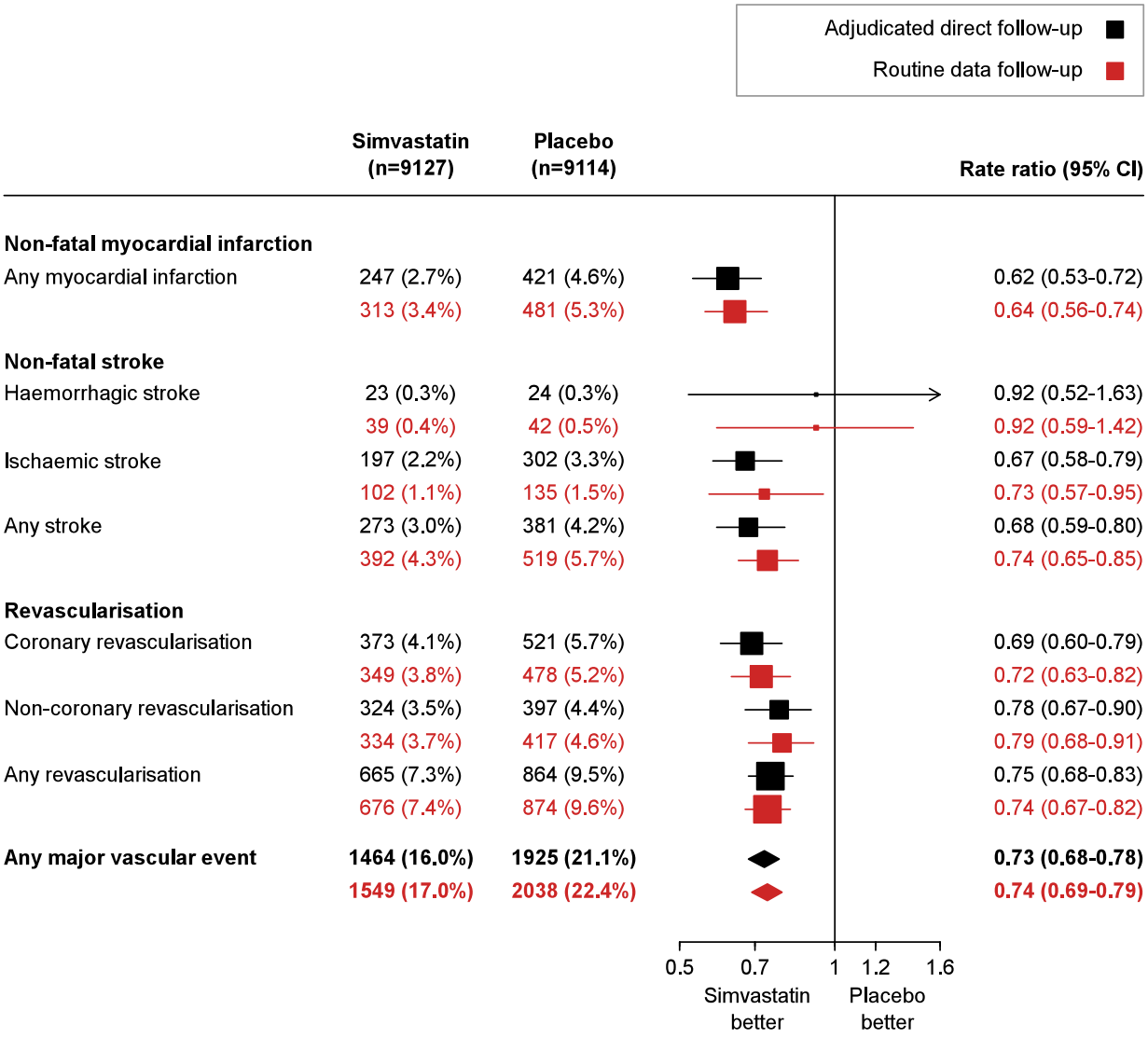
⁑ WOSCOPS restricted its adjudicated follow-up data to hospitalised events only. *HEART-PPCI restricted both methods of follow-up to events that occurred during the participant's index admission. **HEART-PPCI defined major bleeding as BARC (Bleeding Academic Research Consortium) score 3-5 in adjudicated direct follow-up and a bleeding code in any diagnostic position on the routine hospital admission record. CI = confidence interval. MI = Myocardial infarction.

Figure 2.1 Flow chart of study selection for systematic review



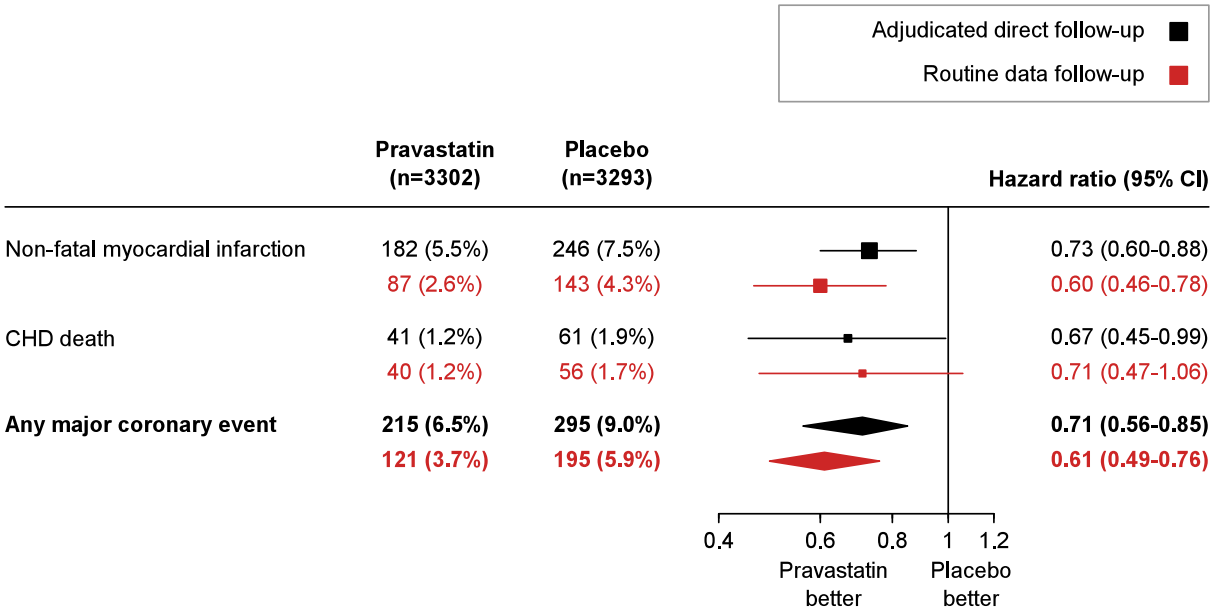
*Titles and abstract screened. **Full text screened. Search included articles published between 1st January 2010 and 12th August 2022. COPD = Chronic obstructive pulmonary disease. RCT = Randomised controlled trial.

Figure 2.2 Effect of allocation to simvastatin vs placebo on any major vascular event using routine data follow−up in the HPS trial



For coronary deaths, the study did not conduct comparisons between adjudicated direct follow-up and routine data follow-up, therefore no rate ratios for this outcome were available. CI = Confidence interval. HPS = Heart Protection Study.

Figure 2.3 Effect of allocation to simvastatin vs placebo on any major coronary event using routine data follow−up in the WOSCOPS trial



CHD = Coronary heart disease. CI = Confidence interval. WOSCOPS = West of Scotland Coronary Prevention Study.

CHAPTER 3 Comparison of routinely collected UK data with adjudicated direct follow-up: analyses of serious vascular events from the ASCEND randomised trial

This chapter was published in JAMA Network Open:

Harper C, et al. Comparison of the Accuracy and Completeness of Records of Serious Vascular Events in Routinely Collected Data vs Clinical Trial–Adjudicated Direct Follow-up Data in the UK: Secondary Analysis of the ASCEND Randomized Clinical Trial. JAMA Network Open 2021; 4(12): e2139748-e.

Chapter summary

Aims: To assess, in people with diabetes and no prior vascular disease, (1) how reliable UK routine data are for ascertaining serious vascular events compared with adjudicated direct follow-up data, and (2) the impact of adjudication on mail-based follow-up of serious vascular events.

Methods: The ASCEND (A Study of Cardiovascular Events in Diabetes) primary prevention trial randomised 15480 UK people with diabetes between 2005-2011 using a 2x2 factorial design to aspirin versus matching placebo, and separately to omega-3 fatty acids versus placebo. Mean participant follow-up was 7.4 years and ended in 2017. The primary efficacy outcome was any serious vascular event (SVE; a composite of non-fatal myocardial infarction, ischaemic stroke or transient ischaemic attack [TIA], or vascular death [excluding intracranial haemorrhage]). Direct participant mail-based follow-up was the main source of outcome data, with more than 90% of the primary outcome events undergoing adjudication. In parallel, more than 99% of participants were linked to routinely collected hospital

admission and death registry data (i.e. routine data), enabling randomised comparisons of different sources of outcome data.

Results: For any SVE, agreement between adjudicated direct follow-up and routine data sources was strong (1401 vs 1127 events; kappa 0.78 [95% CI 0.76-0.80]; sensitivity 72.0% [95% CI 69.7%-74.4%]; specificity 99.2% [95% CI 99.0%-99.3%]), and sensitivity improved when adjudicated follow-up was restricted to fatal and hospitalised events only (1099 vs 1127 events; sensitivity 86.4% [95% CI 84.4%-88.5%]). Rate ratios [RRs] were similar for the aspirin randomised comparison for adjudicated direct follow-up versus follow-up solely through routine data alone (adjudicated data: 658 [8.5%] aspirin vs 743 [9.6%] placebo; rate ratio [RR] 0.88, 95% CI 0.79-0.97; versus routine data: 537 [6.9%] vs 590 [7.6%]; RR 0.91, 95% CI 0.81-1.02), and almost identical when adjudicated follow-up was restricted to fatal and hospitalised events only (adjudicated data: 526 [6.8%] vs 573 [7.4%]; RR 0.91, 95% CI 0.81-1.03). Results were similar for the omega-3 fatty acid comparison, and adjudication had little impact on the RRs.

Conclusions: The findings from these analyses of UK routinely collected hospital admission and death registry data linked to the ASCEND trial show how such data could be used to provide reliable estimates of the treatment effects of interventions on myocardial infarctions, hospitalised ischaemic strokes, arterial revascularisations, and vascular deaths in streamlined primary prevention trials, perhaps without the need for verification by clinician adjudicators. However, event rates from routine data follow-up were marginally lower than adjudicated direct follow-up as these routine datasets were not appropriate for capturing transient ischaemic attacks and non-hospitalised strokes. Finally, in settings where routine data are not available, this study has also demonstrated the reliability of streamlined mail-based follow-up methods, without the need for adjudication.

3.1 Introduction

Large randomised trials (RCTs) in cardiovascular disease (CVD) have provided reliable evidence for interventions whose widespread use has contributed to the secular declines in CVD mortality.[17,82] However, increasing costs of conducting trials threaten our ability to generate new randomised evidence.[19,22] Observational analyses in routinely collected healthcare (i.e. "real-world") data are increasingly used to make inferences about the safety and efficacy of interventions, but such non-randomised analyses cannot fully account for confounding and are unreliable.[2,3] More appropriately, routinely collected data could be used to streamline trial follow-up conduct/design by offering a cost-efficient method to ascertain outcomes.[83,84]

The National Health Service (NHS) in the United Kingdom (UK) has long established, routinely collected, national level mortality and hospital admission records, including Hospital Episode Statistics[31] in England, with equivalents in Wales[29] and Scotland[30]. These records cover all NHS hospitals and have been collected since the 1990s using standardised clinical coding practices[34-36], but have rarely been adopted into trial designs.[43-45,85] This is perhaps due to the limited evidence of their accuracy and completeness for specific outcomes.[24]

In **CHAPTER 2** results were presented from three studies[62,65,67] that compared cardiovascular outcomes ascertained via UK routine data with adjudicated direct follow-up, two of which assessed old routine data collected over two decades ago (i.e. 1989-2001), and the other followed-up participants for only 28 days. The findings from these studies suggested that revascularisations could be reliably ascertained through linkage to UK routine data sources, but the results for MI and stroke were far less clear, with agreement varying substantially between studies.

In **CHAPTER 3** this hypothesis is further tested by assessing "modern" UK routine data recorded from 2005 to 2017 against adjudicated direct follow-up data from the ASCEND[7-9] (A Study of Cardiovascular Events iN Diabetes) randomised trial. This large double-blind

placebo-controlled trial in people with diabetes but no evidence of atherosclerotic vascular disease at recruitment (i.e. a primary prevention population). A total of 15480 participants were randomised and followed up by regular mail-based questionnaires, and reports of possible serious vascular events (SVEs, a composite of non-fatal myocardial infarction, ischaemic stroke, transient ischaemic attack [TIA], or vascular death [excluding intracranial haemorrhage]) or revascularisations underwent clinical adjudication. Participants were also linked to national death and cancer registries, and routinely collected hospital admission data.

This study aimed to assess whether such routinely collected data are sufficiently reliable to be the sole source of serious vascular outcome ascertainment. Analyses compared the ASCEND results using adjudicated direct participant follow-up with two alternative hypothetical scenarios: (a) following participants solely through routine data; and (b) performing streamlined mail-based follow-up without adjudication.

3.2 Methods

3.2.1 The ASCEND trial

Analyses were conducted using data from the ASCEND trial whose design and findings were outlined in **CHAPTER 1** (**Figure 3.1**).[7-9] Briefly, between 2005-2011, 15480 UK participants were randomised using a 2x2 factorial design to low-dose aspirin (100 mg once per day) versus placebo, and separately to omega-3 fatty acids (1 g once per day) versus placebo. The primary efficacy assessment was time to first serious vascular event (a composite of non-fatal myocardial infarction, presumed ischaemic stroke or transient ischaemic attack [TIA], and vascular death [excluding intracranial haemorrhage]). A secondary assessment was time to first serious vascular event or any arterial revascularisation. Mean follow-up was 7.4 years and ended in 2017. The study found that compared to placebo, allocation to aspirin reduced SVE by 12% (95% confidence interval [CI] 3% to 21%; 658 participants [8.5%] aspirin vs. 743 [9.6%] placebo).[8] For the fatty acids

comparison, no difference in SVEs was observed between the two groups (rate ratio [RR] 0.97; 95% CI 0.87 to 1.08).[9] The protocol was approved by the North West Multicentre Research Ethics Committee, and participants provided written informed consent.

3.2.2 Mail-based participant follow-up and adjudication

The principal method of direct participant recruitment and follow-up was by mailed questionnaires (paper and, in later years of the follow-up period for some participants, electronic) every 6-months until the end of the trial, referred to in this chapter as "pre-adjudicated direct follow-up" (**Supplemental Figure 3.1**). The primary and secondary outcomes and whether participants were hospitalised or not were adjudicated by study clinicians blind to treatment allocation. For non-fatal vascular outcomes, the coordinating centre requested documentation (e.g. hospital discharge summaries) from primary care. For the 1,076 non-fatal SVEs reported in the primary results, 514 (47.8%) events were ascertained via participant mail-based follow-up; 260 (24.2%) via telephone contact with the participant; 255 (23.7%) from contact with their general practitioner or hospital doctor; 28 (2.6%) collected via other means; and 19 (1.8%) via routine hospital admission data where (in a small number of cases) participants were lost to follow-up. For fatal outcomes, the primary source of information was the Office for National Statistics (ONS) death certification data, which included deaths occurring both within and out of hospital. The reported primary and subsidiary causes of death were then reviewed by study clinicians along with available hospital admissions data and any information from the trial questionnaires. The published results were based on all unrefuted serious adverse event reports (i.e. a reported event for which supporting documentation could not be obtained was counted in the analysis). More than 90% of the SVEs included in the analysis were verified by adjudicators. These outcome data from adjudicated follow-up presented in the ASCEND primary publications[8,9] are referred to as "adjudicated direct follow-up" in this chapter (**Supplemental Figure 3.1**). Details of the adjudication definitions (including how deaths were ascribed to their underlying cause using information from review of other

clinical information alongside death certificates) are provided in Appendix A of the main publication.[8]

3.2.3 Routinely collected hospital admission and death registry data

Written informed consent was obtained from participants to allow access to their routinely collected data. During follow-up, death records were obtained from the Office for National Statistics via NHS England for England/Wales[32] and National Health Scotland Central Register[33]. These data included date of death, underlying and other contributing causes of death. Participants were also linked to their routinely collected hospital admission records; these data were obtained from NHS England (Hospital Episode Statistics Admitted Patient Care)[31] for England, Public Health Scotland (Scottish Morbidity Records 01)[30], and Welsh SAIL Databank (Patient Episode Database for Wales)[29]. Information used for these analyses included: primary and any secondary diagnoses; any operations and procedures; and admission dates. Linkage between the trial participants and routine datasets were carried out by the national registries, but was not possible for 44 participants (44/15480 [0.3%]) residing in Northern Ireland, among whom only one serious vascular event occurred during follow-up.

All deaths in the UK are included in the mortality registry, irrespective of location of death, and medically certified by a physician or pathologist. For routine data analyses, deaths with vascular disease as the underlying cause were identified from the register, which uses an automated algorithm based on principles set out in the International Classification of Diseases 10th-revision (ICD-10) (categorisation is provided in **Supplemental Table 3.1**). Non-fatal SVEs were identified in hospital admission data using ICD-10 diagnosis codes in any diagnostic position (i.e. primary or secondary). The event date was assumed to be the admission date as there was no diagnosis date in the available routine data. Any arterial revascularisation procedures (including dates) were identified using Office of Population Censuses Surveys Classification of Surgical Operations and Procedures 4th-revision procedure codes. Myocardial infarctions and ischaemic strokes were assumed to be non-

fatal unless they were within 30 days of such deaths ascribed to such causes. Events identified through routinely collected data did not undergo clinical adjudication and are referred to as “routine data” in this thesis (**Supplemental Figure 3.1**).

3.2.4 Statistical analyses

Analyses included only those events that occurred between randomisation and date of death or censoring, except for Northern Ireland participants where routine data follow-up was censored at day zero. Outcomes identified using routine data were compared to those from adjudicated direct participant follow-up. For each outcome, participants were categorised as having: “outcome in both datasets” (i.e. routine data identified an outcome which had also been ascertained by adjudicated direct follow-up); “outcome in routine data only” (i.e. not ascertained in adjudicated direct follow-up); “outcome in adjudicated direct follow-up alone” (i.e. outcomes not recorded in routine data); or “no such outcome in either dataset”. Levels of agreement between the two data sources were assessed using sensitivity and specificity with 95% confidence intervals (see **Appendix 1 Statistics of agreement**).[86] Overall levels of agreement were estimated using the kappa statistic[56] with 95% confidence intervals, and interpreted using an established approach[57] where 0.01-0.20 is considered to represent slight agreement, 0.21-0.40 fair agreement, 0.41-0.60 moderate agreement, 0.61-0.80 strong agreement, and 0.80-0.99 very strong agreement.

Differences in kappa statistics between subgroups were assessed using heterogeneity testing[56,87], including analyses by participants’ mean age (<63 vs ≥63-years), sex (male vs female), vascular risk score (low vs medium vs high [ASCEND generated its own risk score based on age, sex, smoking, systolic blood pressure, body mass index, duration of diabetes and glycated haemoglobin; see ASCEND data analysis plan[7] for further details]), and country of residence (England vs Other UK). Where there was agreement between the two sources of outcome data (i.e. outcome in both datasets), the event dates were compared. Differences were presented as: exact match, 1-7 days, 8-30 days, 31-90 days, 91-180 days, and >180 days. Where an SVE was reported in adjudicated follow-up alone, routine data

were searched to identify whether there was a corresponding hospitalisation (within 90 days of the adjudicated event date) or death record for the same participant. A similar process was carried out for events recorded in routine data only.

All randomised comparisons used standard log-rank methods[53,54] following the intention-to-treat approach. Rate ratios with 95% confidence intervals are provided. The main randomised comparisons were those based on the adjudicated direct follow-up data versus those based on the alternative scenario that ASCEND had only used routine data to identify outcomes. Sensitivity analyses included restricting adjudicated follow-up to fatal and hospitalised events only (in which events where adjudicated hospitalisation status was "unknown" were assumed to have not led to an admission), and analyses restricted to only using diagnoses in the primary position of the hospital admission record. For the main randomised comparisons, differences between the rate ratios for adjudicated follow-up versus routine data were calculated; with 95% confidence intervals derived using bootstrap methods. These methods used 1000 resamplings (as per standard practice[58]), with replacement, where the difference in point estimate was recalculated in each bootstrap sample. Secondary analyses included randomised comparisons using outcomes derived from pre-adjudicated direct follow-up. Analyses were conducted using SAS version 9.4, and R version 4.2.2.

3.3 Results

3.3.1 Baseline characteristics of participants in the ASCEND trial

A total of 15480 participants were included in this study; mean age was 63 years (standard deviation 9.2); 9684 (62.6%) were men and 5796 (37.4%) were women (**Supplemental Table 3.2**). Hypertension was reported in 9533 participants (61.6%). The median duration of diabetes was 7-years (interquartile range 3-13), 2668 (17.2%) participants had a high vascular risk score, and 13960 participants (90.2%) resided in England.

3.3.2 Accuracy and completeness of routine data ascertained SVEs

There were 1401 unrefuted SVEs within adjudicated direct follow-up, of which 1009 were also identified in the routine data, so sensitivity of routine data was 72.0% (95% confidence intervals [CI] 69.7%-74.4%; **Table 3.1**). There were 118 events that were recorded in the routine data but not confirmed in adjudicated direct follow-up, and 13961 participants having no event in either dataset; hence, specificity of routine data was 99.2% (95% CI 99.0%-99.3%). Overall agreement between adjudicated follow-up and routine data for any SVEs was strong (1401 versus 1127 events; kappa 0.78, 95% CI 0.76-0.80), and further improved when arterial revascularisations were added to the composite outcome (kappa 0.83, 95% CI 0.82-0.85).

Specificity of routine data was very strong (>99%) for all components of any SVEs and for arterial revascularisations, but sensitivity varied. It was highest for any arterial revascularisation, at 94.6% (685 of 724 events; 95% CI 93.0%-96.3%) and vascular death at 88.2% (365 of 414 events; 95% CI 85.1%-91.3%; **Table 3.1**). For myocardial infarction it was 78.8% (304 of 386 events; 95% CI 74.7%-82.8%), non-fatal presumed ischaemic stroke was lower (288 of 431 events; 66.8%, 62.4%-71.3%), and lowest for transient ischaemic attack (109 of 365 events; 29.9%, 25.2%-34.6%).

Levels of agreement for any SVE were generally consistent by subgroups, with the exception of vascular risk score where participants with a low score showed slightly lower agreement (kappa 0.71, 95% CI 0.66-0.75) compared to a medium (0.79, 95% CI 0.76-0.81) or high score (kappa 0.80, 0.77-0.84; **Supplemental Table 3.3**). Sensitivity was lower when restricting to diagnosis codes in the primary position only (primary position: 68.2%; vs any: 72.0%) without no apparent improvements in specificity (primary: 99.4%; vs any: 99.2%; **Supplemental Table 3.4**). For the 1009 SVEs in which there was agreement between adjudicated direct follow-up and routine data, date of hospital admission was accurate to within one week for 857 (85%) events (**Supplemental Table 3.5**).

There were 188 SVEs recorded in routine data only of which only a small proportion were reported by participants but refuted during the adjudication process: for non-fatal myocardial infarction 7/54 (13%) were refuted events, 3/40 (8%) for non-fatal presumed ischaemic stroke, and 2/13 (15%) for TIA (**Supplemental Table 3.6**). As the ASCEND trial collected and adjudicated all deaths, the 11 vascular deaths recorded in routine data only were all refuted during adjudication. When the 392 SVEs reported in adjudicated direct follow-up alone were investigated, the corresponding routine data information varied substantially by SVE component (**Supplemental Table 3.7**). For non-fatal myocardial infarction, 44/70 (63%) had a corresponding routinely collected hospitalisation record within 90 days, but with a non-myocardial-infarction ischaemic heart disease code recorded (see **Supplemental Table 3.7** for code list). When routine data missed non-fatal cerebrovascular events, most commonly there was no corresponding routinely collected hospitalisation record within 90 days (61/93 [66%] for presumed ischaemic stroke; 130/199 [65%] for TIA). For vascular death, the majority of events had a corresponding routinely collected death certificate but the vascular code was not recorded as the underlying cause (21/30 [70%]).

The sources of routine data used for these analyses were limited to capturing mostly fatal and hospitalised events, and only 78% (1099/1401) of adjudicated follow-up reported SVEs were recorded in the trial as involving hospitalisation or death. When adjudicated follow-up were restricted to fatal and hospitalised events only, sensitivity of routine data was found to be higher (86.4%, 95% CI 84.4%-88.5%) while maintaining strong specificity (98.8%, 95% CI 98.6%-98.9%; **Supplemental Table 3.8**). Of the adjudicated follow-up reported ischaemic strokes and transient ischaemic attacks, 71% (306/431) and 33% (120/365), respectively, led to hospitalisation or death. When adjudicated follow-up were restricted to fatal and hospitalised events only, sensitivity of routine data was very strong for ischaemic strokes (260 of 306 events; 85.0%, 95% CI 81.0%-89.0%) and moderate for transient ischaemic attacks (71 of 120 events; 59.2%, 95% CI 50.4%-68.0%).

3.3.3 Randomised comparisons for SVEs using routine data

Overall, the randomised comparisons using routine data alone provided similar estimates of treatment effect to those based on adjudicated direct follow-up for SVEs (adjudicated data: 658 [8.5%] aspirin vs 743 [9.6%] placebo; rate ratio [RR] 0.88, 95% CI 0.79-0.97; versus routine data: 537 [6.9%] vs 590 [7.6%]; RR 0.91, 95% CI 0.81-1.02; **Figure 3.2**). Similar findings were apparent for analyses of the effect of omega-3 fatty acids versus placebo for SVEs (adjudicated follow-up: 689 [8.9%] omega-3 fatty acids vs 712 [9.2%] placebo; RR 0.97, 95% CI 0.87-1.08; versus routine data: 547 [7.1%] vs 580 [7.5%]; RR 0.94, 95% CI 0.84-1.06). For SVEs, the difference in estimated rate ratios between adjudicated direct follow-up and routine data was +0.03 (95% CI -0.04, +0.10) for the aspirin randomised comparison, and -0.03 (95% CI -0.10, +0.05) for the omega-3 fatty acids comparison (**Supplemental Table 3.9**). When adjudicated follow-up was restricted to fatal and hospitalised SVEs only, treatment effects for both the aspirin (526 [6.8%] vs 573 [7.4%]; RR 0.91, 95% CI 0.81-1.03) and omega-3 fatty acids comparisons (536 [6.9%] vs 563 [7.3%]; RR 0.95, 95% CI 0.85-1.07) were almost identical to the above routine data follow-up results (**Figure 3.3**).

When revascularisations were included in the composite SVE outcome, both adjudicated direct follow-up (RR 0.88, 95% CI 0.80-0.97) and routine data (RR 0.90, 95% CI 0.81-0.99) appeared to identify a significant benefit of aspirin. Broken down into its components, treatment effects were almost identical for any arterial revascularisation and non-fatal myocardial infarction, and similar for non-fatal presumed ischaemic stroke and vascular death. However, for the transient ischaemic attack outcome, RRs derived from routine data follow-up were somewhat closer to the null with much wider confidence intervals (adjudicated follow-up: 0.85, 95% CI 0.69-1.04; versus routine data: 0.96, 95% CI 0.69-1.33; difference in point estimate: +0.11, 95% CI -0.14, +0.44), when adjudicated follow-up was restricted to fatal and hospitalised TIAs only, these differences appeared to diminish.

3.3.4 Effect of adjudication on mail-based direct follow-up

Adjudication verified approximately three-quarters of participant mail-based reported myocardial infarctions (363 of 494 [73%]) and presumed ischaemic strokes (371 of 488 [76%]), and more than 80% of reported coronary revascularisations (521 of 638 [82%]; **Table 3.2**). Of the 355 reported non-coronary revascularisations, only 189 (53%) were verified by adjudication. Reviewing 968 reports of hospitalised angina led to only 285 (29%) being confirmed and identified 86 (9%) non-fatal myocardial infarctions, whereas reviewing 576 TIAs, confirmed 338 (59%) and identified 110 (19%) presumed ischaemic strokes.

When randomised analyses were conducted using mail-based pre-adjudicated direct follow-up, compared to adjudicated direct follow-up, the relative effects of treatment for SVE remained largely unchanged for both the aspirin (adjudicated direct follow-up: 658 [8.5%] aspirin vs 743 [9.6%] placebo; RR 0.88, 95% CI 0.79-0.97; versus pre-adjudicated direct follow-up: 753 [9.7%] vs 835 [10.8%]; RR 0.90, 95% CI 0.81-0.99) and omega-3 fatty acids comparisons (adjudicated follow-up: 689 [8.9%] omega-3 fatty acids vs 712 [9.2%] placebo; RR 0.97, 0.87-1.08; versus pre-adjudicated follow-up: 798 [10.3%] vs 790 [10.2%]; RR 1.01, 0.92-1.12; **Figure 3.4**).

3.4 Discussion

In these analyses from the ASCEND trial, this study found that routinely collected information about serious vascular events (SVEs, a composite of non-fatal myocardial infarction, presumed ischaemic stroke or transient ischaemic attack [TIA], and vascular death [excluding intracranial haemorrhage]) and arterial revascularisations based on hospital admission and death registry data in the UK provided similar estimated treatment effects to adjudicated follow-up for both the aspirin and omega-3 fatty acids randomised comparisons. Agreement between adjudicated follow-up and routine data was strong for these SVEs, with sensitivity substantially improved when adjudicated follow-up were restricted to fatal and hospitalised SVEs only. Sensitivity was highest for revascularisations,

presumably because the tariff associated with such procedures yields high rates of reimbursement for the hospital. Furthermore, treatment effects were similar when mail-based pre-adjudicated direct participant follow-up reports were compared to adjudicated outcome data. Sensitivity of routine data to ascertain both non-fatal presumed ischaemic strokes and particularly TIAs was lower than for vascular death and cardiac components of SVEs. This could be partly owing to approximately one-third of strokes being managed within outpatient stroke clinics within the UK[80,88]. Routine data did appear to identify some additional events potentially missed by direct participant follow-up. Therefore, although future trials could ascertain myocardial infarctions, hospitalised ischaemic strokes, arterial revascularisations, and vascular deaths solely from routine data, if a wider range of cerebrovascular outcomes is pre-specified to be recorded, this will require direct participant follow-up and/or additional datasets such as UK primary care data.

The findings for non-fatal myocardial infarctions confirm the hypotheses raised by similar analyses of the WOSCOPS[64,65] and HPS[62,66] trials (presented in **CHAPTER 2**). Using more "modern" hospital data from the UK, the ASCEND trial suggests that myocardial infarction events may be recorded more completely than these earlier data, with adjudication no longer being necessary. For non-fatal MIs (including both hospitalised and non-hospitalised events), sensitivity of routine data in primary prevention trails was found to be about 50% in WOSCOPS and 79% in ASCEND, while the HPS secondary prevention trial had 74% sensitivity (with some indications these data were picking up pre-randomisation events). For non-fatal ischaemic strokes, recent observations from Oxfordshire hospital admissions data also corroborate our findings: about two-fifths of ischaemic strokes were not hospitalised and of those that were, only approximately 70% were identifiable using Hospital Episode Statistics.[80] The presented randomised findings from ASCEND in the UK also mirror findings from similar analyses performed using data from randomised trials linked to insurance claims data from the US[58] and separately to public registers in Denmark[89].

The point estimates in the randomised comparisons based solely on routine data in ASCEND showed no evidence of being materially altered for the primary outcome of SVEs or components of that composite outcome based on adjudicated direct follow-up (i.e. there was little evidence of bias, at least for these outcomes in this single trial; **Supplemental Table 3.9**). However, the 22% reduction in the numbers of SVEs from 1401 identified by participant reporting which were then adjudicated versus 1127 SVEs identified through record linkage alone has implications for sample size calculations. If ASCEND had been designed to solely use routinely collected death and hospitalisation data for follow-up, we estimate it would have been necessary for approximately 3500 more participants to be randomised to retain its intended power (if followed-up for the same duration). However, the efficiency of using routine data arguably outweighs this disadvantage and enables low-cost long-term follow-up. Investigators and trial steering committees can pay particular attention to early event rates in trials reliant on routine data for their follow-up to ensure reassessment of design assumptions and recommend protocol modification, where necessary. Indeed monitoring of event rates in ASCEND led to the important decision to expand the trial from its originally intended sample size of 10000.[7]

The rate ratios for the analyses of the primary serious vascular event outcome were similar when serious vascular outcomes were ascertained solely from mail-based questionnaire follow-up without adjudication. The presented analyses provide evidence that participant reported SVEs are sufficiently reliable that confirmation by clinician adjudicators is unnecessary for most vascular events. This is consistent with meta-analysis results from methodological studies demonstrating that adjudication had little effect on observed rate ratios compared to vascular outcomes ascertained at research clinic.[90] ASCEND differs from these other studies because it relied on participant reports from mailed questionnaires rather than the more resource-intensive clinic-based follow-up with trained research staff.

Depending on the nature of the trial, data on events needs to be collected in a timely manner.[91] If the only source of such information is routine data, it is important that this is

received soon enough for relevant clinical decisions to be made. For example, pharmacovigilance reporting timelines may require alternative methods of rapid reporting of suspected serious adverse reactions (e.g. 24-hour phone services). Nevertheless, information from routinely collected sources may be sufficiently timely for data monitoring committees' serial reviews of randomised comparisons, despite outcomes not be recorded immediately. Current collaborative efforts in the UK between researchers and data providers through the recently founded Health Data Research UK[92] mean that a greater amount of routinely collected healthcare data may be accessible to trialists within the timeframes needed to monitor the safety of trial participants.[26,28] In addition, in several non-UK countries, cardiovascular trials have been successfully embedded into routine healthcare data systems, such as TASTE[93] in Sweden[94] and ADAPTABLE[95] in the US[25].

3.4.1 Strengths and limitations

This study has shown how reliable "modern" routine data are at ascertaining serious cardiovascular events across England, Scotland, and Wales, in people with no evidence of atherosclerotic vascular disease. The re-running of ASCEND's randomised comparisons using routine data highlighted how small-to-moderate errors in event ascertainment have little impact on relative treatment effect estimates due to the protection afforded by large-scale randomisation. This study has also shown how streamlined methods of sending questionnaires to participants via mail without adjudication can still provide a reliable means of ascertaining information about serious cardiovascular events during trial follow-up.

There are some limitations of the current study. First, ASCEND was a primary prevention population limiting generalisability to secondary prevention trials, in which a distinction must be made between an incident serious vascular event and the recording of pre-randomisation diseases. Secondly, ASCEND participants were not linked to their primary care records which may have improved the ascertainment of non-hospitalised SVEs and there was no information in the routine data about stroke severity.[96] Finally, the ASCEND trial was carried out in the UK which has a national health service, the study therefore was

not able to assess the reliability of routine data in non-UK countries, including those that do not have a single national healthcare provider nor if trials are conducted across different countries. However, the ADAPTABLE team of trialists in the US[95] have demonstrated the feasibility of a collaborative network (known as the PCORnet[25]) to perform a large randomised trial of different aspirin doses using electronic healthcare records as the key method of participant identification and follow-up in a non-centralised healthcare system.

3.5 Summary

The findings from these analyses of "modern" UK routinely collected hospital admission and death registry data linked to the ASCEND trial show how such data could be used to provide reliable estimates of the treatment effects of interventions on myocardial infarctions, hospitalised ischaemic strokes, arterial revascularisations, and vascular deaths in streamlined primary prevention trials, perhaps without the need for verification by clinician adjudicators. However, event rates from routine data follow-up were marginally lower than adjudicated direct follow-up as these routine datasets were not appropriate for capturing transient ischaemic attacks and non-hospitalised strokes. Finally, in settings where routine data are not available, this study has also demonstrated the reliability of streamlined mail-based follow-up methods, without the need for adjudication.

Chapter 3 Tables and figures

Table 3.1 Agreement of routine data versus adjudicated direct follow-up for serious vascular events in the ASCEND trial

Outcome	Outcome in both datasets	Outcome in routine data only	Outcome in adjudicated follow-up alone	No such outcome in either dataset	Sensitivity (95% CI)	Specificity (95% CI)	Kappa (95% CI)
Non-fatal myocardial infarction	304 (2%)	79 (1%)	82 (1%)	15015 (97%)	78.8% (74.7%-82.8%)	99.5% (99.4%-99.6%)	0.79 (0.75-0.82)
Non-fatal presumed ischaemic stroke	288 (2%)	65 (<1%)	143 (1%)	14984 (97%)	66.8% (62.4%-71.3%)	99.6% (99.5%-99.7%)	0.73 (0.69-0.76)
Vascular death excluding intracranial haemorrhage	365 (2%)	18 (<1%)	49 (<1%)	15048 (97%)	88.2% (85.1%-91.3%)	99.9% (99.8%-99.9%)	0.91 (0.89-0.93)
Any serious vascular event excluding TIA	**910 (6%)**	**116 (1%)**	**219 (1%)**	**14235 (92%)**	**80.6% (78.3%-82.9%)**	**99.2% (99.0%-99.3%)**	**0.83 (0.82-0.85)**
Transient ischaemic attack	109 (1%)	30 (<1%)	256 (2%)	15085 (97%)	29.9% (25.2%-34.6%)	99.8% (99.7%-99.9%)	0.43 (0.36-0.49)
Any serious vascular event including TIA	**1009 (7%)**	**118 (1%)**	**392 (3%)**	**13961 (90%)**	**72.0% (69.7%-74.4%)**	**99.2% (99.0%-99.3%)**	**0.78 (0.76-0.80)**
Any arterial revascularisation	685 (4%)	51 (<1%)	39 (<1%)	14705 (95%)	94.6% (93.0%-96.3%)	99.7% (99.6%-99.7%)	0.94 (0.92-0.95)
Any serious vascular event or revascularisation	**1404 (9%)**	**127 (1%)**	**365 (2%)**	**13584 (88%)**	**79.4% (77.5%-81.3%)**	**99.1% (98.9%-99.2%)**	**0.83 (0.82-0.85)**

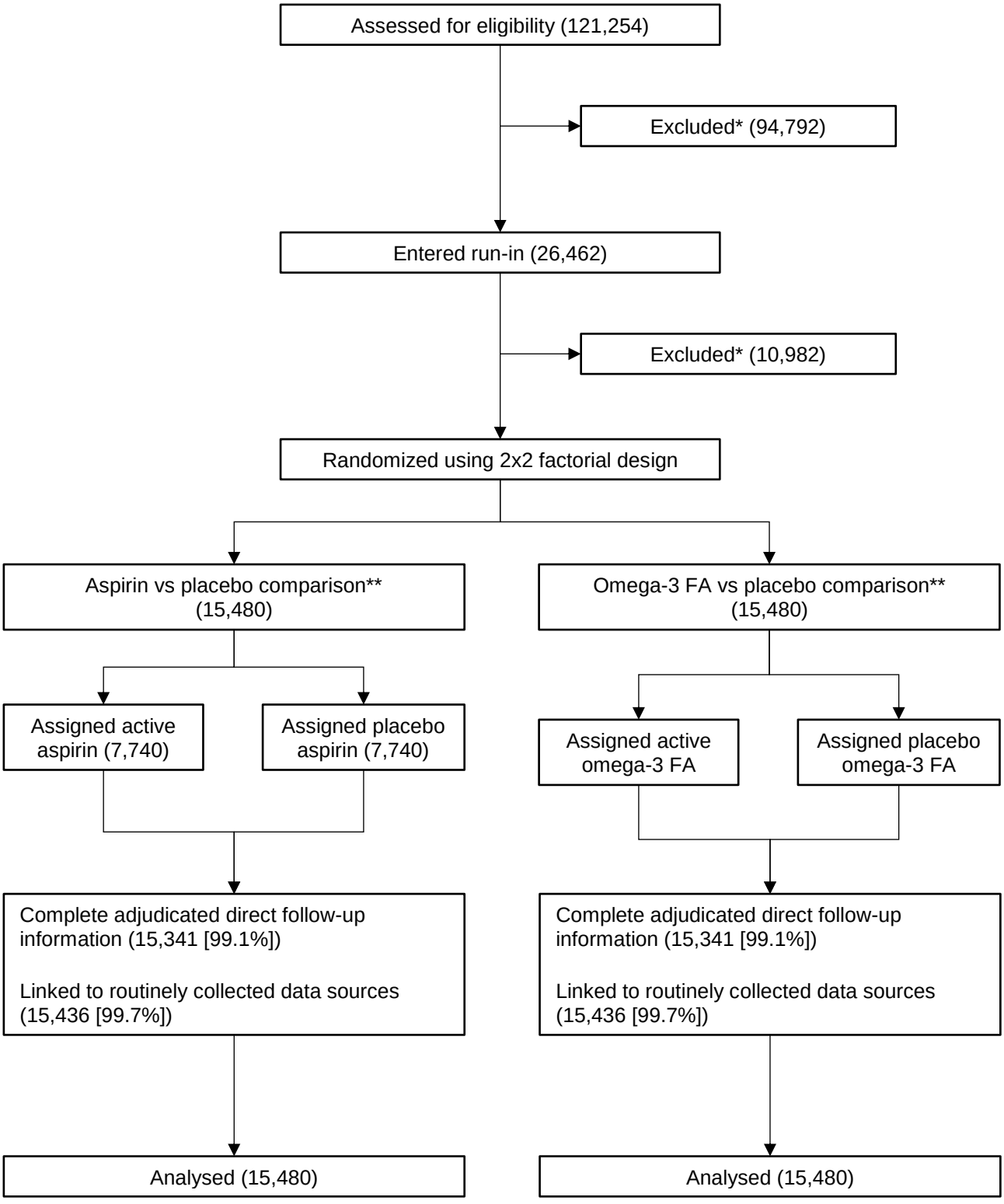
Percentages in parentheses are % of total number of ASCEND participants. CI = Confidence intervals. TIA = Transient ischaemic attack.

Table 3.2 Comparison between pre and post adjudicated direct follow-up for serious vascular events in the ASCEND trial

	Outcome after adjudication									
Outcome before adjudication	**Non-fatal myocardial infarction**	**Angina hospital-lisation**	**Coronary revasc-ularisation**	**Non-coronary revasc-ularisation**	**Transient ischaemic attack**	**Presumed ischaemic stroke**	**Haemorr-hagic stroke**	**Subdural haemorr-hage**	**Other**	**Total**
Non-fatal myocardial infarction	**363 (73%)**	13 (3%)	2 (<1%)	0 (0%)	1 (<1%)	1 (<1%)	0 (0%)	0 (0%)	114 (23%)	**494**
Angina hospitalisation	86 (9%)	**285 (29%)**	9 (1%)	0 (0%)	1 (<1%)	0 (0%)	0 (0%)	0 (0%)	587 (61%)	**968**
Coronary revascularisation	0 (0%)	0 (0%)	**521 (82%)**	37 (6%)	0 (0%)	1 (<1%)	0 (0%)	0 (0%)	79 (12%)	**638**
Non-coronary revascularisation	0 (0%)	0 (0%)	22 (6%)	**189 (53%)**	0 (0%)	0 (0%)	0 (0%)	0 (0%)	144 (41%)	**355**
Transient ischaemic attack	0 (0%)	0 (0%)	0 (0%)	0 (0%)	**338 (59%)**	110 (19%)	4 (1%)	2 (<1%)	122 (21%)	**576**
Presumed ischaemic stroke	0 (0%)	0 (0%)	0 (0%)	1 (<1%)	38 (8%)	**371 (76%)**	7 (1%)	3 (1%)	68 (14%)	**488**
Haemorrhagic stroke	0 (0%)	0 (0%)	0 (0%)	0 (0%)	0 (0%)	1 (4%)	**21 (84%)**	2 (8%)	1 (4%)	**25**
Subdural haemorrhage	0 (0%)	0 (0%)	0 (0%)	0 (0%)	0 (0%)	0 (0%)	0 (0%)	**29 (94%)**	2 (6%)	**31**
Total*	**386**	**296**	**544**	**202**	**365**	**431**	**32**	**38**		

Percentages in parentheses are % of total number of ASCEND participants with the outcome reported before adjudication. The total (for each row or column) counts only the first event that occurred for each participant. *The sum of each column does not equal the total because participants may have had more than one event occur (e.g. during the follow-up period a participant may have had a non-fatal myocardial infarction and angina hospitalisation). The column “other” includes all events that were adjudicated to categories not represented by the table's other columns.

Figure 3.1 Consort diagram of analyses in the ASCEND trial



FA = Fatty acids. *A complete breakdown of exclusions can be found in ASCEND's main publications. **All 15,480 participants were included in both the aspirin and omega-3 fatty acid comparisons.

Figure 3.2 Effect of allocation to aspirin and omega−3 fatty acids vs matching placebos on any serious vascular event or revascularisation using routine data follow−up in the ASCEND trial

Log−rank methods were used to calculate the rate ratio and 95% confidence intervals. CI = Confidence intervals. TIA = Transient ischaemic attack.

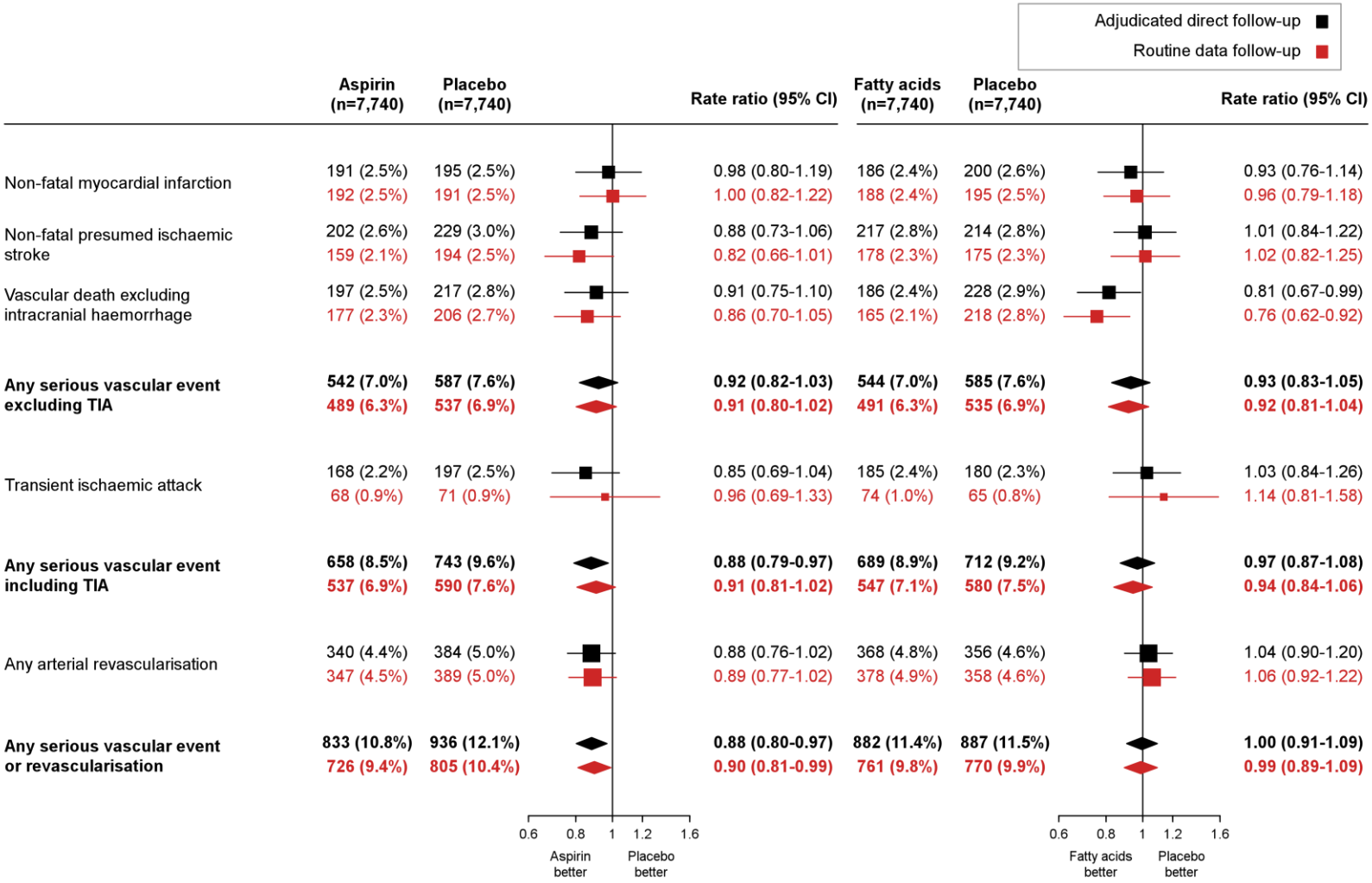


Figure 3.3 Effect of allocation to aspirin and omega-3 fatty acids vs matching placebos on any serious vascular event or revascularisation using routine data follow-up in the ASCEND trial, where adjudicated data restricted to fatal and hospitalised events only

Log-rank methods were used to calculate the rate ratio and 95% confidence intervals. Hospitalised events defined as any event where adjudicated follow-up reported that the participant was admitted to hospital. CI = Confidence interval. TIA = Transient ischaemic attack.

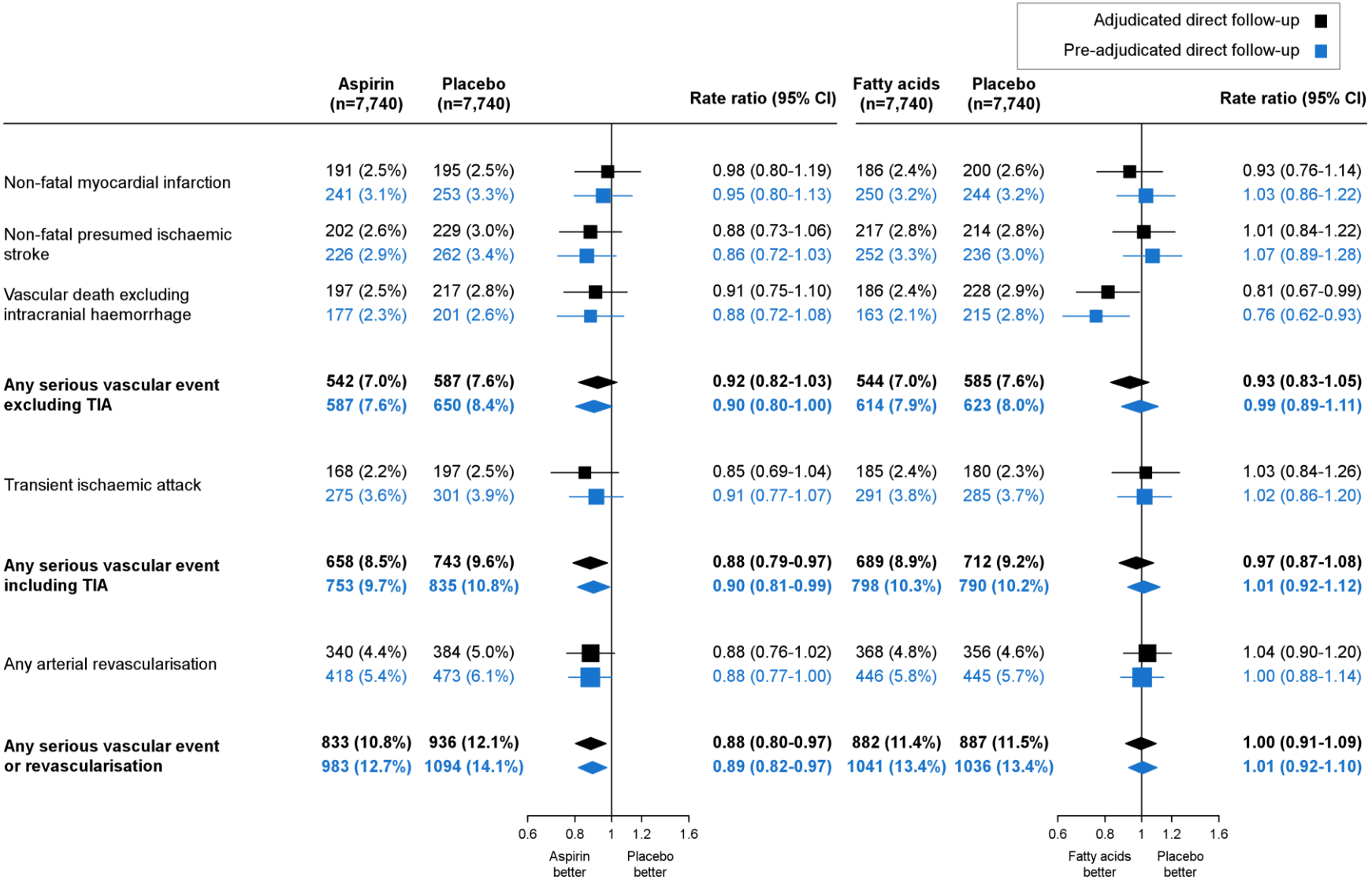
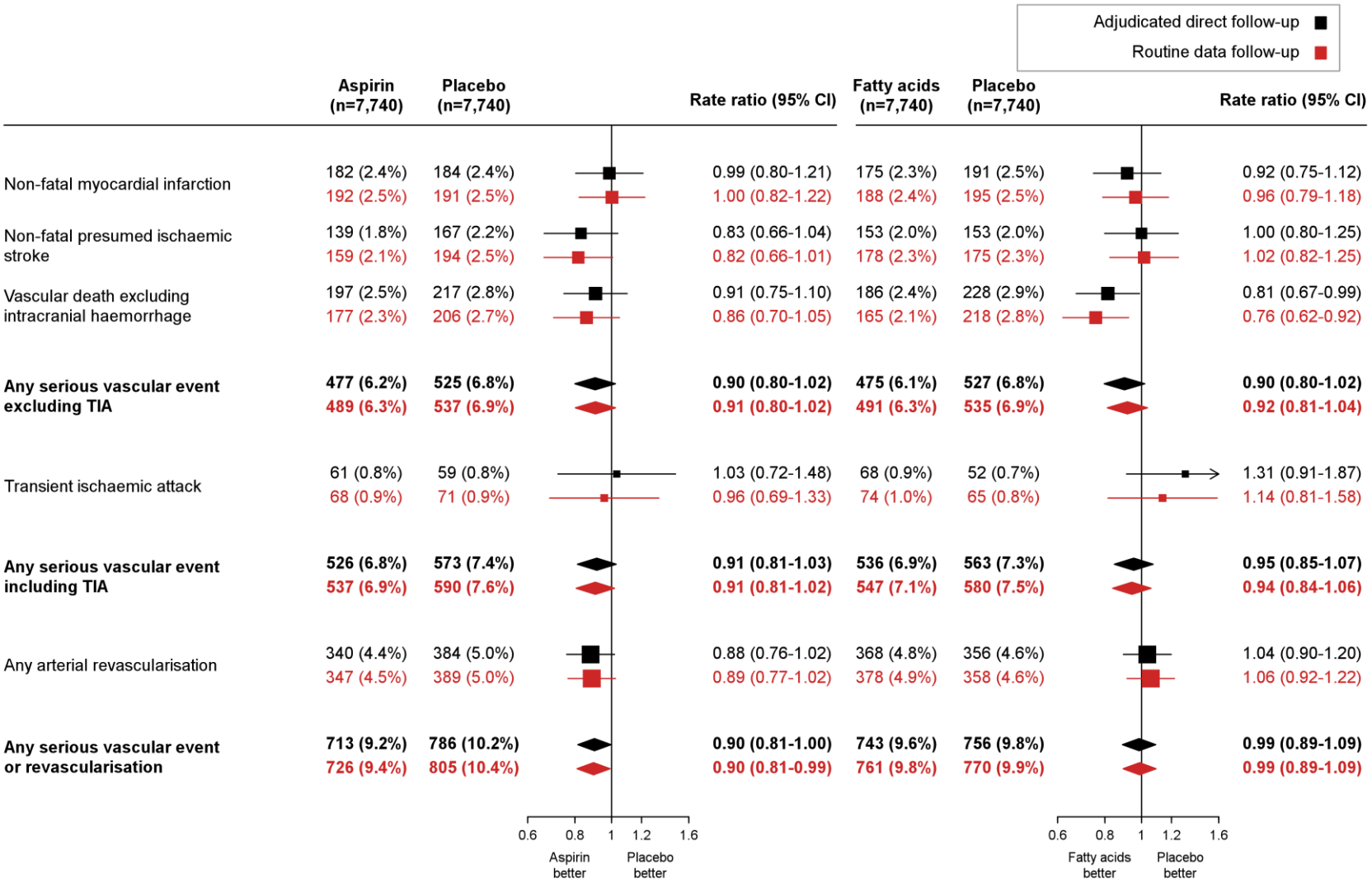


Figure 3.4 Effect of allocation to aspirin and omega−3 fatty acids vs matching placebos on any serious vascular event or revascularisation using pre-adjudicated direct follow-up in the ASCEND trial

Log−rank methods were used to calculate the rate ratio and 95% confidence intervals. CI = Confidence intervals. TIA = Transient ischaemic attack.



CHAPTER 4 Comparison of routinely collected UK data with adjudicated direct follow-up: analyses of major bleeding events from the ASCEND randomised trial

This chapter was published in BMJ Heart:

Harper C, et al. Reliability of major bleeding events in UK routine data vs clinical trial adjudicated follow-up data. BMJ Heart 2023; 109:1467-1472.

Chapter summary

Aims: To assess, in people with diabetes and no prior vascular disease, (1) how reliable UK routine data are for ascertaining major bleeding events compared with adjudicated direct follow-up data, and (2) the impact of adjudication on mail-based follow-up of major bleeding events.

Methods: The ASCEND (A Study of Cardiovascular Events in Diabetes) primary prevention trial randomised 15480 UK people with diabetes using a 2x2 factorial design to aspirin versus matching placebo. The primary safety outcome was major bleeding (a composite of intracranial haemorrhage, sight-threatening eye bleeding, serious gastrointestinal bleeding, and other major bleeding [including epistaxis, haemoptysis, haematuria, vaginal, and other bleeding]) ascertained by direct-participant mail-based follow-up, with >90% of reported major bleeding outcomes undergoing adjudication. Nearly all participants were linked to routinely collected hospitalisation and death registry data (i.e. routine data), where an algorithm was applied to categorise routine data identified bleeding events as major or minor. Kappa statistics were used to assess agreement between follow-up data sources and randomised comparisons were re run solely using routine data.

Results: When adjudicated direct follow-up and routine data were compared, there was agreement for 318 major bleeding events, with routine data identifying an additional 281 potential major bleeding events not previously reported and not identifying 241 participant-reported events (kappa 0.53, 95% confidence interval [CI] 0.49-0.57). By components of major bleeding, agreement was highest for intracranial haemorrhage (0.73, 95% CI 0.67-0.80), moderate for serious gastrointestinal bleeding (0.56, 95% CI 0.50-0.62), and lowest for sight-threatening eye bleeding (0.45, 0.35-0.54) and other major bleeding (0.36, 0.26-0.45). Repeating ASCEND's randomised comparisons using routine data follow-up alone found the estimated relative and absolute effects of allocation to aspirin versus placebo on major bleeding were similar compared to adjudicated follow-up (adjudicated follow-up: aspirin 314 [4.1%] vs placebo 245 [3.2%]; rate ratio [RR] 1.29, 95% confidence interval [CI] 1.09-1.52; absolute excess +6.3/5000 person-years [mean standard error ± 2.1]; versus routine data: 327 [4.2%] vs 272 [3.5%]; RR 1.21, 95% CI 1.03-1.41; absolute excess +5.0/5000 [± 2.2]). Relative effects between data sources were almost identical for all components of major bleeding except other major bleeding, where routine data appeared to somewhat underestimate the effects of aspirin (adjudicated follow-up: RR 1.70, 95% CI 1.18-2.44; vs, routine data: 1.08, 0.79-1.46). For pre-adjudicated participant reported mail-based follow-up, although relative effects were similar, such methods appeared to substantially overestimate the absolute hazard of aspirin (aspirin 668 [8.6%] vs placebo 549 [7.1%]; RR 1.22, 95% CI 1.09-1.37; absolute excess +11.3/5000 [± 2.1]).

Conclusions: In summary, these analyses of the ASCEND randomised trial found that major bleeding events ascertained via applying an algorithm to UK routinely collected hospital admission and death registry data provided relative and absolute treatment effects similar to adjudicated direct participant follow-up. These results suggest UK routine data represent a potentially cost-efficient method to ascertain major bleed trial outcomes during both within and post-trial periods, either to supplement existing methods or to be used on its own. But, further work is needed to validate this algorithm in other trial populations.

Finally, this study found that clinical adjudication of mail-based direct-participant follow-up may still be necessary in order to adequately weigh up the absolute harms and benefits of anti-platelet therapy.

4.1 Introduction

Large randomised trials in cardiovascular disease have provided reliable evidence for interventions whose widespread use has contributed to declines in mortality due to cardiovascular disease.[4,17] However, increasing costs and complexity of conducting trials threaten our ability to generate new randomised evidence.[19,22,97] Linkage to routinely collected healthcare data sources could substantially streamline trial follow-up by offering a cost-efficient method to ascertain outcomes.[83,84] The United Kingdom (UK) is particularly well placed as the National Health Service (NHS) has long-established routinely collected national level mortality and hospitalization records.[29,30,79] These data cover all NHS hospitals and have been collected since the 1990s using standardised clinical coding practices.[34-36]

Major bleeding is a common safety outcome for anti-thrombotic cardiovascular trials, where the assessment of whether such treatments prescribed to at-risk populations requires clinicians to balance the benefits of reducing cardiovascular risk against the increased bleeding hazard. Despite this endpoint being frequently utilised by trials, there is no single standardised definition of what is meant by “major” bleeding, with the three most common definitions including: Bleeding Academic Research Consortium (BARC), International Society on Thrombosis and Haemostasis (ISTH), and Thrombosis in Myocardial Infarction (TIMI).[77,98-101] Such definitions are complex and often require detailed information about the number of units of blood transfused, drops in haemoglobin, haemodynamic compromise, and any required medical/surgical interventions.

In **CHAPTER 2** only one study was identified (the HEART-PPCI trial) that compared major bleeding outcomes ascertained via UK routine data against adjudicated direct follow-up (defined as BARC score 3-5), finding agreement between data sources to be low.[67] But the study had substantial limitations, including: (1) no randomised comparisons were re-run using routine data only; (2) no differentiation was made between major and minor bleeding events recorded in the routine data; (3) follow-up was limited to 28 days, most of which was

during the index admission; and (4) the routine data sources currently available for trialists does not include the information to categorise bleeding events using the BARC score.

In the ASCEND[8] trial, major bleeding was the primary safety outcome for ASCEND's aspirin versus placebo randomised comparison, in which 559 events were identified using adjudicated direct mail-based follow-up. The outcome included intracranial haemorrhage, sight-threatening eye bleeding, serious gastrointestinal bleeding, and other major bleeding (including epistaxis, haemoptysis, haematuria, vaginal, and other bleeding). For gastrointestinal and other bleeding events to be considered "major" the event must have either led to or prolonged a hospitalisation, a far simpler approach than the bleeding categorised discussed above, and one that is directly applicable to the routine data sources currently available in the UK.

The analyses presented in this report aimed to use the ASCEND-linked data to assess whether, by developing and applying an algorithm to UK routine data, these data can reliably follow-up participants for major bleeding outcomes in future streamlined trials. Analyses include: (1) measuring the agreement of events identified from routine data compared to adjudicated direct-participant follow-up, and (2) re-running the aspirin versus placebo randomised comparisons using the hypothetical scenario that follow-up had been solely through linkage to routinely collected data systems. As subsidiary aims, the study re-assessed the absolute benefits of aspirin versus the bleeding hazard using routine data only, and investigated the impact of adjudication on major bleeding outcomes ascertained from mail-based direct participant follow-up.

4.2 Methods

4.2.1 The ASCEND trial

Full details of the ASCEND trial design have been reported in **CHAPTER 1** and **CHAPTER 3**.[7-9] Briefly, between 2005 and 2011, 15480 UK participants were randomised using a 2x2 factorial design to aspirin once daily versus matching placebo, and separately to omega-3 fatty acids versus placebo. For the aspirin comparison (**Supplemental Figure 4.1**), the

primary efficacy assessment was time to first serious vascular event (SVE; a composite of non-fatal myocardial infarction, presumed ischaemic stroke or transient ischaemic attack, and vascular death [excluding intracranial haemorrhage]). The primary safety assessment was time to first major bleeding event (a composite of any intracranial haemorrhage, sight-threatening bleeding in eye, serious gastrointestinal bleeding [including upper, lower or unspecified bleeding, and perforation], and other major bleeding [including epistaxis, haemoptysis, haematuria, vaginal, and other bleeding]). For gastrointestinal or other major bleeding events to meet the ASCEND criteria of major bleeding they would have either led to hospitalisation or death; full details of the bleeding subcategorization used in ASCEND are provided in **Supplemental Figure 4.2**. Sight-threatening eye bleeds included any clinically significant retinal bleeds presenting with symptoms outside of routine retinopathy screening which required laser photocoagulation, surgery or intra-ocular injections or any intra-ocular bleeding resulting in permanent visual loss. All intracranial bleeds were considered to be major. During a mean follow-up of 7.4 years, ASCEND showed that compared to placebo, allocation to aspirin reduced SVE by 12% (95% confidence interval [CI] 3% to 21%; 658 participants [8.5%] aspirin vs. 743 [9.6%] placebo), but increased the major bleeding risk by 29% (95% CI 9% to 52%; 314 [4.1%] vs 245 [3.2%]), thus the benefits of aspirin were largely counterbalanced by the bleeding hazard.[8]

4.2.2 Mail-based participant follow-up and adjudication

The principal method of direct participant recruitment and follow-up was by mailed questionnaires (on paper and in later years of the follow-up period for some participants, electronic) every 6-months until the end of the trial. On each questionnaire, participants were asked to indicate if they had suffered “Bleeding for which you saw a doctor” and to record a date and a description of the body site of the bleeding. Study clinicians blind to treatment allocation coded the bleeding site using the free-text information on the returned questionnaire, and these data are referred to in this thesis as “pre-adjudicated direct follow-up” (**Supplemental Figure 4.3**). Study clinicians blind to treatment allocation adjudicated

the primary and secondary outcomes. For non-fatal vascular and bleeding outcomes, the coordinating centre requested documentation (e.g. hospital discharge summaries) from primary care. In a small number of cases where participants were lost to follow-up, routine hospital admission data were used to ascertain 18 serious vascular events and 24 major bleeds. For fatal outcomes, the primary source of information was the Office for National Statistics death certification data, which included deaths occurring in the UK both in and out of hospital. The reported primary and subsidiary causes of death were then reviewed by study clinicians along with available hospital admissions data and any information from the trial questionnaires. The published results were based on all confirmed and unrefuted serious vascular and major bleeding event reports (i.e. a reported event for which supporting documentation could not be obtained [i.e. was unrefuted] was counted in the analysis). Over 90% of the serious vascular events and major bleeding events included in the analysis were confirmed by adjudication. Details of the adjudication definitions (including how deaths were ascribed to their underlying cause using information from review of other clinical information alongside death certificates) are provided in the Supplement accompanying the main publication.[8] These outcome data were used in ASCEND's primary publications[8,9] and are referred to as "adjudicated direct follow-up" in this thesis (**Supplemental Figure 4.3**).

4.2.3 Routinely collected death records and hospital admission data

Consent was obtained from participants to allow access to their routinely collected data. During follow-up, death records were obtained from the Office for National Statistics via NHS England for England/Wales[32] and National Health Scotland Central Register[33]. These data included date of death, underlying and other contributing causes of death. Participants were also linked to their routinely collected hospital admission records; these data were obtained from NHS England (Hospital Episode Statistics Admitted Patient Care)[31] for England, Public Health Scotland (Scottish Morbidity Records 01)[30], and the Welsh SAIL Databank (Patient Episode Database for Wales)[29]. Information used for these analyses

included: the primary and any secondary admission diagnoses; any operations and procedures; admission method; admission date; and the number of nights the participant stayed in hospital. Linkage between the trial participants and routine datasets were carried out by the national registries, but was not possible for 44 participants (44/15480 [0.3%]) residing in Northern Ireland, among which only one serious vascular event and no major bleeding event occurred during follow-up.

Bleeding episodes were identified in death records and hospital admission data using a clinician specified set of International Classification of Diseases (ICD-10) diagnosis codes (**Supplemental Table 4.1**). The date of event was assumed to be the date of admission or death. The bleeding code must have been recorded as the underlying cause on the death record to be considered fatal (**Supplemental Table 4.2**).[37] While for bleeding codes recorded in the hospital admissions data, potential indicators of event severity that were considered based on previous observational studies[102-106] (**Supplemental Table 4.3**) included: bleeding in the primary (i.e. first) or second diagnostic position on the admission record (where up to twenty diagnoses can be recorded), patients having stayed at least one or two nights in hospital, and admission via emergency means. Severity factors that were not considered due to missing or incomplete information on the routinely collected hospitalisation record included blood transfusions, medical or surgical interventions, and falls in haemoglobin. Fifteen algorithms were derived using either one-severity-factor (e.g. emergency admission), two-severity-factors in combination (e.g. primary diagnosis and emergency admission), or three-severity-factors. For each possible algorithm, kappa, sensitivity and specificity statistics were calculated using adjudicated direct follow-up as the comparator.

Of the fifteen possible algorithms, one was selected for the primary analyses by clinicians before any randomised comparisons were re-run (i.e. blinded). The selection was based on the agreement statistics, simplicity of the algorithm, and a definition broadly similar to the ASCEND trial criteria set out in **Supplemental Figure 4.2**. For any intracranial

haemorrhage or serious eye bleed, the selected algorithm defined this as an ICD-10 code in any diagnostic position (i.e. primary or secondary diagnosis); while for gastrointestinal or other bleeding to be classified as major, records were restricted to bleeding codes in the first (i.e. primary) diagnostic position and patients having stayed at least one night in hospital (full details of the selected algorithm can be found in **Supplemental Table 4.2**). Previously published methods were used to ascertain serious vascular events (**Supplemental Table 4.4**).[63] Events identified through routinely collected data did not undergo clinical adjudication, and are referred to as "routine data" in this thesis (**Supplemental Figure 4.3**).

4.2.4 Statistical analyses

Analyses included only those events that occurred between randomisation and date of death or censoring, except for participants living in Northern Ireland where routine data follow-up was censored at day zero. Outcomes identified using routine data were compared to those from adjudicated direct participant follow-up. For the major bleeding outcome (and its components), participants were categorized as having: "outcome in both datasets"; "outcome in routine data only"; "outcome in adjudicated follow-up alone"; or "no such outcome in either dataset", irrespective of the relative timings of the events. Comparisons between the data sources (i.e. routine data versus adjudicated direct participant follow-up) were performed using sensitivity and specificity statistics with 95% confidence intervals (CI; see **Appendix 1 Statistics of agreement**).[86] Overall levels of agreement were estimated using the kappa statistic[56] with 95% CI.

Subgroups were assessed using heterogeneity testing[56,87], including analyses split by median age (<63 vs ≥63-years), sex (male vs female), vascular risk score (low vs medium vs high [see ASCEND data analysis plan[7]]), and country of residence (England vs Other UK). Where there was agreement between two sources of outcome data (i.e. outcomes in both datasets), the event dates were compared. Differences were presented as: exact match (same day), 1-7 days, 8-30 days, 31-90 days, 91-180 days, and >180 days. Where

an event was reported in adjudicated follow-up alone, routine data was searched to identify whether there was a corresponding hospitalisation (within 90 days of the adjudicated event date) or death record for the same participant. A similar process was carried out to search adjudicated data for events recorded solely in routine data.

All randomised comparisons used standard log-rank methods following the intention-to-treat approach.[53,54] Rate ratios (RR) with 95% CI are provided. The main randomised comparisons were those based on the adjudicated direct follow-up data versus those based on the alternative scenario that ASCEND had only used routine data to identify outcomes. For the main randomised comparisons, differences between the rate ratios for adjudicated follow-up versus routine data were calculated; with 95% confidence intervals derived using bootstrap methods. These methods used 1000 re-samplings, with replacement, where the difference in point estimate was recalculated in each bootstrap sample. Observed absolute treatment effects were also calculated, in which results were presented as number of events per 5000 person-years including the mean standard error. Sensitivity analyses included re-running randomised comparisons using the 14 other proposed algorithms to distinguish major from minor bleeding. Secondarily, the impact of adjudication on randomised comparisons was assessed using outcomes derived from pre-adjudicated direct follow-up. Analyses were conducted using SAS version 9.4, and R version 4.2.2.

4.3 Results

4.3.1 Accuracy and completeness of routine data ascertained major bleeding events

Of the 1690 participants with a bleeding events identified from routine data sources, 137 of these were intracranial haemorrhages, 97 serious eye bleeds, 698 gastrointestinal and 911 other bleeds (**Supplemental Figure 4.3**). Applying the routine data algorithm to the gastrointestinal and other bleeding events categorised 227 (32.5%) and 164 (18%) as major, respectively. In total 559 of 1690 (35.4%) bleeding events recorded in the routine hospital and death registry data were categorised as major.

Adjudicated direct follow-up identified 559 participants with a major bleeding event, of which 318 also had a major bleeding event recorded in the routine data (sensitivity 56.9%, 95% confidence interval [CI] 52.8%-61.0%; **Table 4.1**). Routine data identified an additional 281 potential major bleeding events not previously reported by adjudicated direct follow-up, and 14640 participants had no major bleeding event in either dataset (specificity 98.1%, 95% CI 97.9%-98.3%). Overall agreement between routine data and adjudicated direct follow-up for the any major bleeding outcome was moderate (kappa 0.53, 95% CI 0.49-0.57), with similar kappa statistics across subgroups of participants (**Supplemental Table 4.5**). For the 318 participants with a major bleeding event in adjudicated direct follow-up and routine data, date of first recorded major bleeding event in both data sources were within 90 days of each other for 260 (81.8%) participants (**Supplemental Table 4.6**).

Agreement statistics for the alternative fourteen major bleeding routine data algorithms not pre-selected for analysis can be found in **Supplemental Table 4.7**. If no algorithm had been applied to the routine data, although completeness would have been good (sensitivity 74.2%, 95% CI 70.6%-77.9%), routine data would have identified 1275 additional possible bleeding events (specificity 91.5%, 95% CI 91.0%-91.9%), therefore overall agreement would have been poor (kappa 0.33, 95% CI 0.30-0.37). In general, two or more severity factors showed better agreement than a single factor. Although some algorithms performed slightly better, the selected algorithm was chosen as its definition was closest to that of the ASCEND trial (i.e. hospitalised due to a bleeding event).

Compared to adjudicated direct follow-up, agreement varied by major bleeding component, it was highest for intracranial haemorrhage with a kappa of 0.73 (95% CI 0.67-0.80) and serious gastrointestinal bleeding at 0.56 (95% CI 0.50-0.62; **Table 4.1**). Other major bleeding and sight-threatening eye bleeding had lower sensitivity at 0.45 (95% CI 0.35-0.54) and 0.36 (95% CI 0.26-0.45), respectively. For most of the major bleeding components, event dates in routine data compared to adjudicated direct follow-up were

similar, except for sight-threatening bleeding in eye where 51.0% (25/49) of events had a difference of more than 90 days.

For the 241 major bleeding events identified in adjudicated follow-up alone, within 90 days, 74 (30.7%) had a bleeding code recorded in the routine data but did not meet the algorithm's definition for major bleeding, while 93 (38.6%) had a hospitalisation but no bleeding code recorded (**Supplemental Table 4.8**). For the 281 major bleeding events identified in routine data only, in the adjudicated direct follow-up database: 11 (3.9%) events were refuted after adjudication, 10 (3.6%) had a minor bleed reported, and 223 (79.4%) had no hospitalisation reported within 90 days.

4.3.2 Randomised comparisons for major bleeding events using routine data

For the any major bleeding outcome, randomised comparisons using routine data alone provided similar estimates of the relative effects of aspirin versus placebo compared to the published results based on adjudicated direct follow-up (adjudicated direct follow-up: aspirin 314 [4.1%] vs placebo 245 [3.2%]; RR 1.29, 95% CI 1.09-1.52; routine data: 327 [4.2%] vs 272 [3.5%]; RR 1.21, 95% CI 1.03-1.41; difference in point estimate: -0.08, 95% CI -0.28, +0.12; **Figure 4.1** and **Supplemental Table 4.9**). When routine data follow-up was broken down by whether there was a corresponding event in adjudicated direct follow-up, events identified in routine data only and not in adjudicated follow-up (RR 1.13, 95% CI 0.89-1.43) were found to have a lower rate ratio than events identified by both data sources (RR 1.27, 95% CI 1.02-1.58; **Supplemental Table 4.10**). Randomised comparisons performed using the routine data algorithms not selected for analyses (by the selection process which occurred prior to un-blinding) found that the fewer severity factors included in the algorithm the lower the relative effects of aspirin. Importantly, the major bleeding hazard would have been missed if no algorithm had been applied to the routine data (RR 1.05, 95% CI 0.96-1.16; **Figure 4.2**) or data were restricted to the primary diagnosis only (RR 1.02, 95% CI 0.91-1.15).

Analyses by bleeding type were limited by low power. Nevertheless, compared to adjudicated direct follow-up, the relative effects of aspirin in the routine data were almost identical for intracranial haemorrhage (adjudicated direct follow-up: RR 1.22, 95% CI 0.82-1.81; versus, routine data: 1.24, 0.89-1.74), sight-threatening bleeding in the eye (RR 0.89, 95% CI 0.62-1.27; vs, 0.90, CI 0.60-1.34), and serious gastrointestinal bleeding (RR 1.36, 95% CI 1.05-1.75; vs, 1.41, 1.09-1.83; **Figure 4.1**). For the other major bleeding outcome, compared to adjudicated direct follow-up (RR 1.70, 95% CI 1.18-2.44), routine data provided a lower estimate (RR 1.08, 95% CI 0.79-1.46) of the relative effect, with the estimate being almost the same as when no algorithm had been applied to the routine data (RR 1.09, 95% CI 0.95-1.24; **Supplemental Table 4.11**).

Figure 4.3 provides estimates of the observed absolute effects of allocation to aspirin versus placebo for the SVE and major bleeding outcomes. Similar to adjudicated direct follow-up (SVE -8.2 per 5000 person-years [mean standard error ± 3.4] vs major bleeding +6.3 [± 2.1]), when events were ascertained solely through routine data the absolute benefits of aspirin were largely counterbalanced by the bleeding hazard (SVE -5.0 [± 3.0] vs major bleeding +5.0 [± 2.2]).

4.3.3 Effect of adjudication on mail-based direct follow-up

Of the 1215 major bleeds reported via mail-based direct follow-up, 58.4% (710) were refuted during adjudication, with 496 being considered to be minor bleeds and 214 not being considered a bleeding event (**Supplemental Table 4.12**). These results differed by bleeding type: adjudication confirmed 92.2% (83/90) of participant reported intracranial haemorrhages and just over two-thirds (69.5% [221/318]) of reported serious gastrointestinal bleeding events (**Table 4.2**). For sub-components of gastrointestinal bleeding, adjudication confirmation rates were far higher for upper (83.3% [135/162]) compared to lower gastrointestinal bleeding (47.6% [71/149]; **Supplemental Table 4.13**). Confirmation rates were low for sight threatening eye bleeding (16.8% [94/560]) and other major bleeding (30.2% [94/311]; where haematuria (23.8% [31/130]) and vaginal bleeding

(13.9% [10/72]) had the poorest confirmation rates for sub-components of the other major bleeding outcome (**Supplemental Table 4.14**).

For the any major bleeding outcome, when randomised analyses were re-run using pre-adjudicated direct follow-up, compared to adjudicated direct follow-up, the relative effects of allocation to aspirin were similar (pre-adjudicated: aspirin 666 [8.6%] vs placebo 549 [7.1%]; RR 1.22, 95% CI 1.09-1.37; **Figure 4.1**). Analyses by bleeding type showed similar relative effect estimates before and after adjudication for intracranial haemorrhage (pre-adjudicated: RR 1.30, 95% CI 0.86-1.97) and serious gastrointestinal bleeding (1.24, 95% CI 0.99-1.54); with larger differences for other major bleeding (1.27, 95% CI 1.02-1.59) and a complete shift in direction for eye bleeding (1.17, 0.99-1.37). For the any major bleeding outcome the absolute effects before adjudication were almost double that of after adjudication (+11.3 per 5000 person-years [mean standard error ±2.1]; **Figure 4.3**).

4.4 Discussion

In these analyses from the ASCEND[7-9] randomised trial, compared to adjudicated direct follow-up, this study found that by applying an algorithm to UK routinely collected hospitalisation and death registry data, major bleeding events (a composite of any intracranial haemorrhage, sight-threatening bleeding in eye, serious gastrointestinal bleeding, and other major bleeding) identified using routine data follow-up provided a similar estimate of the relative effects of aspirin versus placebo. Between the two data sources this study found moderate levels of agreement, with both routine data and adjudicated direct follow-up missing some major bleeding events. Analysis by bleeding type found that for intracranial haemorrhage, sight-threatening bleeding in eye, and serious gastrointestinal bleeding the rate ratios were almost identical; however, for other major bleeding (which included epistaxis, haemoptysis, haematuria, vaginal, and other bleeding) routine data provided a lower estimate of the relative effects of aspirin compared to adjudicated direct follow-up. Despite these differences, the interpretation of ASCEND's aspirin randomised comparison (i.e. the benefits of aspirin being largely counterbalanced by the bleeding

hazard) would have not changed had follow-up been solely via routinely collected data. These important results complement findings from ASCEND's analyses of serious vascular outcomes[63] in **CHAPTER 3** which also suggested UK routine data represent a potentially cost-efficient method to ascertain trial outcomes during both within and post-trial periods, and does not necessarily require clinician adjudicator review to provide reliable estimates of an intervention's effect.

For major bleeding events, ASCEND is one of the first published studies to compare relative and absolute treatment effects from a large-scale randomised trial using adjudicated-direct follow-up versus the hypothetical scenario that follow-up had been solely through routinely collected data. A previous study using data from the HEART-PPCI[68] trial (see **CHAPTER 2** for further details) in England concluded that, due to the low specificity of routine data identified major bleeding events (sensitivity 64.6, 95% CI 57.9–70.4; specificity 88.6%, 95% CI 86.7%-90.0%), randomised trials should avoid using such sources of follow-up data. However, the findings from this study was limited by the absence of an appropriate algorithm to subcategorise bleeding events into major or minor, our findings suggested that when no algorithm was applied to the routine data the significant bleeding hazard of aspirin would have been missed. In contrast, ADAPTABLE trial[95] in the US developed a simple routine data algorithm to define major bleeding (i.e. hospitalisation for bleeding with associated blood transfusion), that was then validated early on in the trial by adjudicating a subset of all events (yielding a positive predictive value 93%).[107] This allowed the trial to follow-up participants almost solely via routine data sources for their primary safety outcome of major bleeding, without the need for verification by clinical adjudication. The findings from ASCEND[63] provide reassurance that approaches used by the ADAPTABLE trial can be extrapolated to large-scale streamlined cardiovascular trials conducted in the UK.

In ASCEND's primary prevention diabetic population the routine data major bleeding algorithm categorised 559 of 1690 (35.4%) bleeding events recorded in the routine hospital

admission and death registry data as major bleeding events. The proportion of all recorded bleeds that were major in ASCEND was higher than a previous study following-up people hospitalised with acute coronary syndrome, where adjudication of all routine hospital admission data recorded bleeding led to 17.8% being categorised as major (213/1194).[108] However, definitions of major bleeding differed with ASCEND, as only bleeding events with a BARC[77] score of 3-5 or ISTH[100] (International Society of Thrombosis and Haemostasis) criteria were considered major, thus people hospitalised due to a bleeding event but with no further severity factors were excluded from these definitions. Additionally, the study selected a wider range of ICD-10 codes to identify bleeding events that also included codes for iron deficient anaemia and acute post haemorrhagic anaemia (i.e. codes not included in this ASCEND study definition). Despite such differences, where adjudication of bleeding subtypes may be necessary (e.g. haematuria, vaginal bleeding, epistaxis, hematemesis), this previous study proposes a viable method of event verification while maintaining the cost saving benefits of not directly contacting participants.

This analysis of the ASCEND trial highlights the challenges of ensuring complete follow-up for major bleeding outcomes when using either adjudicated direct mail-based follow-up or routinely collected data. When comparing these two methods of follow-up, we found that although there was moderate agreement, over two-hundred major bleeding were either reported in adjudicated direct follow-up alone or routine data only. Adjudicated mail-based follow-up may have missed some events, at least for intracranial haemorrhage and serious gastrointestinal bleeding, however, as these routine data only identified events were not adjudicated we cannot say for certain whether these were in fact true events. For the major bleeding events only identified in adjudicated direct follow-up, the majority had a corresponding routinely collected hospitalisation record, but either the record gave no indication of bleeding severity or no bleeding code was recorded. Although there has been an expansion in sources of routine data available for UK clinical trials,[92,109] secondary care datasets that could help categorise bleeding events are not yet collected nationally.[102]

Expanding existing hospital data flows to also include within-hospital prescribing, laboratory results, and standardised patient discharge summaries could improve our algorithm's performance at ascertaining major bleeding events from routine data.

Although this study found that routine data provided similar estimates of the relative effects of aspirin on major bleeding, these data failed to identify almost 40% of the bleeding events identified by adjudicated direct follow-up. There are clear benefits to efficiency and the streamlining of trial conduct by using routine data but this study has also identified some potential risks, in particular by missing important safety events this limits the ability for the data monitoring committee (DMC) to assess the safety of the intervention and may lead to the underestimation of absolute harm to patients. Additionally, with considerable delays in trials receiving routine data, trial teams may not able to reliably and rapidly report safety concerns to the trial chief investigators and thence to the DMC.

When components of the major bleeding outcome were analysed, agreement and treatment effect estimates between adjudicated direct follow-up and routine data varied substantially, with intracranial haemorrhage and serious gastrointestinal bleeding appearing to be reliably recorded in the routine data. For sight-threatening eye bleeding routine hospital admission data appeared to be a poor source of information, with events dates at least 6 months after adjudicated direct follow-up, or more often, no record of eye bleeding was present. These findings were corroborated by ASCEND's investigators who found that during the study most people requiring laser photocoagulation, surgery or intra-ocular injections or any intra-ocular bleeding resulting in permanent visual loss were seen in retinopathy clinics outside of the hospital inpatient setting. For other major bleeding, where a hospitalisation had to have occurred, the routine hospital admission data either didn't have a bleeding code recorded or was not recorded in the primary position. Similar to findings in the pre-adjudicated data, it remains to be challenging to identify serious haematuria and vaginal bleeding events.

A number of studies have reported on the impact of clinical adjudication on investigator reported and participant reported cardiovascular events.[63,110-113] However, the accuracy of mail-based participant reported bleeding events compared to clinical adjudication has not been previously reported. Despite the fact that a large number of pre-adjudication bleed events were refuted by the adjudication process, the relative effects of aspirin on pre-adjudicated outcomes was only slightly lower than the effects on adjudicated outcomes, presumably because aspirin also increases the risk of minor bleeding. However, as over half the participant reported major bleeding events were refuted during adjudication, the absolute effects of aspirin were found to be materially different, with pre-adjudicated follow-up providing an absolute bleeding hazard almost double that of the adjudicated data. Thus clinical adjudication may still be necessary in order to adequately weigh up the absolute harms and benefits of anti-platelet therapy (see **Figure 4.3**).

4.4.1 Strengths and limitations

This is an important study, as previous to this, little was understood of the reliability of UK routine data at ascertaining major bleeding events, in particular whether the use of such streamlined methods impact relative and absolute bleeding hazard estimates. Compared to the complex major bleeding definitions outlined in the introduction, ASCEND had a simple pragmatic definition that was directly transferable to the routine data. In addition, with over five-hundred major bleeding events, this study was able to assess the reliability of routine data ascertained major bleeding by site, thus providing insights into which major bleeding components may need to be supplemented by direct-participant follow-up. As the study assessed the impact of adjudicating mail-based participant follow-up, these results are also important for trials aiming to streamlined follow-up in non-UK countries where routine data may not be available/reliable.

There are some limitations of the current analyses. First, the ASCEND database did not have the associated medical notes to verify or refute the 281 major bleeds identified in routine data only. Secondly, ASCEND used innovative mail-based methods, so

comparisons with site-investigator reported major bleeding were not possible. Thirdly, findings are based on the particular definition of major bleeding used in ASCEND; these definitions were an adaptation of the BARC[77] and are still likely to be widely relevant. Additionally, the algorithm developed in ASCEND has not yet been validated in an external cohort, this is an important step to assess its performance before wider use. Finally, the ASCEND trial was carried out in the UK which has a national health service, the study therefore was not able to assess the reliability of routine data in non-UK countries, including those that do not have a single national healthcare provider nor if trials are conducted across different countries.

4.5 Summary

In summary, these analyses of the ASCEND randomised trial found that major bleeding events ascertained via applying an algorithm to UK routinely collected hospital admission and death registry data provided relative and absolute treatment effects similar to adjudicated direct participant follow-up. These results suggest UK routine data represent a potentially cost-efficient method to ascertain major bleed trial outcomes during both within and post-trial periods, either to supplement existing methods or to be used on its own. But, further work is needed to validate this algorithm in other trial populations. Finally, this study found that clinical adjudication of mail-based direct-participant follow-up may still be necessary in order to adequately weigh up the absolute harms and benefits of anti-platelet therapy.

Chapter 4 Tables and figures

Table 4.1 Agreement between routine data and adjudicated direct follow-up for first major bleeding event in the ASCEND trial

Outcome	Outcome in both datasets	Outcome in routine data only	Outcome in adjudicated follow-up alone	No such outcome in either dataset	Sensitivity (95% CI)	Specificity (95% CI)	Kappa (95% CI)
Intracranial haemorrhage	87 (0.6%)	50 (0.3%)	13 (0.1%)	15330 (99.0%)	87.0% (80.4%-93.6%)	99.7% (99.6%-99.8%)	0.73 (0.67-0.80)
Sight-threatening bleeding in eye	49 (0.3%)	48 (0.3%)	72 (0.5%)	15311 (98.9%)	40.5% (31.7%-49.2%)	99.7% (99.6%-99.8%)	0.45 (0.35-0.54)
Serious gastrointestinal bleeding	131 (0.8%)	96 (0.6%)	107 (0.7%)	15146 (97.8%)	55.0% (48.7%-61.4%)	99.4% (99.2%-99.5%)	0.56 (0.50-0.62)
Other major bleeding	51 (0.3%)	113 (0.7%)	66 (0.4%)	15250 (98.5%)	43.6% (34.6%-52.6%)	99.3% (99.1%-99.4%)	0.36 (0.26-0.45)
Any major bleeding	**318 (2.1%)**	**281 (1.8%)**	**241 (1.6%)**	**14640 (94.6%)**	**56.9% (52.8%-61.0%)**	**98.1% (97.9%-98.3%)**	**0.53 (0.49-0.57)**

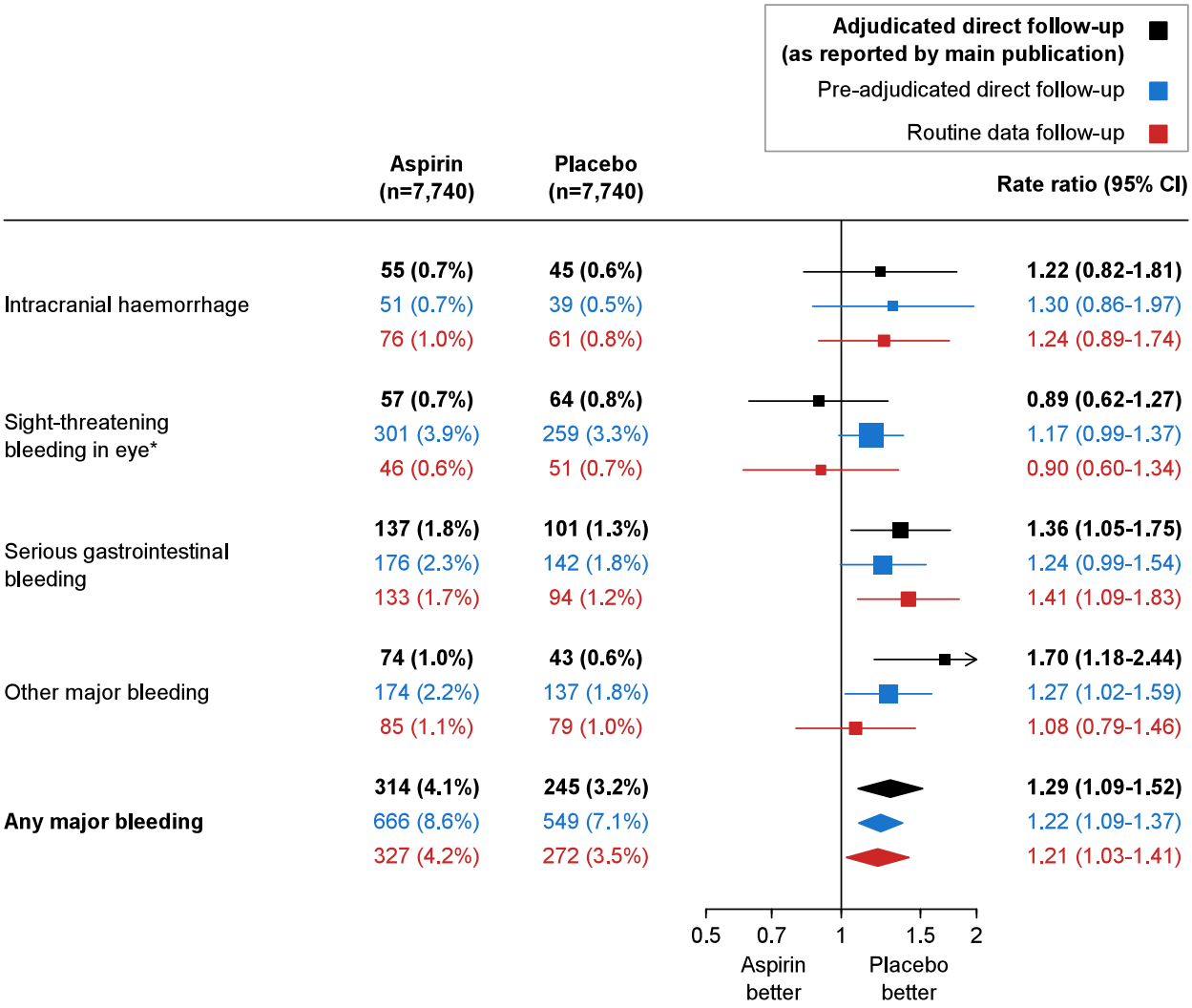
Percentages in parentheses are % of total number of ASCEND participants. Sensitivity and specificity statistics calculated using adjudicated direct follow-up as the reference dataset. CI = Confidence interval.

Table 4.2 Comparison between pre and post adjudicated direct follow-up for first major bleeding event in the ASCEND trial

	Bleeding category after adjudication						
Bleeding category before adjudication	**Intracranial haemorrhage**	**Sight threatening bleeding in eye**	**Serious gastrointestinal bleeding**	**Other major bleeding**	**Minor bleeding**	**No bleeding**	**Total**
Intracranial haemorrhage	**83 (92.2%)**	0 (0.0%)	0 (0.0%)	0 (0.0%)	0 (0.0%)	7 (7.8%)	**90**
Any eye bleeding	3 (0.5%)	**94 (16.8%)**	0 (0.0%)	0 (0.0%)	248 (44.3%)	215 (38.4%)	**560**
Serious gastrointestinal bleeding	1 (0.3%)	0 (0.0%)	**221 (69.5%)**	4 (1.3%)	65 (20.4%)	27 (8.5%)	**318**
Other major bleeding	2 (0.6%)	0 (0.0%)	3 (1.0%)	**94 (30.2%)**	190 (61.1%)	22 (7.1%)	**311**
Total*	**100**	**121**	**238**	**117**			

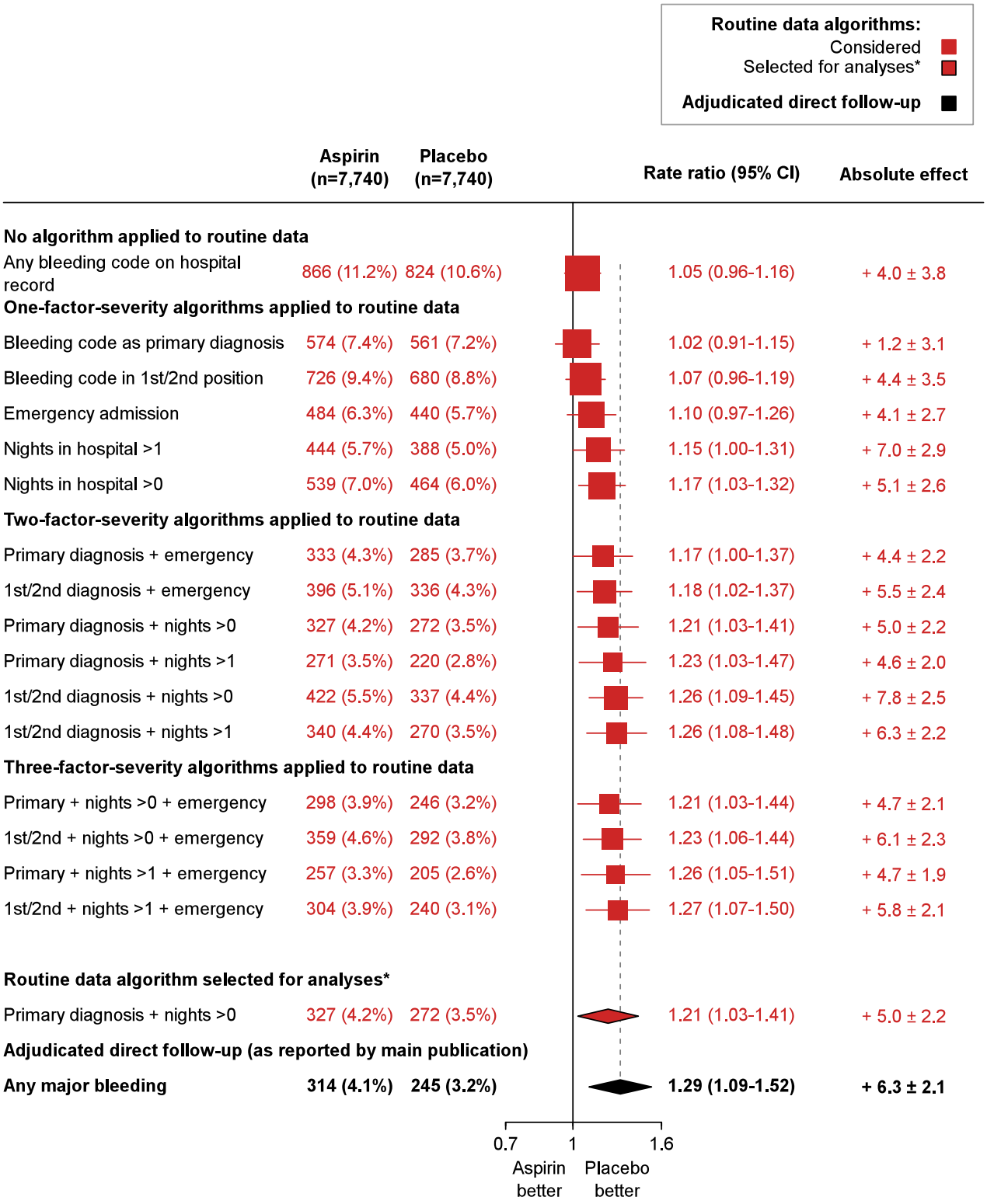
Percentages in parentheses are % of total number of ASCEND participants with the bleeding category before adjudication. The total (for each row or column) counts only the first event that occurred for each participant. The sum of each column does not equal the total because a small number of minor and non-bleeding events were categorised as major bleeding events after adjudication (see Supplemental Table 4.11).

Figure 4.1 Effect of allocation to aspirin versus placebo on any major bleeding in the ASCEND trial



Log−rank methods were used to calculate the rate ratio and 95% confidence intervals. *The pre−adjudicated direct follow−up outcome included all eye bleeds. CI = Confidence interval.

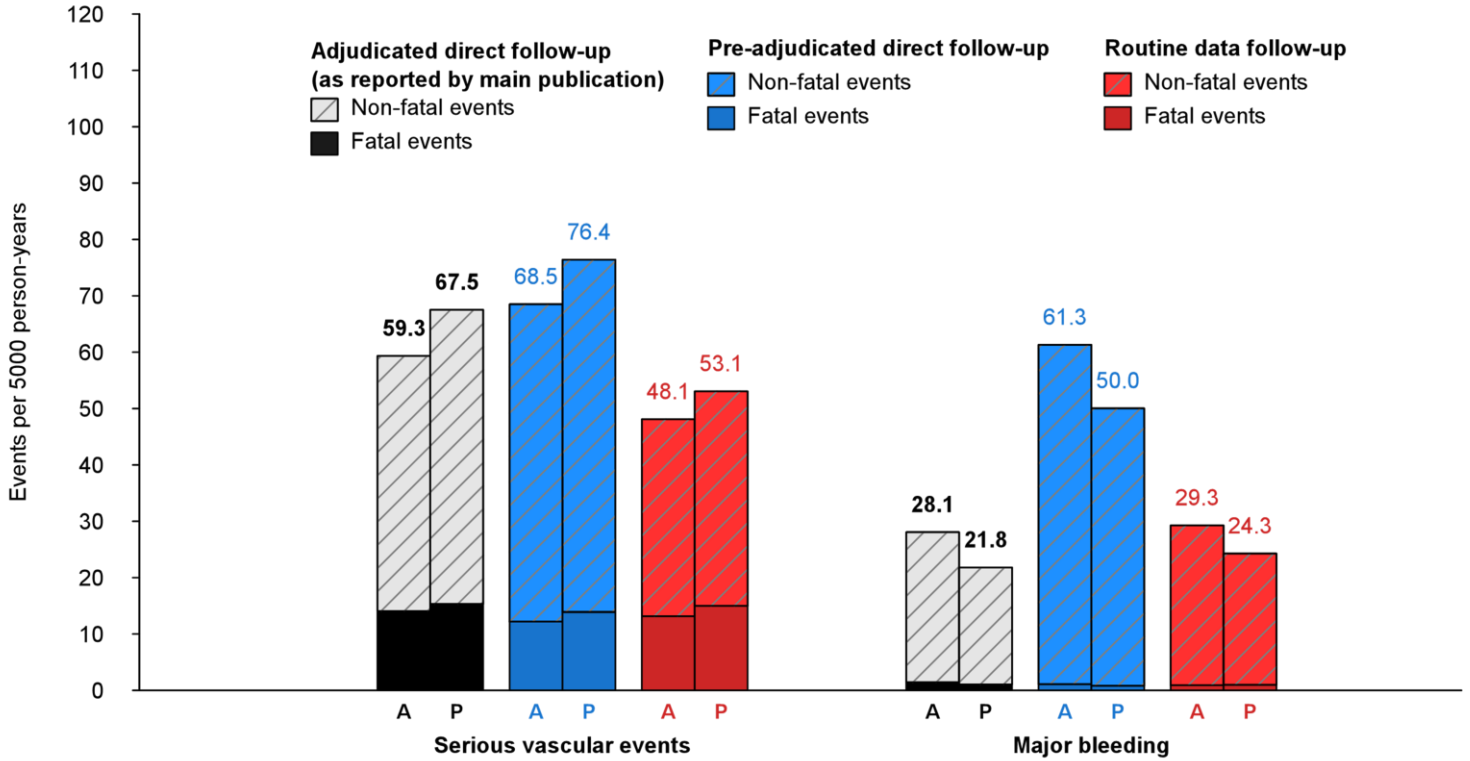
Figure 4.2 Effect of allocation to aspirin versus placebo on any major bleeding in ASCEND, by routine data algorithm



Log−rank methods were used to calculate the rate ratio and 95% confidence intervals. Absolute effects expressed as numbers of events per 5000 person−years, plus−minus values are means ± standard error. *All algorithms presented were considered, selection of algorithm for major bleeding in routine data was conducted by clinicians when blinded to results of the randomised comparison. 1st/2nd = Bleeding code in 1st or 2nd position. CI = Confidence interval. Emergency = Emergency admission. Nights = Nights in hospital. Primary = Bleeding code as primary diagnosis.

Figure 4.3 Observed absolute effects of allocation to aspirin (A) versus placebo (P) in ASCEND

Data expressed as numbers of events per 5000 person−years. Plus−minus value are means standard errors. Serious vascular events include non−fatal myocardial infarction, ischaemic stroke, transient ischaemic attack, or vascular death (excluding intracranial haemorrhage).



Absolute effects of allocation to aspirin vs placebo	Number of events per 5000 person-years in aspirin vs placebo group	
Adjudicated follow-up	**- 8.2 ± 3.4**	**+ 6.3 ± 2.1**
Pre-adjudicated follow-up	- 7.9 ± 3.6	+ 11.3 ± 2.1
Routine data	- 5.0 ± 3.0	+ 5.0 ± 2.2

CHAPTER 5 Comparison of England routinely collected data with adjudicated direct follow-up: analyses of major atherosclerotic events and kidney replacement therapy in people with CKD from the SHARP trial

The pre-adjudicated direct follow-up analysis in this chapter was published in Kidney International Reports:

Herrington WG*, Harper C*, et al. Impact of Outcome Adjudication in Kidney Disease Trials: Observations from the Study of Heart and Renal Protection. Kidney International Reports 2023; 8(8): 1489-1495. **Joint first authors.*

Chapter summary

Aims: To assess, in people with moderate-to-advanced chronic kidney disease (CKD): (1) how reliable England routine data are for ascertaining major atherosclerotic events (MAEs) and kidney replacement therapy (KRT) outcomes compared with adjudicated direct follow-up data, and (2) the impact of adjudication on clinic-based follow-up.

Methods: The Study of Heart and Renal Protection (SHARP) was a large double-blind placebo-controlled trial in 9270 people with moderate-to-advanced CKD conducted between 2003 and 2010 comparing simvastatin plus ezetimibe versus matching placebos. The primary outcome was any MAE (a composite of non-fatal MI, coronary death, non-haemorrhagic stroke, or arterial revascularisation [excluding dialysis access]), with the secondary outcome being initiation of KRT (including maintenance dialysis and kidney transplantation). The SHARP trial used clinic-based participant follow-up (i.e. pre-adjudicated direct follow-up), in which the majority of reported primary and secondary outcomes underwent adjudication (i.e. adjudicated direct follow-up).

(1) Routine data analysis

1574 SHARP participants living in England were linked to their routinely collected hospital admission and death registry data (i.e. routine data). Primary and secondary outcomes ascertained using these data were compared to adjudicated direct follow-up. For the MAE outcome, all 1574 participants were included in the analysis, while for KRT, 1109 participants not on dialysis at randomisation were included. Kappa statistics were used to assess overall agreement. There were too few events in the SHARP England cohort to re-run the trial's randomised comparisons using routine data only.

(2) Pre-adjudicated direct follow-up analysis

All 9270 SHARP participants were included in the pre versus post adjudication comparisons for MAE, while for KRT, 6245 participants not on dialysis at randomisation were included. Clinical outcome classification were compared before versus after adjudication, and then SHARP's randomised comparisons of the effects of simvastatin/ezetimibe versus placebo were re-ran using pre-adjudicated direct follow-up data.

Results:

(1) Routine data analysis

For the 1109 SHARP participants living in England and not on dialysis at baseline, 386 KRT outcomes were reported, agreement between adjudicated direct follow-up and routine data was very strong (kappa 0.87, 95% confidence interval [CI] 0.84-0.90). In the 1574 SHARP participants living in England, agreement between both data sources was moderate-to-strong for the 178 reported MAEs (kappa 0.67, 95% CI 0.61-0.72), in which agreement was stronger in participants not on dialysis at baseline. Routine data identified 100 additional MAEs, these were mostly non-coronary revascularisations related to dialysis access, and major coronary events adjudicated to non-atherosclerotic causes.

(2) Pre-adjudicated direct follow-up analysis

For the 6245 SHARP participants not on dialysis at baseline, adjudication had little impact on the KRT outcome, with only 1% of events being refuted (28/2115). Consequently, randomised comparisons using pre-adjudicated direct follow-up found almost identical results (pre-adjudication: simvastatin/ezetimibe 1038 vs placebo 1077; risk ratio [RR] 0.95, 95% CI 0.88-1.04; post-adjudicated: 1057 vs 1084; RR 0.97, 95% CI 0.89-1.05). In all 9270 SHARP participants, for MAEs, about one-quarter of patient reports were refuted (324/1275 [25%]), and reviewing 3538 other potential vascular events and death reports identified only 194 additional MAEs. Nevertheless, randomised analyses using SHARP's pre-adjudicated data alone found similar results to analyses based on adjudicated outcomes (pre-adjudication: 573 vs 702; RR 0.80, 95% CI 0.72-0.89; post-adjudicated: 526 vs 619; RR 0.83, 95% CI 0.74-0.94), and also suggested refuted MAEs were likely to represent atherosclerotic disease (RR for refuted MAEs 0.80, 95% CI 0.65-1.00).

Conclusions: In summary, for populations with moderate-to-severe CKD, routinely collected hospital admission and death registry data in England may be reliable enough to ascertain kidney replacement therapy and major atherosclerotic event outcomes, but adjudication of non-coronary revascularisations may be necessary to exclude procedures related to dialysis access. For clinic-based follow-up of KRT and MAEs, verification via clinical adjudication made no material difference to SHARP's primary results, and over strict adjudication criteria for MAEs may also have led to the trial underestimating the relative and absolute treatment effects among people on dialysis.

5.1 Introduction

Analyses presented in **CHAPTER 2** and **CHAPTER 3** have consistently shown that UK routinely collected hospital admission and death registry data are a reliable source of follow-up for non-fatal MIs, hospitalised stokes, vascular deaths, and revascularisations. Yet these analyses were limited to populations (i.e. diabetes, hypercholesterolemia, prior coronary disease) that did not include people with moderate-to-severe chronic kidney disease (CKD; i.e. estimated glomerular filtrations rate <60ml/min). People with CKD are an understudied high risk group and may be as prevalent as 10% in the UK adult general population.[114] Atherosclerotic risk is high in CKD, but non-atherosclerotic cardiac events (e.g., heart failure and sudden cardiac death) also become increasingly common as kidney function decreases, and predominate over atherosclerotic cardiovascular disease once patients require kidney replacement therapy.[59,60] More trials assessing potential interventions to reduce cardiovascular risk in CKD are required, and such trials need to test interventions that may modify atherosclerotic and non-atherosclerotic disease processes, yet they rarely attract funding at the level available for industry-funded trials.[115] Streamlined approaches to participant follow-up in kidney trials, such as utilising health data collected during a patient's routine care (i.e. routine data) and identifying where adjudication of clinic-based follow-up for particular outcomes may not necessary, could help increase the number practice-changing trials in nephrology.

In addition to people with CKD being a high risk understudied population, presentation of atherosclerotic cardiac events are often atypical and more complex to diagnose. Key reasons for this include: (1) CKD being associated with heart muscle wall thickening (ventricular hypertrophy) which can make interpretation of the electrocardiogram challenging; (2) troponin, a key marker for the diagnosis of myocardial infarction, is often elevated in people with CKD in the absence of infarction; (3) patients with CKD often develop small blood vessel disease which are not always evident on angiogram; and (4) atypical symptoms are more common than the classical presentation with chest pain.[59,60]

Therefore it is important that the reliability of UK routine data to ascertain atherosclerotic events is assessed in people with CKD, as there may be particular differences to non-CKD populations.

This chapter uses data from the SHARP[10,11] (Study of Heart and Renal Protection) trial of simvastatin plus ezetimibe versus matching placebo in people with moderate-to-severe CKD. The trial's primary outcome was major atherosclerotic events (MAEs; a composite of non-fatal MI, coronary death, non-haemorrhagic stroke, or arterial revascularisation) and secondary outcome was kidney replacement therapy (KRT; a composite of maintenance dialysis and kidney transplantation). The aim of the analyses presented in this chapter was to assess the reliability: (1) of routinely collected hospital admission and death registry data at ascertaining MAEs and KRT in SHARP participants living in England; and (2) of KRT and MAEs reported by patients in clinic compared to outcomes derived following adjudicated direct follow-up.

5.2 Methods

5.2.1 The SHARP trial

SHARP's design and results have been reported previously in **CHAPTER 1**.[10,11] Briefly, between 2003 and 2006, 9270 people with CKD were randomly assigned to simvastatin 20 mg plus ezetimibe 10 mg daily versus matching placebos. The primary composite outcome was major atherosclerotic events (MAEs) defined as non-fatal myocardial infarction (MI) or coronary death (together referred to as major coronary events), non-haemorrhagic stroke (ischaemic and unknown subtypes combined), or arterial revascularisation (coronary or non-coronary, excluding dialysis access). SHARP's secondary outcome of kidney replacement therapy (KRT) was defined as initiation of maintenance dialysis (defined as planned dialysis initiation or dialysis status for ≥3 months) or receipt of a kidney transplant. During a median follow-up of 4.9 years, compared to placebo, allocation to simvastatin plus ezetimibe reduced low density lipoprotein cholesterol (LDL) by an average of 0.85 mmol/L (standard error 0.02) and reduced risk of MAE by 17% (95% confidence interval [CI] 6% to

26%; 526 [11.3%] simvastatin/ezetimibe vs 619 [13.4%] placebo). In the subset of 6247 participants not on dialysis at baseline, study treatment had no effect on risk of needing to start KRT (1057 [33.9%] vs 1084 [34.6%]; risk ratio [RR] 0.97, 95% CI 0.89-1.05).

5.2.2 Clinic-based follow-up and adjudication

The main method of follow-up was clinic-based systematic interviews of participants (or with relatives/clinicians in the event of death) every 6 months. Participants' answers were recorded and coded in real-time directly into clinical terms on a computer-system by trained local research coordinators. Specific descriptive codes for initiation of maintenance haemodialysis and peritoneal dialysis were used, with temporary dialysis distinguished using the code "acute-on-chronic renal failure requiring dialysis". Free text descriptions of events, recoded by central study clinicians, were only used when local research coordinators (usually a trained nurse) were uncertain about an appropriate clinical term. There was no requirement for direct review of reported events by a local physician investigator to confirm the report, and are referred to as "pre-adjudicated direct follow-up" in this chapter. All deaths, potential MAEs and KRTs were then subject to clinical adjudication. The central coordinating office requested documentation (e.g. hospital discharge summaries, death certificates, clinical investigation reports, vascular access procedures) for each of these potential outcomes. Any mentioned use of lipid-modifying drugs and laboratory measurements of lipids were redacted before documents were made available to nephrologist adjudicators blind to treatment allocation. They followed a standard operating procedure containing pre-specified diagnostic criteria. Adjudication documents were available in >95% of potential outcomes and are referred to as "adjudicated direct follow-up" in this chapter.

5.2.3 Routinely collected death records and hospital admission data in England

For the SHARP participants living in England, death records were also obtained from the Office for National Statistics via NHS England[32]. These data included date of death, underlying and other contributing causes of death. Participants were also linked to their

routinely collected hospital admission records, these data were obtained from NHS England (Hospital Episode Statistics Admitted Patient Care)[31]. Information used for these analyses included: primary and any secondary diagnoses; any operations and procedures; and admission dates.

Events related to kidney replacement therapy (e.g. initiation of dialysis, recipient of kidney transplant, insertion of arteriovenous fistula or graft) were identified in hospital admission data using ICD-10 diagnosis codes in any diagnostic position (i.e. primary or secondary) and OPCS-4 procedure codes (**Supplemental Table 5.1**). A previously developed algorithm[40] by the Oxford CTSU Renal Studies Group's was then applied to the data to identify KRT events, including its components of kidney transplantation and maintenance dialysis (see **Supplemental Table 5.2** for full-details of the algorithm). In brief, the algorithm excluded kidney transplantations where a removal code (i.e. excision of rejected transplanted kidney) was recorded within 90 days, and initiation of dialysis where a code of acute kidney injury was present (with additional criteria related to an insertion of an arteriovenous fistula/graft and a diagnosis of end-stage kidney disease).

Coronary heart disease and non-haemorrhagic stroke deaths were identified from the ICD-10 underlying cause of death on the mortality record (**Supplemental Table 5.3**). Non-fatal myocardial infarctions and non-haemorrhagic strokes were identified in hospital admission data using ICD-10 diagnosis codes in any diagnostic position (i.e. primary or secondary). The event date was assumed to be the admission date as there was no diagnosis date in the available routine data. Any arterial revascularisation procedures (including dates) were identified using OPCS-4 codes (**Supplemental Table 5.3** and **5.4**). Myocardial infarctions and non-haemorrhagic strokes were assumed to be non-fatal if the date of event was between randomisation and 30 days prior to death. Events identified through routinely collected data did not undergo clinical adjudication and are referred to as "routine data" in this chapter.

5.2.4 Statistical analyses

For the routine data analysis, the MAEs comparison included all randomised participants living in England whom did not subsequently withdraw from the study (**Figure 5.1**). While the kidney replacement therapy analysis included a subset of these England participant not reported to be on dialysis at baseline (i.e. date of randomisation). Both analyses included only those events that occurred between randomisation and date of death or censoring. Outcomes identified using routine data were compared to those from adjudicated direct follow-up. For each outcome, participants were categorised as having: "outcome in both datasets", "outcome in routine data only", "outcome in adjudicated follow-up alone", or "no such outcome in either dataset". Levels of agreement between the two data sources were assessed using sensitivity and specificity with 95% confidence intervals (see **Appendix 1 Statistics of agreement**).[86] Overall levels of agreement were estimated using the kappa statistic with 95% confidence intervals.[56]

Differences in kappa statistics between subgroups were assessed using heterogeneity testing,[56,87] including analyses by participants' mean age (<62 vs ≥62-years), sex (male vs female), diabetes at baseline, prior vascular disease, and dialysis at baseline (for the MAE analysis only). Where there was agreement between the two sources of outcome data (i.e. outcome in both datasets), the event dates were compared. Differences were presented as: exact match (same day), 1-7 days, 8-30 days, 31-90 days, 91-180 days, and >180 days. Where events did not agree, adjudicated direct follow-up and routine data records (within 1 year for KRT outcomes and 90 days for MAEs) were searched in order to ascertain why the event may have been missed. Sensitivity analyses restricted to only using diagnoses in the primary position of the hospital admission record. There were two few events in the England subset of the SHARP cohort to be able to re-run the trial's randomised comparisons using routine data only.

The pre-adjudicated direct follow-up analysis included all SHARP participants randomised (**Figure 5.1**). Tabulations of pre-adjudication versus post-adjudication outcome categories

(based on the first relevant event in each category) were constructed and the proportion of pre-adjudicated outcomes refuted by adjudication calculated. Agreement statistics were not presented as the two outcome datasets were not independent. Differences between event dates before and after adjudication were presented as categories, including: exact match, 1-7 days, 8-30, 31-90, 91-180, and >180 days. Randomised comparisons of the effect of allocation to study intervention using intention-to-treat time-to-first event log-rank methods were re-run, presenting rate ratios (RRs) with 95% confidence intervals.[11,53,54,116] These analyses were conducted for maintenance KRT and MAE outcomes based on pre-adjudication participant-reports and compared with the published post-adjudication outcome data.[11,116] The 95% confidence interval for the difference between RRs was estimated using bootstrap methods. These methods used 1000 re-samplings, with replacement, where the difference in point estimate was recalculated in each bootstrap sample. For MAE, analyses were also performed using refuted, unrefuted, and "identified by adjudication" events only; and considering separately patients not on dialysis or on dialysis at randomisation. All analyses were conducted using SAS version 9.4 and R version 4.2.2.

5.3 Results

5.3.1 Baseline characteristics of SHARP participants

Of the 9270 SHARP participants, 1574 (17.0%) participants living in England were included in the routine data analyses (**Figure 5.1**). Compared to the 7696 participants not included in the routine data analysis, baseline characteristics were mostly similar, except for baseline diabetes (11.4% England vs 24.9% not England), sex (70.1% male vs 61.0%), and age ≥70 years (32.5% vs 27.3%; **Supplemental Table 5.5**).

5.3.2 Routine data analysis of KRT based on 1109 SHARP participants living in England and not on dialysis at randomisation

Of the 1574 SHARP participants living in England, the kidney replacement therapy analyses included 1109 participants not reported to be on dialysis at baseline (**Figure 5.1**). During

the trial follow-up period there were 386 participants with an unrefuted KRT event within adjudicated direct follow-up, of which 351 were also identified in the routine data, so sensitivity was 90.9% (95% confidence interval [CI] 88.1%-93.8%; **Table 5.1**). Twenty-nine KRT events were recorded in routine data alone and 694 participants had no event in either dataset, thus specificity was 96.0% (95% CI 94.6%-97.4%). Overall agreement between routine data and adjudicated direct follow-up for KRT was very strong (kappa 0.87, 95% CI 0.84-0.90) and consistent across subgroups (**Supplemental Table 5.6**). For the 351 KRT events identified in both adjudicated direct follow-up and routine data, 288 (82%) had dates accurate to within 6 months of one another (**Supplemental Table 5.7**).

When broken down by KRT subcomponents, agreement was very strong for the kidney transplant outcome (kappa 0.94, 95% CI 0.91-0.97). Of the 8 kidney transplant events reported in adjudicated direct follow-up only, one had a kidney transplant code in the routine data but did not satisfy the algorithm, one had a code for “hypertensive renal disease with renal failure”, and 6 had no renal code recorded within 1 year (**Supplemental Tables 5.8**). For the 6 transplant events recorded in routine data only, 3 had a clinic-based report for “creation of permanent arteriovenous fistula”, one had a report of acute-on-chronic renal failure requiring dialysis, and 2 had no renal event reported within 1 year.

For maintenance dialysis, agreement was also strong (kappa 0.85, 95% CI 0.88). Thirty-five dialysis events were reported in adjudicated direct follow-up alone, of which 5 (14%) had a dialysis code in the routine data but did not satisfy the algorithm, 15 (43%) had a non-dialysis renal code recorded (e.g. hypertensive renal disease with renal failure, chronic renal failure unspecified), and 15 (43%) with no renal code (**Supplemental Tables 5.9**). Of the 37 dialysis events recorded in routine data only, none were clinic reported events refuted during adjudication, 23 (62%) had a non-dialysis renal code reported (e.g. creation of permanent arteriovenous fistula, living donor renal transplantation, acute on chronic renal failure requiring dialysis), and 14 (38%) with no renal code within 1 year.

5.3.3 Routine data analysis of MAE based on 1574 SHARP participants living in England

All 1574 participants living in England were included in the MAE routine data analysis (**Figure 5.1**). Of the 178 MAEs identified in adjudicated direct follow-up, routine data identified 153, thus sensitivity was high (86.0%, 95% CI 80.9%-91.1%; **Table 5.1**). Routine data identified 100 additional MAEs not reported by adjudicated direct follow-up (including 47 non-coronary revascularisations, 31 coronary deaths, 14 non-fatal myocardial infarctions, 7 non-haemorrhagic strokes, and 1 coronary revascularisation), thus specificity was moderate (92.8%, 91.5%-94.2%). Overall agreement between the two data sources was moderate-to-strong (kappa 0.67, 95% CI 0.61-0.72). For the 153 MAEs recorded in routine data and adjudicated direct follow-up, 128 (83%) had dates in routine data accurate to within 90 days of adjudicated direct follow-up (**Supplemental Table 5.7**).

Agreement was mostly consistent within subgroups, except for small differences by age (<62 years: kappa 0.56, 95% CI 0.44-0.67; vs ≥62 years: 0.71, 95% CI 0.65-0.78), and dialysis status at baseline (dialysis: 0.58, 95% CI 0.48-0.68; vs no-dialysis: 0.72, 95% CI 0.65-0.79; **Supplemental Table 5.10**), where routine data accuracy was far higher in participants not on dialysis at baseline (dialysis: specificity 86.6%, 95% CI 83.3%-90.0%; vs no-dialysis: 95.3%, 95% CI 94.0%-96.6%). Additionally, there were no differences in agreement when MAEs were restricted to codes recorded in the primary diagnosis only (**Supplemental table 5.11**).

By components of the MAE outcome, routine data sensitivity was high for coronary deaths (87.2%, 95% CI 77.7%-96.8%) and revascularisations (93.0%, 87.6%-98.4%), but was lower for non-haemorrhagic strokes (72.2%, 57.6%-86.9%) and lowest for non-fatal MIs (63.5%, 50.4%-76.5%). For the 19 non-fatal MIs not recorded in the routine data, only one had no routine hospital record within 90 days, and 11 (58%) had a code for ischaemic heart disease (e.g. atherosclerotic heart disease and angina; **Supplemental Table 5.12**). Further

details about the MAEs reported only in adjudicated direct follow-up can be found in **Supplemental Tables 5.13 to 5.17**.

The 100 MAEs recorded in routine data only were mostly non-coronary revascularisations (specificity 96.3%, 95% CI 95.4%-97.2%), coronary deaths (97.4%, 96.6%-98.2%), and non-fatal MIs (98.6%, 98.0%-99.2%). When these events were investigated further, for the 57 non-coronary revascularisations: 6 (11%) were patient-reported non-coronary revascularisations refuted after adjudication (e.g. infected toe, transplant renal artery angioplasty); 27 (47%) had a procedure related to dialysis access (i.e. arteriovenous fistula/graft) or kidney transplant; 11 (19%) with a non-KRT related event; and 13 (23%) with no event reported within 90 days (**Supplemental Table 5.17**). For the 40 coronary deaths in routine data only: 8 (20%) were reported as coronary deaths then adjudicated to another cause (e.g. cardiac death, heart failure, limb ischaemia); 28 (70%) were both reported and adjudicated as a non-coronary cause of death (e.g. death out-of-hospital, cardiac death, heart failure); and 4 (10%) had no death record in the adjudicated follow-up database (**Supplemental Table 5.13**). Finally, for the 22 non-fatal myocardial infarctions: 6 (27%) were patient-reported myocardial infarctions refuted after adjudication; 12 (55%) had a non-MI event (e.g. angina, atrial fibrillation, coronary angioplasty, heart failure, pericarditis); and 4 (18%) had no event within 90 days (**Supplemental Table 5.12**).

5.3.4 Pre-adjudicated direct follow-up analysis of KRT based on 6245 SHARP participants not on dialysis at randomisation

For the 6245 SHARP participants not on dialysis at randomisation, 2115 KRT events were reported, of which less than 1% (n=28) were refuted by adjudication (**Table 5.2**). Only one of the 504 reported kidney transplant events was refuted. Adjudicating 312 events reported as acute-on-chronic renal failure events requiring dialysis identified an additional 54 maintenance KRT events. Differences in reported dates of start of maintenance KRT between datasets of >30 days were uncommon (<7%, **Supplemental Table 5.18**).

Randomised analyses using pre-adjudicated direct follow-up showed that allocation to simvastatin/ezetimibe had no effect on risk of maintenance KRT (simvastatin/ezetimibe 1038 vs placebo 1077; RR 0.95, 95% CI 0.88-1.04), almost identical to results from adjudicated direct follow-up (1057 vs 1084; RR 0.97, 95% CI 0.89-1.05; **Figure 5.2**). The 95% CI for the -0.01 difference between these RRs was -0.03 to 0.00. Similarly, RRs estimated from pre- and post-adjudicated direct follow-up were almost identical for the outcomes of initiation of maintenance dialysis and receipt of a kidney transplant considered separately.

5.3.5 Pre-adjudicated direct follow-up analysis of MAE based on 9270 SHARP participants

Among the full 9270 participant SHARP cohort (including patients on dialysis), 1275 first MAEs were reported, of which 25% (324/1275) were refuted by adjudication (**Table 5.2**). By components of the MAE outcome, major coronary events had the highest proportion of refuted events (214/512 [42%]), with coronary death reports (111/198 [56%]; **Supplemental Table 5.19**) being more likely to be refuted than non-fatal MI reports (132/346 [38%]). Non-haemorrhagic stroke reports (102/340 [30%]), and revascularisations (102/646 [16%]) were least likely to be refuted. Of the MAEs which were refuted, 71% led to hospitalisation versus 96% of unrefuted MAEs, but when a hospitalisation did result, median stay duration was similar (7 [interquartile range 2-14] versus 7 [4-15] days, respectively; **Supplemental Table 5.20**). After adjudicating 3538 other potentially relevant event reports (e.g. non-coronary CV deaths and all non-CV deaths, angina events, transient ischaemic attacks, coronary angiograms; see **Supplemental Table 5.21** for the full list of terms), 194 additional first MAEs that had not been reported by sites were identified from collected adjudication source documents. These included 48 non-fatal myocardial infarctions and 58 CHD deaths (**Table 5.2** footnote).

Reported MAEs were more likely to be refuted among patients on dialysis at baseline for non-fatal MIs (dialysis vs non-dialysis: 45% vs 34%) and coronary revascularisations (25%

vs 10%; **Supplemental Table 5.22**). Differences of >30 days for start dates of MAE between datasets were uncommon (11% of MAE; **Supplemental Table 5.18**).

Analyses comparing the effects of allocation to simvastatin/ezetimibe versus placebo on pre-adjudication MAE reports demonstrated similar findings to those using adjudicated follow-up data (**Figure 5.2**). Allocation to simvastatin/ezetimibe reduced the risk of pre-adjudicated MAE by 20% (simvastatin/ezetimibe 573 vs placebo 702; RR 0.80, 95% CI 0.72-0.89), similar to the 17% reduction in risk of adjudicated MAE (526 vs 619; RR 0.83, 95% CI 0.74-0.94; difference between RRs -0.03, 95% CI -0.10, +0.03). RRs based on adjudicated and pre-adjudication follow-up were similar across the components of MAE. The RR for the effect of treatment allocation on the 324 MAEs refuted by adjudication was 0.80 (95%CI 0.65-1.00), and 1.02 (95%CI 0.77-1.35) for the 194 MAEs identified by adjudication of other potentially relevant event reports (**Supplemental Table 5.20**).

In analyses comparing patients by baseline dialysis status, adjudicated data showed that allocation to simvastatin plus ezetimibe provided a significant reduction in MAEs for the 6245 participants' not on dialysis (RR 0.78, 95% CI 0.67-0.91) but not for the 3025 people on dialysis (0.90, CI 0.75-1.08; **Figure 5.3**). Pre-adjudicated data showed that the 20% relative risk reduction observed overall in the SHARP trial was consistent regardless of dialysis status (no-dialysis: RR 0.81, 95% CI 0.70-0.94; dialysis: 0.78, 0.66-0.92).

5.4 Discussion

Analyses from the SHARP trial, in people with moderate-to-severe CKD, found that information on maintenance kidney replacement therapy (i.e. maintenance dialysis or kidney transplantation) was highly accurate and complete both in England routine data and reported by participants to trained research coordinators (i.e. pre-adjudicated direct follow-up), with pre-adjudicated data providing almost identical relative treatment effects estimates to adjudicated direct follow-up. For major atherosclerotic events, while both routine data and pre-adjudicated follow-up successfully identified the majority of MAEs, these data sources appeared to identify additional coronary events (i.e. non-fatal MI and coronary

death) that adjudicators considered to be non-atherosclerotic (e.g., heart failure and sudden cardiac death). Additionally, routine data were not able to be able to differentiate between non-coronary revascularisations related to dialysis access vs not. Despite these disagreements, relative treatment effect estimates from pre-adjudicated data were very similar to adjudicated direct follow-up, with benefits seen both in people on and not on dialysis at time of randomisation.

This is the first study to have assessed, in people with moderate-to-severe CKD, how reliable England routine data are at ascertaining KRT and MAE outcomes, compared to adjudicated direct follow-up. The validation of the KRT algorithm[40] presented in this report provides a robust and cost-effective method of ascertaining this renal outcome both for randomised trials and observational studies. With UK Biobank[43] already using a simplified version of this algorithm and NHS England's Hospital Episode Statistics Admitted Patient Care data[31] linked to other large UK research cohorts,[44,85,117,118] there is the potential to have a single definition of KRT throughout the UK research community. Furthermore, the assessment of the reliability of routine data ascertained cardiovascular outcomes in people with CKD, extends the work conducted in **CHAPTER 2** and **CHAPTER 3** in people with diabetes, hypercholesterolemia, and prior coronary disease. This SHARP analyses highlights the challenges of diagnosing atherosclerotic cardiac events in people with advanced CKD, particular those on maintenance dialysis, and this directly impacts the recorded coded information in the routine data.

These analyses form the SHARP trial found that both routine data and clinic-based follow-up identified many additional coronary events that adjudicators classified as non-atherosclerotic. But, re-run randomised comparisons using pre-adjudicated direct follow-up showed a strong treatment effect for the refuted MAEs, indicating that these may have been "real" atherosclerotic events excluded by overly strict adjudication criteria. This also may be true for the events identified in routine data only, and similar to the pre-adjudicated data, most of these occurred in participants on maintenance dialysis. Therefore routine data may

be reliable enough to be the sole source of follow-up for major coronary events, non-haemorrhagic strokes and coronary revascularisations in people with moderate-to-severe CKD.

SHARP's definition of non-coronary revascularisation excluded procedures related to dialysis access (i.e. creation and maintenance of arteriovenous fistula/graft) as kidney replacement therapy was a separate secondary outcome. Analysis of routine hospital admission data in England found that differentiating between non-coronary revascularisation procedures related to and not related to dialysis access (by carefully selecting OPCS-4 procedure codes) was very challenging. Therefore for streamlined kidney trials in the UK, it may be that a two stage process is required, where (1) routine data are used to flag all non-coronary revascularisations that have occurred, and (2) source documentation (i.e. hospital discharge summary) are collected in order for an adjudicator to assess whether the procedure was related to dialysis access.

In the SHARP analysis of pre-adjudicated data, these observations based on populations from North America, Europe and Asia provide the largest and most generalisable experience of adjudicating KRT and cardiovascular outcomes in a CKD trial. The SHARP MAE data also confirm that previous outcome adjudication experiences from diabetes and cardiology trials are relevant to CKD trials. SHARP results mirror those from the ASCEND trial in diabetes[119] and from 42 cardiology trials (including over 200000 participants) which together have consistently reported that that relative treatment effects on cardiovascular outcomes were not materially changed by clinical adjudication.[120,121]

These SHARP data may also provide a rationale for not requiring adjudication of MAEs. Despite the fact that substantial re-categorisation occurred during adjudication, this had no material impact on the RRs for confirmed, refuted and newly identified outcomes, which were similar. In fact, it is possible that SHARP's approach to cardiovascular outcome adjudication may have reduced SHARP's power to detect treatment effects, especially in subgroups. The similar sized RRs for the 1145 confirmed MAEs and the 324 refuted MAEs

suggest that many of the cardiovascular events that did not fulfil the pre-specified criteria for confirming MAEs, the majority of which led to hospitalisation for several days, were prevented by lowering LDL-cholesterol and hence represented modifiable atherosclerotic disease. It was notable that agreement between pre- and post-adjudicated datasets was less strong for major coronary events than other MAE components (non-haemorrhagic stroke and revascularisation procedures). This may reflect the challenges of distinguishing between atherosclerotic and non-atherosclerotic cardiac events based on clinical findings in patients with advanced CKD.[59,60] The corollary of these SHARP observations is that if cardiovascular outcome adjudication is deemed scientifically important in a CKD trial, then outcome definitions for major coronary events should allow for inclusion of atypical presentations and incomplete information. Avoiding overly strict and overly specific adjudication criteria could reduce the chances of "real" outcomes being refuted and maintain trial power. This may have been particularly true among patients on dialysis who's reported MAEs were more likely to be refuted. Both the relative and absolute effects of lowering LDL-cholesterol on MAEs in patients on dialysis may have been underestimated in SHARP.

5.4.1 Strengths and limitations

The strengths of these analyses are already discussed above, but they do have some limitations. First, routine data linkage was only possible for a subset of the trial population and this substantially limited the power of the study to investigate how accurately and completely routine data ascertained MAEs. Secondly, SHARP had limited clinical information on some of the MAEs identified in routine data only, thus we were not able to definitively confirm or refute whether these were "true" events. Finally, the routine data study did not have enough power to re-analyse SHARP's randomised comparisons, therefore it was not possible to assess whether the disagreement with the adjudicated direct follow-up database would have led to routine data follow-up generating different relative treatment effect estimates to SHARP's published results.[11]

5.5 Summary

In summary, for populations with moderate-to-severe CKD, routinely collected hospital admission and death registry data in England may be reliable enough to ascertain kidney replacement therapy and major atherosclerotic event outcomes, but adjudication of non-coronary revascularisations may be necessary to exclude procedures related to dialysis access. For clinic-based follow-up of KRT and MAEs, verification via clinical adjudication made no material difference to SHARP's primary results, and over strict adjudication criteria for MAEs may also have led to the trial underestimating the relative and absolute treatment effects among people on dialysis.

Chapter 5 Tables and figures

Table 5.1 Agreement of routine data versus adjudicated direct follow-up for kidney replacement therapy and major atherosclerotic events in SHARP

Outcome	Outcome in both datasets	Outcome in routine data only	Outcome in adjudicated follow-up alone	No such outcome in either dataset	Sensitivity (95% CI)	Specificity (95% CI)	Kappa (95% CI)
Kidney transplant*	129 (11.6%)	6 (0.5%)	8 (0.7%)	968 (87.1%)	94.2% (90.2%-98.1%)	99.4% (98.9%-99.9%)	0.94 (0.91-0.97)
Maintenance dialysis*	308 (27.7%)	37 (3.3%)	35 (3.2%)	731 (65.8%)	89.8% (86.6%-93.0%)	95.2% (93.7%-96.7%)	0.85 (0.81-0.88)
Kidney replacement therapy*	**351 (31.6%)**	**29 (2.6%)**	**35 (3.2%)**	**696 (62.6%)**	**90.9% (88.1%-93.8%)**	**96.0% (94.6%-97.4%)**	**0.87 (0.84-0.90)**
Non-fatal myocardial infarction	33 (2.1%)	22 (1.4%)	19 (1.2%)	1500 (95.3%)	63.5% (50.4%-76.5%)	98.6% (98.0%-99.2%)	0.60 (0.48-0.72)
CHD death	41 (2.6%)	40 (2.5%)	6 (0.4%)	1487 (94.5%)	87.2% (77.7%-96.8%)	97.4% (96.6%-98.2%)	0.63 (0.52-0.73)
Any major coronary event	**71 (4.5%)**	**52 (3.3%)**	**20 (1.3%)**	**1431 (90.9%)**	**78.0% (69.5%-86.5%)**	**96.5% (95.6%-97.4%)**	**0.64 (0.56-0.72)**
Non-fatal non-haemorrhagic stroke	15 (1.0%)	9 (0.6%)	9 (0.6%)	1541 (97.9%)	62.5% (43.1%-81.9%)	99.4% (99.0%-99.8%)	0.62 (0.44-0.79)
Fatal non-haemorrhagic stroke	9 (0.6%)	7 (0.4%)	3 (0.2%)	1555 (98.8%)	75.0% (50.5%-99.5%)	99.6% (99.2%-99.9%)	0.64 (0.42-0.86)
Any non-haemorrhagic stroke	**26 (1.7%)**	**10 (0.6%)**	**10 (0.6%)**	**1528 (97.1%)**	**72.2% (57.6%-86.9%)**	**99.3% (98.9%-99.8%)**	**0.72 (0.59-0.84)**
Coronary revascularisation	50 (3.2%)	3 (0.2%)	3 (0.2%)	1518 (96.4%)	94.3% (88.1%-100.0%)	99.8% (99.6%-100.0%)	0.94 (0.89-0.99)
Non-coronary revascularisation	31 (2.0%)	57 (3.6%)	3 (0.2%)	1483 (94.2%)	91.2% (81.6%-100.0%)	96.3% (95.4%-97.2%)	0.49 (0.37-0.62)
Any revascularisation	**80 (5.1%)**	**56 (3.6%)**	**6 (0.4%)**	**1432 (91.0%)**	**93.0% (87.6%-98.4%)**	**96.2% (95.3%-97.2%)**	**0.70 (0.63-0.77)**
Any major atherosclerotic event	**153 (9.7%)**	**100 (6.4%)**	**25 (1.6%)**	**1296 (82.3%)**	**86.0% (80.9%-91.1%)**	**92.8% (91.5%-94.2%)**	**0.67 (0.61-0.72)**

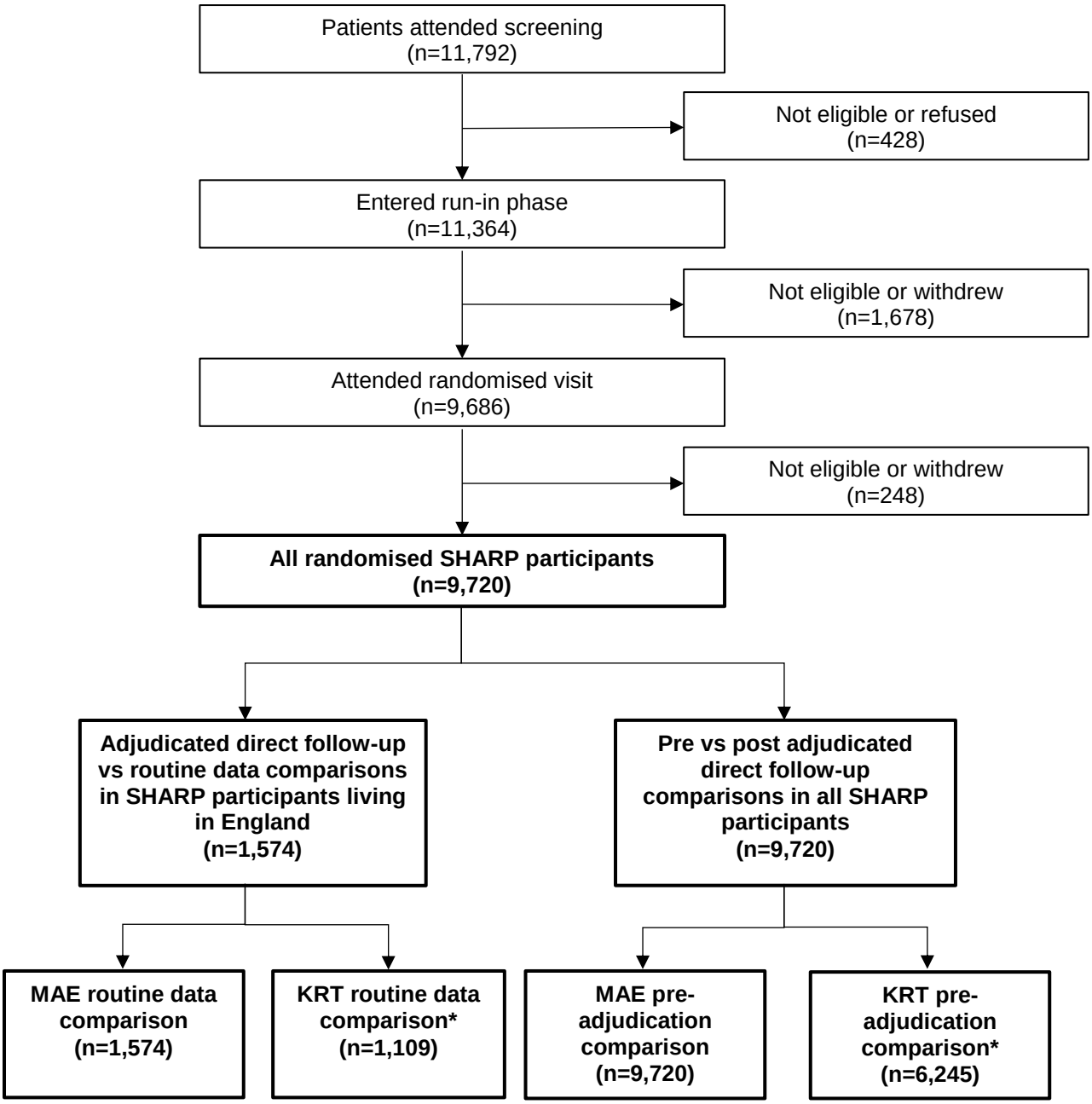
Percentages in parentheses are % of total number of SHARP participants living in England. *Excluding those on maintenance dialysis at baseline. CI = Confidence interval. CHD = Coronary heart disease.

Table 5.2 Tabulation of kidney replacement therapy and major atherosclerotic events before and after adjudication in SHARP

Outcome	Total events before adjudication	Unrefuted by adjudication	Refuted by adjudication	Identified by adjudication of other reported events	Total events after adjudication
Kidney transplant*	**504**	503 (>99%)	1 (<1%)	4	**507**
Maintenance dialysis*	**1959**	1930 (99%)	29 (1%)	61	**1991**
Kidney replacement therapy*	**2115**	**2087 (99%)**	**28 (1%)**	**54**	**2141**
Non-fatal myocardial infarction	**346**	214 (62%)	132 (38%)	79	**293**
CHD death	**198**	87 (44%)	111 (56%)	94	**181**
Any major coronary event	**512**	**298 (58%)**	**214 (42%)**	**145**	**443**
Ischaemic stroke	**297**	210 (71%)	87 (29%)	61	**271**
Unknown stroke	**57**	6 (11%)	51 (89%)	31	**37**
Any non-haemorrhagic stroke	**340**	**238 (70%)**	**102 (30%)**	**67**	**305**
Coronary revascularisation	**340**	283 (83%)	57 (17%)	69	**352**
Non-coronary revascularisation	**356**	285 (80%)	71 (20%)	38	**323**
Any revascularisation	**646**	**544 (84%)**	**102 (16%)**	**92**	**636**
Any major atherosclerotic event	**1275**	**951 (75%)**	**324 (25%)**	**194**	**1145**

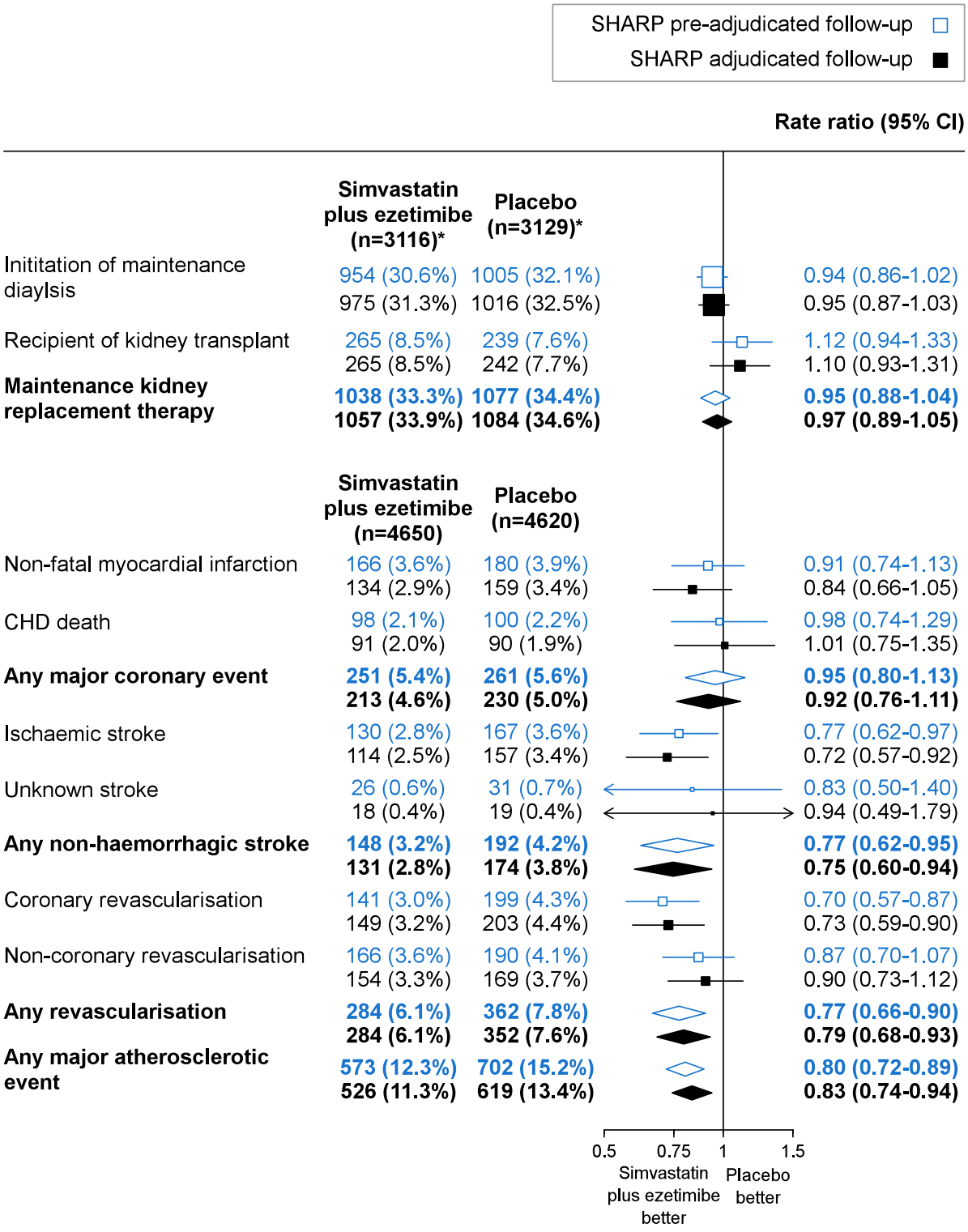
Percentages in parentheses are % of total number of SHARP participants with the outcome reported before adjudication. *Excluding those on maintenance dialysis at baseline. The 194 first major atherosclerotic events identified by adjudication of 3538 other reported events included: 48 non-fatal myocardial infarctions, 58 CHD deaths, 36 ischaemic strokes, 19 unknown strokes, 12 coronary revascularisations, and 21 non-coronary revascularisations. CHD = coronary heart disease.

Figure 5.1 Consort diagram of analyses in the SHARP trial



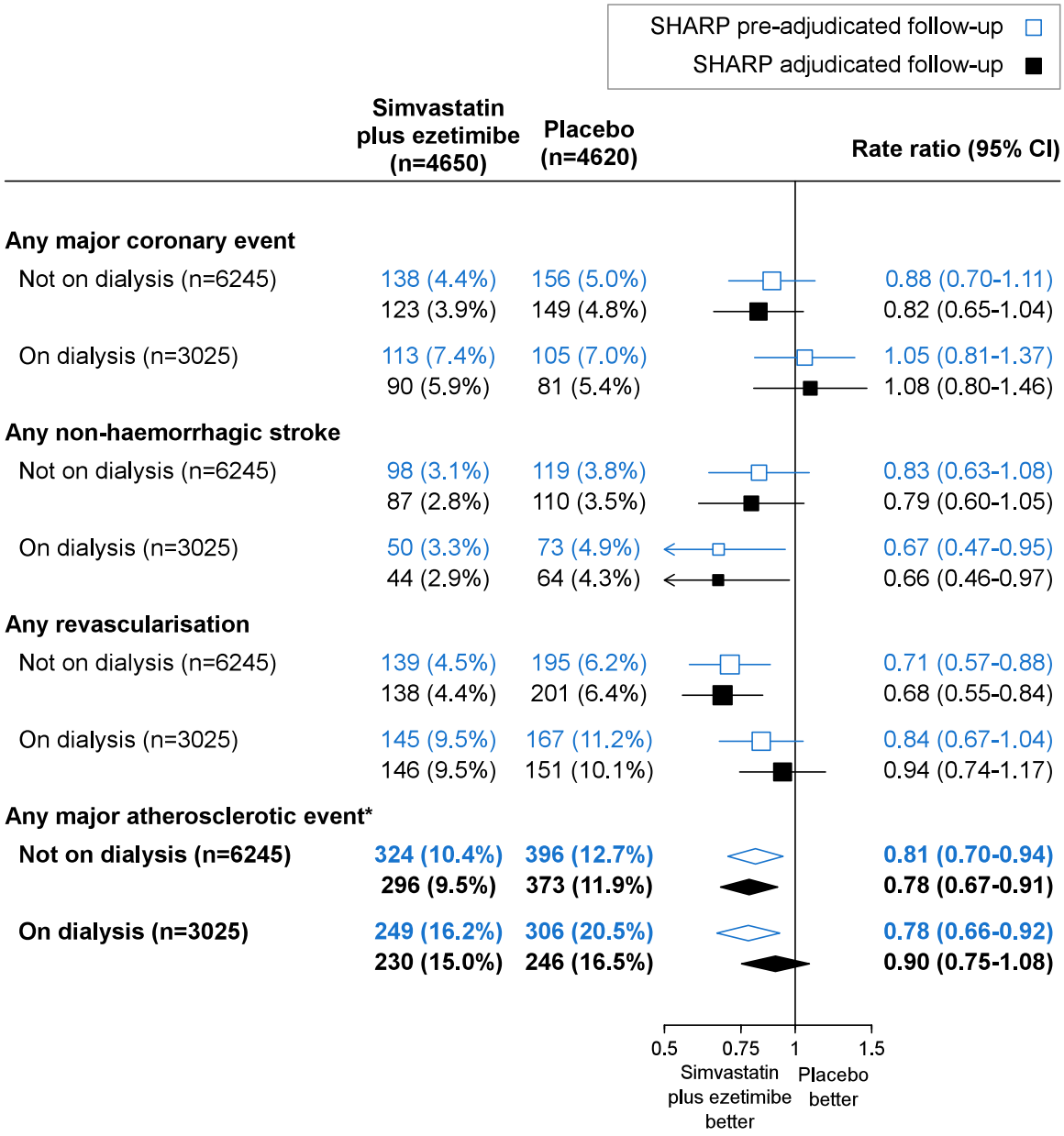
*Participants reported to be on dialysis at baseline were excluded from these analyses.

Figure 5.2 Effect of allocation to simvastatin/ezetimibe vs placebo on maintenance kidney replacement therapy and major atherosclerotic events using SHARP pre−adjudicated follow−up



Log−rank methods were used to calculate the rate ratio and 95% confidence intervals. *Excluding those on maintenance dialysis at baseline. CI = confidence interval. CHD = coronary heart disease.

Figure 5.3 Effect of allocation to simvastatin/ezetimibe vs placebo on major atherosclerotic events using SHARP pre−adjudicated follow−up, by dialysis status



Log−rank methods were used to calculate the rate ratio and 95% confidence intervals. *Heterogeneity between patients on dialysis and not on dialysis: p = 0.24 for adjudicated follow−up and p = 0.74 for pre−adjudicated follow−up. CI = confidence interval.

CHAPTER 6 Simulation study assessing the impact of removing true events and adding false events on randomised trials

Chapter summary

Aims: This simulation study aimed to investigate the impact on randomised controlled trials of removing true events and adding false events, as may happen if routine data are used for follow-up.

Methods: Multiple parallel group randomised trials (i.e. active intervention vs placebo) with a binary outcome were generated based on two parameters: the relative effect of allocation to the active intervention arm estimated by a risk ratio (RR) and overall trial sample size (n). The number of events in the placebo arm were pre-determined using the assumption each trial would have 90% power at a 5% significant level to detect the selected RR. These data were considered the ground-truth and assumed to be equivalent to "best available data". To compare the results to previous chapters, the parameters for the generated randomised trials were initially set to match those of ASCEND's (n=15480) aspirin comparisons for any serious vascular event (SVE, RR=0.88) and major bleeding (RR=1.29). Scenarios using smaller (n=1000 and 5000) and even larger (n=25000) sample sizes were then considered with different treatment effects, these included ground truths based on different degrees of relative benefit (RR 0.70, 0.80, 0.90) and harm (RR 1.10, 1.20, 1.30).

To simulate errors that might occur when using routine data follow-up, true events were removed and/or false events added in a range of different scenarios. Each participant had an equal probability of having an error (i.e. errors were random rather than systematic). The following two parameters were used: the proportion of true events removed (TER; calculated by dividing the number of true events removed by the total number of true events before any were removed) and proportion of false events added (FEA; calculated by dividing the number of false events added divided by the total number of true events before

any were removed). FEA was used in preference to specificity as it is independent of the event rate. The different scenarios considered errors of between 0% and 40%. The primary performance measure was the mean difference between the ground-truth RR versus routine data RR. The secondary performance measure was the difference in power between the best available data versus simulated routine data with errors.

Results: For the ASCEND SVE (n=15480, RR=0.88) and major bleeding event (n=15480, RR=1.29) scenarios, the simulated removal of true events had no impact on the estimated RR, while adding false events modestly biased the treatment effect towards the null (SVE: mean difference in RR: +0.01 [FEA=10%] to +0.04 [FEA=40%]; major bleeding: -0.03 to -0.10). When both false events were added and true events removed, the proportion of all trial events that were false marginally increased, further biasing the routine data RR towards the null (SVE: +0.01 [FEA & TEA=10%] to +0.05 [FEA & TEA=40%]; major bleeding: -0.03 to -0.13). Each error type individually reduced trial power, with true events removed (SVE: difference in power: -3.5% [TER=10%] to -21.4% [TER=40%]; major bleeding: -3.9% to -21.0%) having a greater impact than false events added (SVE: -3.3% [FEA=10%] to -14.9% [FEA=40%]; major bleeding: -3.2% to -12.5%). Notably, the effects on power more than doubled when both error types were present (SVE: -7.5% [FEA & TEA=10%] to -48.9% [FEA & TEA=40%]; major bleeding: -6.8% to -42.4%).

For simulations where only small-to-moderate random errors were introduced (defined as FEA≤20% and TER≤20%) there appeared to be little impact on relative treatment effect estimates across all trial sample size and treatment effect scenarios. Furthermore, only moderate reductions in power were observed (n=5000 and 25000), with the exception of smaller trials (n=1000).

Conclusions: Analysing simulated trial data from a wide range of trial scenarios showed that routine data sources with small-to-moderate random errors (i.e. true event removed and false events added ≤20%, which are equally distributed between trial arms) have little impact on relative treatment effect estimates in randomised trials. In the situation where

routine data sources do introduce large amounts of error into a trial (i.e. true event removed and false events added >20%), larger trials (n≥5000) are better protected against random errors biasing the treatment effect towards the null and reducing trial power. These results provide further rationale to consider using UK routine data as the sole source of follow-up for future cardiovascular trials.

6.1 Introduction

For randomised controlled trials (RCTs), current gold-standard methods of event ascertainment and verification are adjudicated clinic-based follow-up, but these processes are highly resource-intensive and complex, limiting the generation of new randomised evidence.[22] Instead, following-up participants via routinely collected health data sources (i.e. routine data) could substantially streamline trial conduct, as it could potentially reduce both the need to regularly interview participants at local research clinics and the manual collection of medical documentation.[2] But the trade-off when using these relatively novel methods to ascertain events is the potential introduction of errors, including missing true events and adding false events.

Thus far, the findings from studies presented in this thesis have been based on real data without knowledge of the ground truth, but suggest that there is considerable protection afforded by large-scale randomisation within appropriately blinded RCTs.[63,122,123] These studies using ASCEND[8,9] and SHARP[11] are important case-studies that highlight the potential for cardiovascular trials to utilise routine data follow-up. Limitations of these case-studies are that only few scenarios have been investigated and actual error rates are unknown. Simulation studies allow for numerous scenarios with varying levels of errors and could use assumptions derived from these case studies to add to the understanding of the potential impact of errors on randomised trial results.

Previous simulation studies have shown that only when a large proportion of trial participants (i.e. >50%) have a random error (i.e. where errors are equally distributed between trial arms) is there any meaningful change in relative treatment effect estimates (**Supplemental Table 6.1**).[124-126] Yet the impact of individual error components (i.e. true events removed and false events added) remains little understood. This is of particular relevance to trials using routine data follow-up, where investigators often have to balance accuracy and completeness. High accuracy may minimise the proportion of false events, but excludes relevant and important events by overly strict definitions. Aiming for high

completeness minimises the proportion of true events that are excluded, but may increase the chance of accepting a false event as an outcome. By way of illustration, in **CHAPTER 4,** the choice of routine data outcome definition for major bleeding showed to be important in whether the important bleeding hazard from aspirin was identified from routine data.[122] If the broadest bleeding definition that maximised completeness had been selected (i.e. any bleeding code recorded in the routine data was considered major), then the bleeding hazard was not identified. This is presumably because the large number of false events which were not influenced by study treatment had been included in the outcome, and this biased the relative treatment effect estimate towards the null. In contrast, a routine data algorithm that was developed to improve accuracy while not substantially compromising completeness, identified the bleeding hazard and provided estimated treatment effects similar to adjudicated direct follow-up.

This simulation study therefore aims to investigate the impact of removing true events and adding false events on relative treatment effect estimates and power in randomised trials. With the secondary aims of assessing: (1) whether such findings are consistent across trials with different risk ratios and samples sizes; and (2) the impact of adding false events that are not allocated equally between trial arms.

6.2 Methods

6.2.1 Data generation and scenarios

Multiple parallel-group randomised trials (i.e. active intervention vs placebo) with a binary outcome (i.e. by the end of the trial a participant either had or did not have an event) were generated based on two parameters: the relative effect of allocation to the active intervention arm estimated by a risk ratio (RR; calculated by dividing the proportions of participants with an event in the treatment arm by the event rate in the placebo arm) and overall trial sample size (n; the total number of trial participants). The number of events in the placebo arm were pre-determined using the assumption each trial would have 90% power at a 5% significant level to detect the selected RR (**Supplemental Table 6.2**). These

data were considered the ground-truth and assumed to be equivalent to “best available data”.

To compare the results to previous chapters, the parameters for the generated randomised trials were initially set to match those of ASCEND’s (n=15480) aspirin comparisons for any serious vascular event (SVE; RR=0.88; see **CHAPTER 3**) and major bleeding (RR=1.29; see **CHAPTER 4**). Scenarios using smaller (n=1000 and 5000) and even larger (n=25000) sample sizes were then considered with different treatment effects, these included ground truths based on different degrees of relative benefit (RR 0.70, 0.80, 0.90) and harm (RR 1.10, 1.20, 1.30).

6.2.2 Simulated errors

To simulate errors that might occur when using routine data follow-up, true events were removed and/or false events added in a range of different scenarios. Each participant had an equal probability of having an error (i.e. errors were random rather than systematic). The following two parameters were used: the proportion of true events removed (TER; calculated by dividing the number of true events removed by the total number of true event before any were removed) and proportion of false events added (FEA; calculated by dividing the number of false events added divided by the total number of true events before any were removed). FEA was used in preference to specificity as it is independent of the event rate, thus comparable across trial scenarios (**Supplemental Table 6.3** illustrates how specificity varies depending on the trial event rate). The results from the ASCEND and SHARP analyses (in **CHAPTER 3**, **CHAPTER 4** and **CHAPTER 5**) were used to help inform which error rates were plausible (**Supplemental Table 6.4** and **Supplemental Table 6.5**), thus this simulation study considered errors (i.e. TER and FEA values) between 0% and 40%.

Secondary analyses investigated scenarios where false events were not equally distributed between trial arms (i.e. were impacted by treatment), in which the effect size of the false

events could be half or double the true relative risk reduction (RRR). For example, if the true risk ratio was set at 0.80 (i.e. a 20% relative risk reduction) and the false events were distributed assuming an effect size of half the true RRR, then the false events RRR would be 10%. A complete list of the parameters used in this simulation study can be found in **Supplemental Table 6.6**.

6.2.3 Performance measures

In this simulation study it was assumed that the "best available data" were equivalent to the ground-truth, where no errors were present. For routine data follow-up, it was assumed that some level of error had been introduced (i.e. the proportion of true events removed and/or false events added were greater than 0%). Similar to previous chapters, routine data performance was measured by comparing it to the "best available data", the closer the routine data estimates were the stronger its performance. The primary performance measure was the mean difference between the "best available data" and routine data risk ratios, with 95% confidence intervals calculated using the 2.5^{th} and 97.5^{th} percentiles. The secondary performance measure was the differences in power (i.e. the proportion of simulated trials that correctly identify the presence of the treatment) between the "best available data" versus simulated routine data with errors. For each scenario, 5000 trials were simulated, as preliminary data indicated this to be more than adequate to yield a precise estimate of the primary performance measure (**Supplemental Table 6.8**). All data were generated and analysed using R version 4.2.2.

6.3 Results

6.3.1 ASCEND trial scenarios

For the ASCEND serious vascular event (n=15480, RR=0.88) and major bleeding event (n=15480, RR=1.29) scenarios, the non-differential removal of true events (i.e. the same proportion of events removed in both arms) had no impact on the estimated RR, while the non-differential addition of false events modestly biased the treatment effect towards the null (SVE: mean difference in RR: +0.01 [FEA=10%] to +0.04 [FEA=40%]; major bleeding:

-0.03 to -0.10; **Figure 6.1**; **Figure 6.2**; **Supplemental Tables 6.9-6.12**). When both false events were added and true events removed, the proportion of all trial events that were false marginally increased, thus further biasing the routine data RR towards the null (SVE: +0.01 [FEA & TEA=10%] to +0.05 [FEA & TEA=40%]; major bleeding: -0.03 to -0.13).

Both error types reduced trial power, with true events removed (SVE: difference in power: -3.5% [TER=10%] to -21.4% [TER=40%]; major bleeding: -3.9% to -21.0%) having a greater impact than false events added (SVE: -3.3% [FEA=10%] to -14.9% [FEA=40%]; major bleeding: -3.2% to -12.5%). Notably, the effects on power more than doubled when both error types were present (SVE: -7.5% [FEA & TEA=10%] to -48.9% [FEA & TEA=40%]; major bleeding: -6.8% to -42.4%).

In secondary analyses, when the false events were impacted by treatment (i.e. were not equally distributed between trials arms), although the overall proportion of false events added was not changed, the weak signal (i.e. false events RRR = half the true RRR) meant that more false events were allocated to the placebo rather than the active arm (**Supplemental Tables 6.9-6.12** and **Supplemental Figure 6.1-6.4**). Therefore, adding these false events with a weak signal had less of an impact on the mean risk ratio than when the false events RRR was null (SVE: mean difference in RR: +0.01 [FEA=10%] to +0.02 [FEA=40%]; major bleeding: -0.02 to -0.05). In contrast, when the false events RRR was set at double the true RRR, the addition of false events led to the routine data slightly overestimating the relative treatment effect.

6.3.2 Other randomised trial scenarios

For the other randomised trial scenarios (i.e. RR=0.70, 0.80, 0.90, 1.10, 1.20, 1.30; N=1000, 5000, 25000), removal of true events had no impact on the mean RR, while the addition of false events impacted the RR by biasing the estimate towards the null (**Figure 6.3; Supplemental Table 6.13**). When both false events were added and true events removed, this further biased the routine data RR towards the null. For a given treatment

effect size, trials with a smaller sample size (n=1000) were greater impacted (i.e. the RR point-estimate was closer to the null) by random errors than larger trials (n≥5000), with these differences between small and large trial scenarios increasing with high error rates. Similarly, for a given sample size, when random errors were introduced, trials with a larger treatment effect (i.e. where the true RR deviates furthest from the null) showed greater differences in the RR point-estimate between routine data and best available data, but the proportional difference in the point-estimate compared to true RR remained the same. For example, in the scenario where n=25000, FEA=40% and TER=40%, the difference in point-estimate for a true RR=0.90 was +0.04 (percentage change +40%), +0.08 (+40%) for RR=0.80, and +0.11 (+37%) for RR=0.70.

Considering the impact of random errors on power, smaller trials (n=1000) with more moderate treatment effects (i.e. where the true RR is closest to the null) showed greater losses in power than larger trials (n≥5000) with stronger treatment effects (**Figure 6.4; Supplemental Table 6.14**). In secondary analyses, when the false events were impacted by treatment (i.e. false events RRR = half or double the true RRR), the impact of errors on RRs and power were more modest (**Supplemental Tables 6.15-6.18**; **Supplemental Figures 6.5-6.8**).

6.3.3 Small-to-moderate random errors

For simulations where only small-to-moderate random errors were introduced (defined as FEA≤20% and TER≤20%), there appeared to be little impact on relative treatment effect estimates both in the ASCEND (SVE: +0.3% [FEA & TEA=20%]; major bleeding: -0.6%) and other trial scenarios, regardless of the sample size and treatment effect. But power was somewhat impacted (SVE: -20.5% [FEA & TEA=20%]; major bleeding: -16.9%), with greater reductions in power for smaller (n=1000) rather than larger (n≥5000) trials.

6.4 Discussion

Analysing simulated trial data from a wide range of trial scenarios showed that routine data sources with small-to-moderate random errors (i.e. true event removed and false events added ≤20%, which are equally distributed between trial arms) have little impact on relative treatment effect estimates in randomised trials. In the situation where routine data sources do introduce large amounts of error into a trial (i.e. true event removed and false events added >20%), larger trials (n≥5000) are better protected against random errors biasing the treatment effect towards the null and reducing trial power. By error type, true events removed did not impact the risk ratio but substantially reduced trial power, while false events added biased the risk ratio towards the null and modestly reduced power. When both false events were added and true events removed, the proportion of all trial events that were false marginally increased, further biasing the risk ratio towards the null.

These analyses complement the findings from previous chapters, and provide further rationale to consider using UK routine data as the sole source of follow-up for future cardiovascular trials. Similar to the findings in **CHAPTER 3** for serious vascular events in ASCEND (where both participants and investigators were blinded to treatment allocation), for simulations where only small-to-moderate random errors were introduced (i.e. false events added and true events removed ≤20%), there appeared to be little impact on relative treatment effect estimates, regardless of the trial's treatment effect and sample size. For simulations with high levels of random error (i.e. false events added and true events removed =40%), approximately the values observed in **CHAPTER 4** for major bleeding events, there was a far greater bias towards the null in the simulated data than was seen in the real ASCEND data. This may indicate that, as was hypothesised in **CHAPTER 4**, many of the supposed "false events" identified in the routine data may have been "true" major bleeds not captured by adjudicated direct follow-up. Thus highlighting the challenges of assessing the reliability of routine data follow-up using real trial data, where the best available data are often not equivalent to the ground truth. In this situation, it may be most

informative to compare relative treatment effect estimates derived using either data source because, as has been illustrated in this simulation study, the non-differential removal of true events should not bias the point-estimate.

This is the first study to use simulation methods to investigate how removing true events and adding false events impacts relative treatment effects and power in the context of trials utilising routine data sources for trial follow-up. Previous simulation studies focused on the impact of adjudication, in particular scenarios where investigators were not blinded to treatment allocation and when errors rates varied by trial-site (**Supplemental Table 6.1**).[124-126] These studies used the overall proportion of trial participants being misclassified as the error parameter, but this was not appropriate for routine data sources where outcome definitions can have different levels of accuracy (i.e. false events added) and completeness (i.e. true events removed). Understanding the mechanisms of these two error types and their relationship together was essential in order to quantify the true impact of routine data utilisation on a trial's treatment effect estimate and power. Additionally, the study that did use sensitivity/specificity as its main parameter, set specificity at implausible levels (i.e. a specificity ≤0.80 meant the trial included more false events than true events), thus limiting its relevance for most trials.[126] Despite these differences in study design, this study confirms the hypothesis set out in these previous studies that when errors are random (i.e. equally distributed between trial arms), large amounts of error are required before there is any meaningful effect on relative treatment effect estimates. The protection afforded by randomisation therefore allows appropriately blinded trials to consider the use of streamlined follow-up methods with the potential for less loss to follow-up, without compromising the precision of the relative treatment effect estimate.[22]

Although moderate random errors were shown to have little impact on relative effects, there were noticeable reductions in power observed in all trial scenarios, in particular for those smaller trials (n=1000) with more moderate treatment effects. Therefore when designing trials that utilise streamlined follow-up methods, trialists must consider approaches to

addressing these potential loses in power. Such methods may involve increasing the trial's sample size or extending the follow-up period. By way of illustration, for serious vascular events in the ASCEND trial (see **CHAPTER 3**), routine data were found to be poor at capturing transient ischaemic attacks, thus routine data event rates were marginally lower than for adjudicated direct follow-up. In the hypothetical scenario that ASCEND has used routine data follow-up only, approximately 3500 (23%) more participants were needed to have been recruited to have maintained the trial's power to detect the relative effects of aspirin on serious vascular events. But, the efficiency of using streamlined follow-up methods for large-scale trials arguably outweighs this disadvantage. Investigators and trial steering committees can pay particular attention to early event rates in trials reliant on routine data for their follow-up to ensure reassessment of design assumptions and recommend sample size/follow-up duration modification, where necessary.[127] Directions for future work include the development of a tool to adjust power calculations to take into account the potential errors introduced by utilising routine data as a means of follow-up.

6.4.1 Strengths and limitations

The strength of this simulation study is that it goes beyond the limited examples in the earlier chapters, allowing for a far more complete picture of the impact of errors on relative treatment effect estimates and power in a wide range of trial scenarios, making the results highly generalizable. Furthermore, by using simulation methods this study was able to compare routine data to the ground-truth, which is often not the case in real trial data where two imperfect sources are compared. Finally, the results from the earlier chapters enabled this study to set parameters at levels that were plausible for cardiovascular trials utilising routine data sources as a means of follow-up, thus making the results highly relevant to trials currently considering the use of such streamlined methods.

There were also a number of limitations. Firstly, this study only considered binary outcomes (i.e. the presence or absence of an event) and did not focus on the timing of the event because previous research has indicated that, for appropriately blinded trials, even very

large event date measurement errors have little impact on a trial's estimated relative treatment effect (**Supplemental Table 6.1**).[128] Re-analysing the ASCEND trial with a binary outcome (i.e. yielding a risk ratio and chi-squared statistic) provided almost identical estimates of the relative effects of aspirin compared to log-rank time-to-event analyses (**Supplemental Table 6.7**). Secondly, despite generating 5000 trials per scenario (far exceeding the minimum number required), the confidence intervals around the mean difference in risk ratios remained wide. Thus this study cannot discount that the observed differences between trial scenarios were due to chance. Thirdly, this simulation focussed only on scenarios where participants and investigators were blinded to treatment allocation, thus assuming errors to be equally distributed between trial arms. But for some trials blinding is not feasible (e.g. where participants are randomly allocated to undergo a surgical procedure or not). For such trials, further simulations where the distribution of errors between trial arms may not be equal would be scientifically valuable. Such analyses would quantify the major systematic bias that can be introduced by unblinded outcome ascertainment.[125]

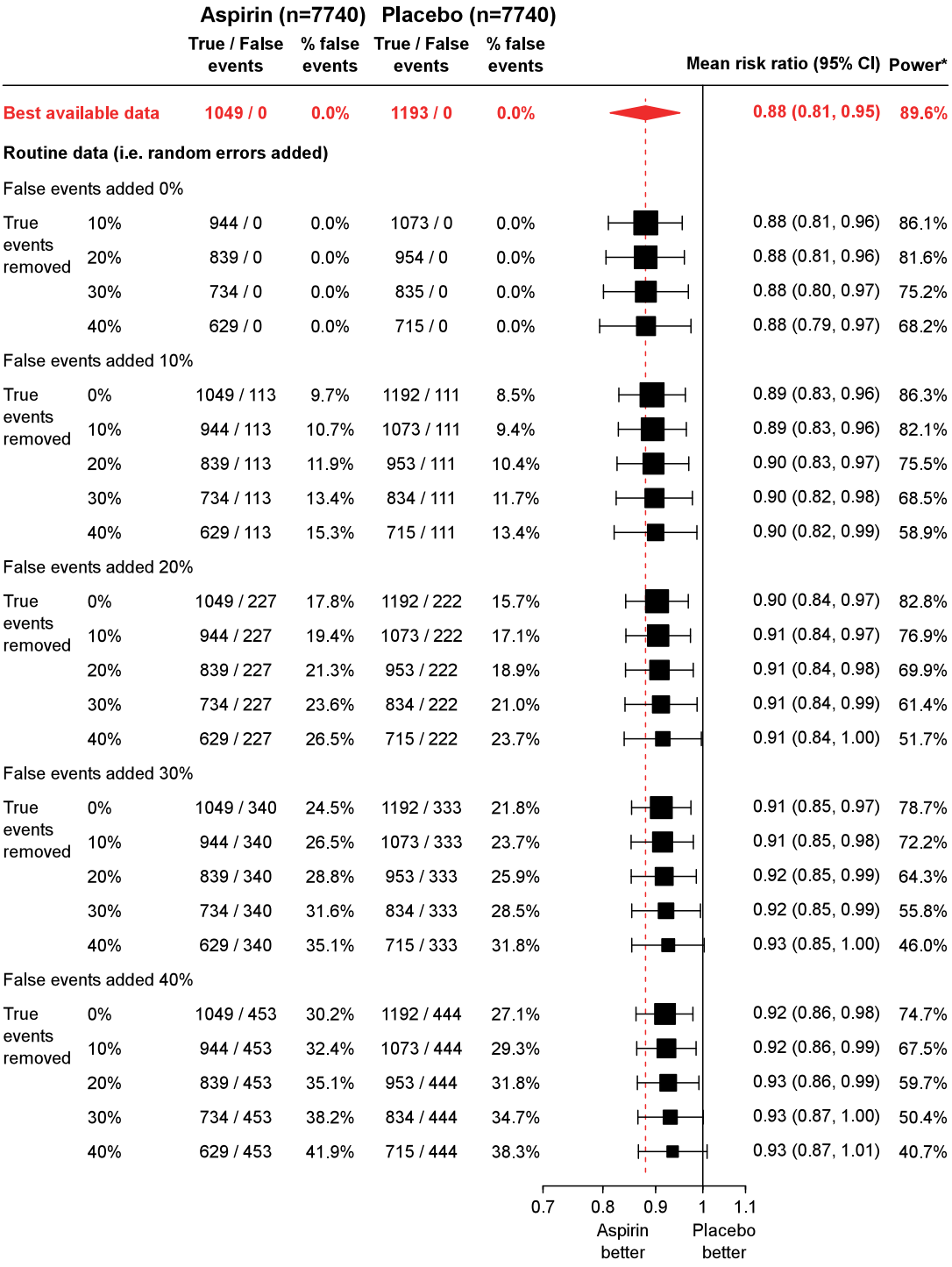
6.5 Summary

Analysing simulated trial data from a wide range of trial scenarios showed that routine data sources with small-to-moderate random errors (i.e. true event removed and false events added ≤20%, which are equally distributed between trial arms) have little impact on relative treatment effect estimates in randomised trials. In the situation where routine data sources do introduce large amounts of error into a trial (i.e. true event removed and false events added >20%), larger trials (n≥5000) are better protected against random errors biasing the treatment effect towards the null and reducing trial power. These results provide further rationale to consider using UK routine data as the sole source of follow-up for future cardiovascular trials.

Chapter 6 Tables and figures

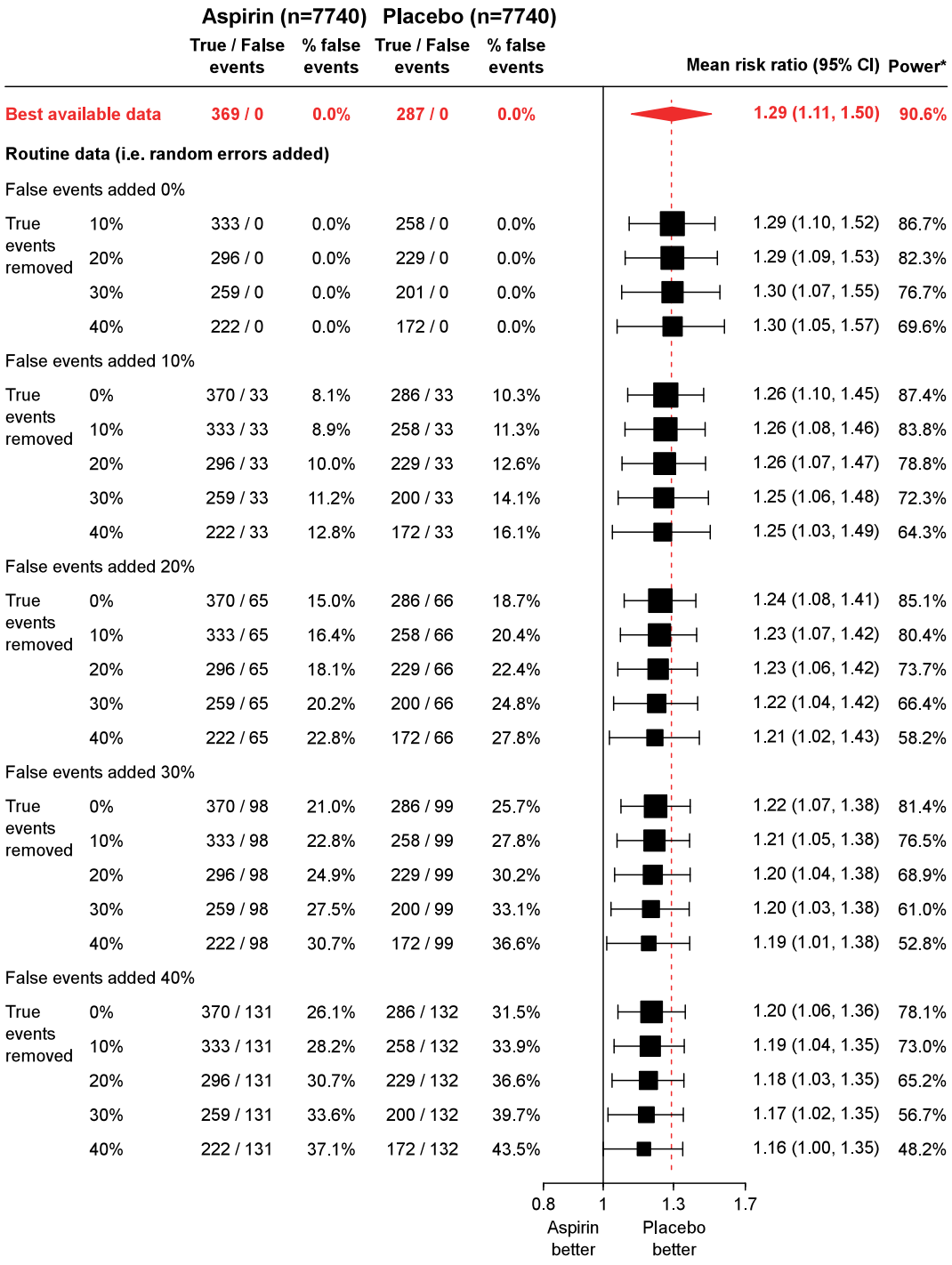
Tables/Figures	Title
Figure 6.1	Simulated ASCEND serious vascular events data measuring the impact of removing true events and adding false events on risk ratio estimates and power
Figure 6.2	Simulated ASCEND major bleeding events data measuring the impact of removing true events and adding false events on risk ratio estimates and power
Figure 6.3	Simulated trial data measuring the impact of removing true events and adding false events on risk ratio estimates, by trial sample size and relative treatment effect
Figure 6.4	Simulated trial data measuring the impact of removing true events and adding false events on power, by trial sample size and relative treatment effect

Figure 6.1 Simulated ASCEND serious vascular events data measuring the impact of removing true events and adding false events on risk ratio estimates and power



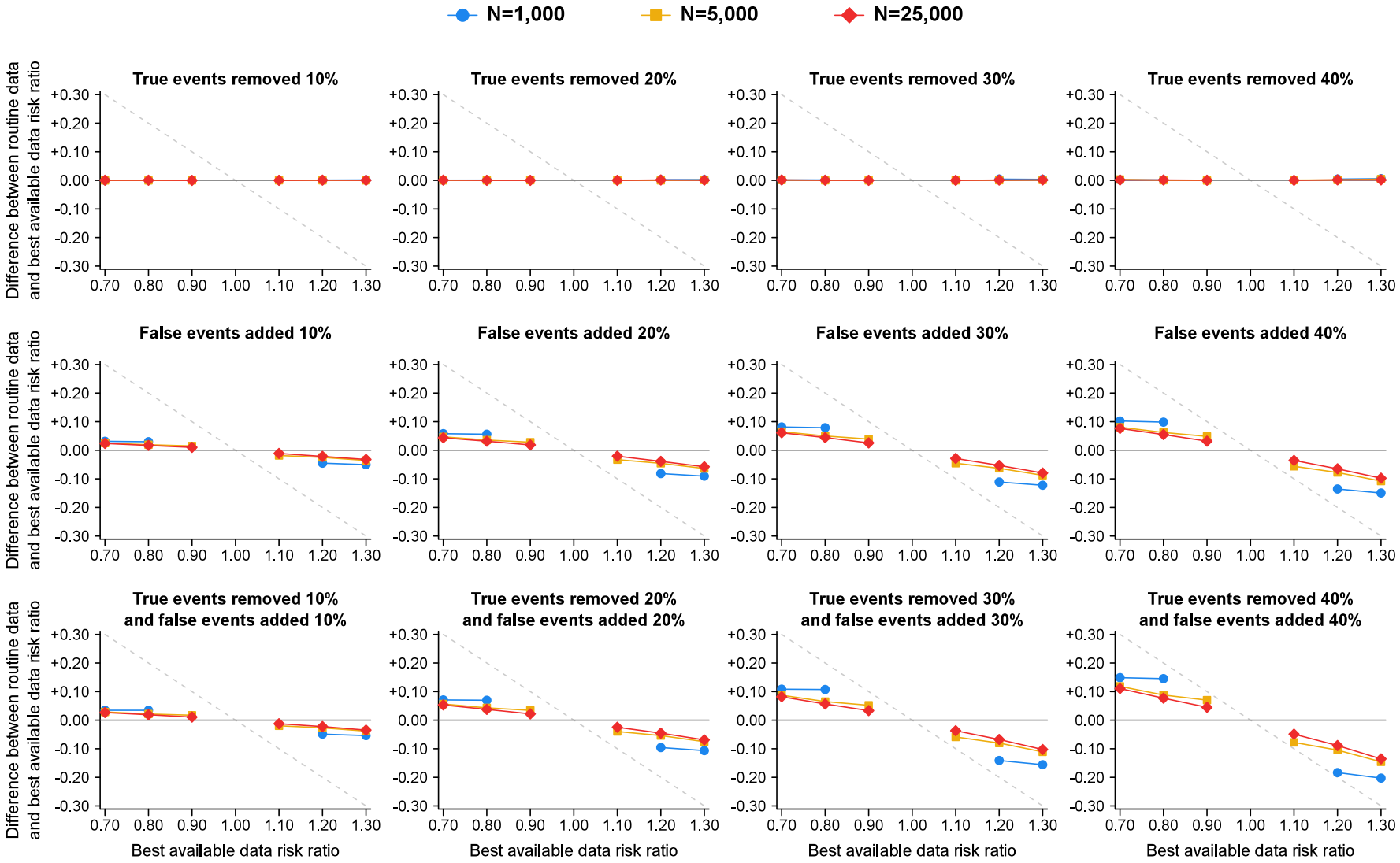
Mean risk ratio (2.5th and 97.5th percentiles) where point size is proportional to trial power. Simulated event rates are larger than in the published ASCEND trial because the trial ended with less than 90% power. Dashed red line represents adjudicated data risk ratio. % false events calculated by dividing the number of false events by the total number of events, by treatment arm. *Significance level set at 5%. CI = Confidence interval.

Figure 6.2 Simulated ASCEND major bleeding events data measuring the impact of removing true events and adding false events on risk ratio estimates and power



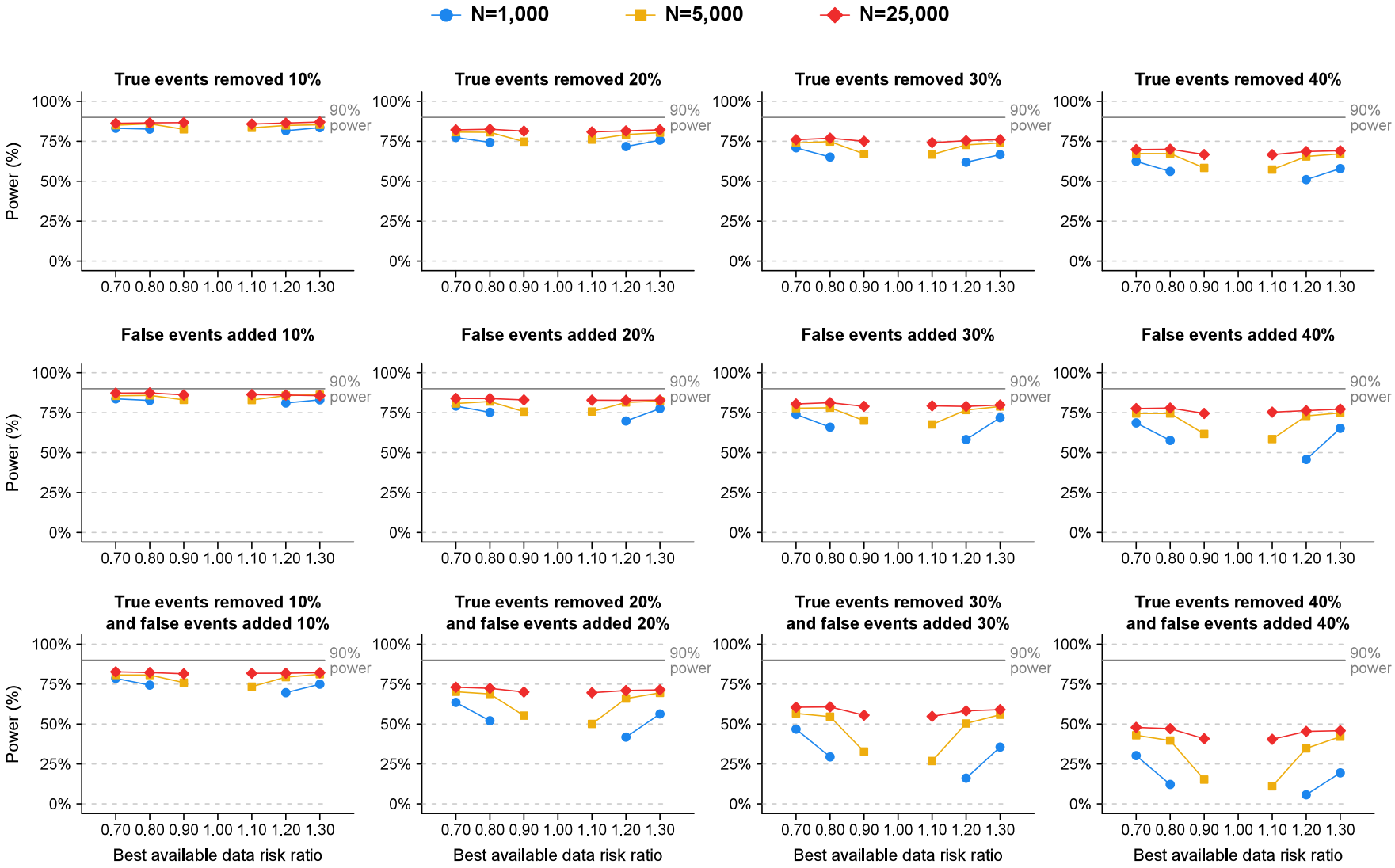
Mean risk ratio (2.5th and 97.5th percentiles) where point size is proportional to trial power. Simulated event rates are larger than in the published ASCEND trial because the trial ended with less than 90% power. Dashed red line represents best available data risk ratio. % false events calculated by dividing the number of false events by the total number of events, by treatment arm. *Significance level set at 5%. CI = Confidence interval.

Figure 6.3 Simulated trial data measuring the impact of removing true events and adding false events on risk ratio estimates, by trial sample size and relative treatment effect



Mean difference in risk ratio between routine data and best available data. Where no difference is shown by the solid grey line and routine data risk ratio = 1.00 (i.e. no effect) by the dashed grey line.

Figure 6.4 Simulated trial data measuring the impact of removing true events and adding false events on power, by trial sample size and relative treatment effect



Power (i.e. the proportion of simulated trials that correctly identify the presence of a treatment effect, with a significance level set at 5%) presented for routine data, where the best available data power is shown by the solid grey line.

CHAPTER 7 Overall conclusions

Chapter summary

Findings from this thesis suggest that UK routinely collected health data may be reliable enough to be the sole method of follow-up for major cardiovascular and bleeding events in primary prevention cardiovascular RCTs, allowing future trials to have a far more streamlined design with efficient low-cost participant follow-up. More generally, the findings from this thesis have consistently shown that when sufficient blinding is in place, the protection afforded by large-scale randomisation means that moderate errors introduced by utilising streamlined follow-up methods (i.e. routine data or pre-adjudicated direct follow-up) have little impact on relative treatment effect estimates. The results from this thesis have been published in a number of open-access journals, but future work is needed to investigate a wider range of outcomes, trial populations, and routine data sources. Other work by the wider UK research and health data community is also required to support implementation of these findings. These include initiatives to reduce the time and complexity of required to establish data access from providers and to reduce the time lag between hospitalisation and availability of information for researchers.

7.1 Introduction

Large randomised trials (RCTs) in cardiovascular disease (CVD) have provided reliable evidence for interventions whose widespread use has contributed to the secular declines in CVD mortality.[17,82] However, increasing costs of conducting trials threaten our ability to generate new randomised evidence.[19,22] Observational analyses in routinely collected healthcare (i.e. "real-world") data are increasingly used to make inferences about the safety and efficacy of interventions, but such non-randomised analyses cannot fully account for confounding and are unreliable.[2,3] More appropriately, routinely collected data could be used to streamline trial follow-up conduct/design by offering a cost-efficient method to ascertain outcomes.[83,84] But there remains considerable uncertainty about whether information derived from routine data follow-up are reliable enough to be used in randomised trials.[24]

The primary aim of this thesis was to investigate, whether for large cardiovascular trials in the UK, follow-up of their primary and secondary outcomes via routinely collected hospital admission and death registry data: (1) are ascertained as accurately and completely as adjudicated direct follow-up, and (2) can yield the same estimated treatment effects as adjudicated direct follow-up. With secondary aims to: (3) assess whether clinical adjudication of direct trial follow-up is necessary; and (4) use simulated data to investigate the impact on randomised trials more generally of removing true events and adding false events, as may happen if routine data are used for follow-up.

7.2 Main findings

7.2.1 Previous studies

The systematic search identified only three previous studies that compared cardiovascular and bleeding outcomes ascertained via UK routine data with adjudicated direct follow-up, two of which[62,65] assessed old routine data collected over two decades ago (i.e. 1989-2001), and the other[67] followed-up participants for only 28 days. Nevertheless, for cardiovascular

outcomes, agreement between data sources was strong for revascularisations, and moderate for MIs and strokes, with better concordance in people with no prior vascular disease. Re-running randomised comparisons using routine data only provided relative treatment effect estimates similar to adjudicated direct follow-up. For major bleeding events, routine data showed moderate completeness but accuracy was low, as the study[67] made no differentiation between major and minor bleeding events recorded in the routine data. No studies were identified that assessed kidney disease outcomes.

The systematic review concluded that further research was needed in order to better inform the design of future cardiovascular trials, in particular to:

1. Investigate the reliability of cardiovascular outcomes ascertained using "modern" up-to-date routine data sources, across UK nations.
2. Develop a routine data algorithm to accurately categorise bleeding events as major or minor.
3. Assess the reliability of routine data ascertained cardiovascular outcomes in populations where presentation of such events are atypical, for example in people with moderate-to-severe chronic kidney disease.
4. Investigate whether the kidney replacement therapy outcome can be reliability ascertained from UK routine data.

7.2.2 Serious vascular events and major bleeding events in the ASCEND trial

In analyses of the ASCEND[7-9] primary prevention trial (conducted between 2005-2017) in people with diabetes, cardiovascular and major bleeding outcomes ascertained via "modern" UK routinely collected hospital admission and death registry data were compared to adjudicated direct follow-up. For serious vascular events (SVEs, a composite of non-fatal myocardial infarction, presumed ischaemic stroke or TIA, and vascular death [excluding intracranial haemorrhage]) and arterial revascularisations, UK routinely data were found to provide similar estimated relative treatment effects to adjudicated direct follow-up for both

the aspirin and omega-3 fatty acids randomised comparisons. Agreement between adjudicated direct follow-up and routine data was strong for most of the components of the SVE outcome, with sensitivity poorest for ischaemic strokes and TIAs, and substantially improved when adjudicated follow-up were restricted to fatal and hospitalised events only. Routine data did appear to identify some additional events potentially missed by direct participant follow-up. Thus UK routine data may be reliable to be the sole methods of follow-up for myocardial infarctions, hospitalised ischaemic strokes, arterial revascularisations, and vascular deaths in streamlined primary prevention trials, perhaps without the need for verification by clinician adjudicators.

For major bleeding events (a composite of any intracranial haemorrhage, sight-threatening bleeding in eye, serious gastrointestinal bleeding, and other major bleeding events) in ASCEND, compared to adjudicated direct follow-up, by applying an algorithm to UK routinely collected hospitalisation and death registry data, routine data follow-up provided similar estimates of the relative and absolute effects of aspirin versus placebo. Between the two data sources this study found moderate levels of agreement, with both routine data and adjudicated direct follow-up missing some major bleeding events. Agreement was poorest for eye bleeding and other major bleeding, where routine data failed to identify a number of important events. Analysis by bleeding type found that for intracranial haemorrhage, sight-threatening bleeding in eye, and serious gastrointestinal bleeding the rate ratios were almost identical; however, for other major bleeding (which included epistaxis, haemoptysis, haematuria, vaginal, and other bleeding) routine data provided a lower estimate of the relative effects of aspirin compared to adjudicated direct follow-up. Despite these differences, the interpretation of ASCEND's aspirin randomised comparison (i.e. the benefits of aspirin being largely counterbalanced by the bleeding hazard) would have not changed had follow-up been solely via routinely collected data. These results suggest UK routine data represent a potentially cost-efficient method to ascertain major bleeding outcomes during both within and post-trial periods, either to supplement existing methods

or to be used on its own. But, further work is needed to validate this algorithm in other trial populations and evaluate further sources of routine data that can better ascertain eye bleeding and other major bleeding events. Additionally, safety data needs to be sent to the trial in a timely manner to satisfy regulatory requirements, currently such lags in reporting risk missing early clinically important safety signals.

The ASCEND trial's primary methods of follow-up was via mailed questionnaires every 6 months, as the majority of the primary and secondary outcome events were adjudicated, this allowed a direct assessment of whether the trial's results would have been altered in the hypothetical scenario that adjudication had not been carried out. For serious vascular events, relative and absolute treatment effects remained similar for pre-adjudicated follow-up. While for major bleeding events, the relative effects of aspirin on pre-adjudicated outcomes were similar, but the absolute effects of aspirin were found to be materially different, with pre-adjudicated follow-up providing an absolute bleeding hazard almost double that of the adjudicated data. ASCEND has illustrated the feasibility of streamlined mail-based follow-up methods, with reported SVEs potentially not requiring verification via clinical adjudication. This is particular relevant to trials being conducted outside of the UK where routine data may not be available.

7.2.3 Major atherosclerotic events and kidney replacement therapy in the SHARP trial

Thus far, routine data has shown to be a reliable method of follow-up in people who present with typical symptoms for atherosclerotic cardiac events, but it remains unclear whether such data are also a reliable means of follow-up in populations with atypical presentation, such as those with chronic kidney disease (CKD).[59] Analyses from the SHARP trial (2003-2010), in people with moderate-to-severe CKD, found that information on maintenance kidney replacement therapy (i.e. maintenance dialysis or kidney transplantation) was highly accurate and complete both in England routine data and reported by participants to trained research coordinators (i.e. pre-adjudicated direct follow-up), with pre-adjudicated data providing almost identical relative treatment effects estimates to adjudicated direct follow-

up. For major atherosclerotic events (MAEs), while both routine data and pre-adjudicated follow-up successfully identified the majority of events, these data sources appeared to identify additional coronary events (i.e. non-fatal MIs and coronary deaths) that adjudicators considered to be non-atherosclerotic, particular for people on dialysis; and routine data were not able to be able to differentiate between non-coronary revascularisations related to dialysis access vs not. Despite these discrepancies, relative treatment effect estimates from pre-adjudicated data were very similar to adjudicated direct follow-up, with benefits seen both in people on and not on dialysis at time of randomisation.

Thus in people with CKD, routine data in England may be reliable enough to ascertain initiation of maintenance dialysis, kidney transplantation, non-fatal myocardial infarctions, coronary deaths, non-haemorrhagic strokes, and coronary revascularisations. However, for non-coronary revascularisations an additional adjudication step might be necessary to exclude dialysis access procedures. In settings where routine data information are not available, findings from SHARP indicate that adjudication of clinic-based follow-up for KRT and MAE outcomes may not be necessary, and over strict adjudication criteria for MAEs may also have led to the trial underestimating the relative and absolute treatment effects among people on dialysis.

7.2.4 Simulation study investigating the impact of routine data errors on treatment effect estimates and power in randomised trials

Analysing simulated trial data from a wide range of trial scenarios showed that routine data sources with small-to-moderate random errors (i.e. true event removed and false events added ≤20%, which are equally distributed between trial arms) have little impact on relative treatment effect estimates in randomised trials. In the situation where routine data sources do introduce large amounts of error into a trial (i.e. true event removed and false events added >20%), larger trials (n≥5000) are better protected against random errors biasing the treatment effect towards the null and reducing trial power. These results provide further

rationale to consider using UK routine data as the sole source of follow-up for future cardiovascular trials.

7.3 Strengths and limitations

The studies conducted in this thesis had a number of strengths. Firstly, using data from the ASCEND trial, which completed follow-up in 2017, meant that "modern" up-to-date routinely collected data were able to be assessed and therefore are highly relevant to trialists currently at the stage of designing streamlined trials in the UK. Secondly, there were enough events in ASCEND and SHARP to not only investigate its primary and secondary outcomes, but also to provide agreement statistics for the individual components. This was particularly important in ASCEND, where due to large numbers of events, it was possible to identify that the reduction in sensitivity for SVEs was driven by non-hospitalised ischaemic strokes and TIAs no being recorded in the current routine data sources. Thirdly, unlike many previous studies that were not able to compare routine data sources to an established gold-standard[96,129], this study was able to compare routine data with clinically adjudicated direct participant follow-up, therefore assessing these novel follow-up methods against current trial approaches. Fourthly, by re-running ASCEND's randomised comparisons, this study was able to consider the hypothetical scenario that routine data had been the sole method of follow-up and assess whether there were any statistically meaningful differences in the point-estimate, thus providing context for agreement statistics that can often have limited meaning on their own.[50,67] In ASCEND, the trial team made particular effort to establish routine data linkage not just in England, but also in the devolved nations, meaning it represents one of the few studies to consider reliability of routine data across multiple countries' data systems. Finally, with the utilisation of simulation methods, this thesis has been able to go beyond the case studies presented in this thesis, allowing for a more complete overview of the impact of routine data errors on treatment effect estimates and power. With results suggesting that small-to-moderate random errors (i.e. false events

added and true events removed ≤20%) have little impact on relative treatment effect estimates in appropriately-blinded large randomised trials.

The studies presented in this thesis also have a number of limitations, firstly, due to the systematic review only identified a small number of trials reporting experiences with routine data, and trials where experiences may have been different may not have been published. Secondly, the ASCEND trial (and the England subset of the SHARP trial) were carried out in the UK which has a national health service, the interpretation may not be generalisable to routine data in non-UK countries. Thirdly, ASCEND and much of SHARP were primary prevention populations limiting generalizability to secondary prevention trials, in which a distinction must be made between an incident serious vascular event and the recording of pre-randomisation diseases (although such work has been conducted in the HPS trial described in **CHAPTER 2**). Fourthly, the ASCEND database did not have the associated medical notes to verify or refute over two-hundred events of major bleeding identified solely from routine data sources. ASCEND also utilised mail-based follow-up, so comparisons with more traditional clinic-based reported events was not possible (although again, such work has been conducted in the HPS trial described in **CHAPTER 2**). ASCEND participants were not linked to their primary care records which may have improved the ascertainment of non-hospitalised events. Finally, in the simulation study, it was assumed that routine data sources can be compared to the ground-truth, however, despite clinical adjudication, perfect follow-up data is an aspiration and so any data comparisons needs to acknowledge there will always be imperfections in the real world.

7.4 Implications for future clinical trials

This thesis has highlighted the potential utility of routine data as a means of participant follow-up. For UK primary prevention cardiovascular trials, linkage to routinely collected hospital admission and death registry data may be a reliable method of streamlining trial follow-up for a variety of outcomes, including: non-fatal myocardial infarction, hospitalised ischaemic stroke, vascular death (including its subtypes), arterial revascularisation, kidney

replacement therapy, intracranial haemorrhage, and serious gastrointestinal bleeding. Furthermore, analyses of real and simulated trial data both suggest that in large trials, where both participants and investigators are blinded to treatment allocation, small-to-moderate errors in event ascertainment appear to have little impact on relative treatment effect estimates. Thus, when designing large cardiovascular trials in the UK, trialists should consider the merits of applying for linkage to routinely collected data, and its potential role as a means of follow-up information for trial participants. Although further studies to confirm the findings presented in this thesis would be scientifically valuable, the presented data provide a strong rationale for use of UK routinely collected data for follow-up of important major cardiovascular and bleeding outcomes, potentially without the need for clinical adjudication.

These relatively novel methods have important benefits both for participants and researchers, as fewer site visits may be required, and efficient participant follow-up directly allows trials to recruit large numbers of participants, thus being better powered to answer important questions that directly impact clinical practice. Indeed, a number of UK streamlined trials are currently being set up with a plan to utilise both routine data and mail-based follow-up methods, including: A Study of Cardiovascular Events iN Diabetes Plus (ASCEND-PLUS)[130], a double-blind placebo-controlled trial investigating whether semaglutide can safely help reduce major cardiovascular events in 20000 people with type 2 diabetes; Aspirin to Target Arterial Events in Chronic Kidney Disease (ATTACK)[131], an open-label trial (with blinded endpoint adjudication) investigating whether taking low dose aspirin, compared to usual care, can reduce the risk of major cardiovascular events in 25210 people with CKD; and Aspirin after hospitalisation with Pneumonia to prevent cardiovascular Events randomised Controlled Trial (ASPECT)[132], a multi-centre open-label parallel group trial investigating whether aspirin reduces the risk of major cardiovascular events in 22600 patients hospitalised with pneumonia. Without such streamlined methods available it is unlikely that such large trials would be feasible or affordable, particularly when

investigating “old” drugs.[97] As the primary outcome for each of these trials is a modified version of the major cardiovascular event outcome and for two trials the primary safety outcome is major bleeding, the published results from **CHAPTER 3, CHAPTER 4** and **CHAPTER 5** should help inform the design of these trials.[63,122,123]

In addition to new streamlined trials being set up, the findings from this thesis are currently being used to help inform the BHF Data Science Centre[133] in developing standardised routine data definitions for core cardiovascular trial outcomes. The BHF’s SCORE-CVD project is currently bringing together multiple stake-holders to agree on consistent outcome definitions that can be adopted by all future UK cardiovascular trials that plan to use routine data follow-up.

In order to provider a stronger evidence base in the understanding of the reliability of routine data follow-up for randomised trials many more Studies Within Trials (SWATs) ought to be conducted.[134] However, as highlighted in **CHAPTER 2**, most SWATs to-date are inconsistent in their analytical approach and presentation of results. Guidelines by an MRC-NIHR Trials Methodology Research Partnership (TMRP) Health Informatics Working Group[135] on the design of routine data utility SWATs are being developed and will include training based on data from this thesis for other trialists wishing to conduct such studies.

Beyond the use of routine data, the ASCEND trial has demonstrated the effective use of mail-based recruitment and follow-up to streamline trial conduct, where no physical sites were needed. In this thesis, post-hoc analysis of these data showed that the trial’s interpretation and main conclusions for the serious vascular event finding would not have been different if mail-based reports had not been adjudicated, thus offering a further step to streamlining trial procedures.

For people with CKD, current guidelines from the International Society of Nephrology’s 1st International Consensus Meeting on Defining Kidney Failure in Clinical Trials concluded that “the role of event adjudication for kidney failure needs to be evaluated within the

specific context of the study planned".[136] Findings from **CHAPTER 5** indicate that adjudication of initiation of dialysis and kidney transplantation may be wholly unnecessary for international trials recruiting patients from nephrology sites. Indeed, the preliminary findings from these analysis informed the design of the EMPA-KIDNEY trial, where KRT events were not adjudicated, and instead investigators reviewed a renal status timeline to record new onset of kidney replacement therapy (a key component of the trial's primary outcome).[137] Additionally, for major atherosclerotic events in people with CKD, SHARP highlighted a potential negative consequences of overly strict criteria for outcome definitions on a trial's power. These findings should prompt more reflection on scientific value of adjudication when designing trials, rather than assuming it is always the most prudent design option, where resources allow.

7.5 Current barriers to utilising routinely collected data

This thesis has demonstrated that routine data may play an important role in the design of future randomised trials. Nevertheless, there are challenges with respect to applications and data processing/analysis that may need to be solved before routine data can be successfully embedded into a trial.

7.5.1 Applying for and receiving linked routine data

At the current time, applying for routine data linkage is a highly time consuming and complex process that can take many months (and sometime years) to complete.[45] A recent commentary by the PATCH trial, described how from the time when the application was submitted, is took 15 months to receive the first extract of data from NHS Digital.[47] With data provider policies continually changing, current applications in development can quickly become out of date and frequently need updating.[138] Indeed, this was true when applying for routine data linkage to the SHARP England cohort, with sections in the application often needing to be amended prior to and after submission. Even across UK nations the application processes can be very heterogeneous and require months of additional staff

time to submit each application.[134] The “correct” wording on consent forms and the information governance procedures in place are essential for routine data linkage approval. Better guidance from data providers would be extremely helpful.[47,138] Finally, data providers require researchers to carefully select the minimum number of datasets and data fields required to conduct their study. Much of the language around such datasets are specific to NHS service-delivery. This could be a barrier for to researchers who do not fully understand what data are available and how to request specific data items.

Once the application is approved, the trial unit can start receiving extracts from the data provider. But, even for well-established datasets, such as hospital admission data, the data received tends to be at least 3 months out-of-date. Clinical trial regulators require the timely reporting (within 7 to 15 days) of any suspected unexpected serious adverse reactions (SUSAR) related to the treatment being investigated.[134] The delayed data flows therefore limit its utility for randomised trials, where other methods of follow-up are needed to identify any potentially SUSARs. Additionally, trials are also depending on the ongoing existence of the dataset and flow of data, this could be seen as a considerable risk to principle investigators.

Current collaborative efforts in the UK between researchers and data providers through the recently founded Health Data Research UK[92] (HDR-UK) are working at trying to reduce the barriers mentioned above.[26,28] In NHS England for example, NHS DigiTrials[109], a new data service for UK trials, are considering methods to simplify the application process, provide datasets that satisfy CDISC trial data standards[139], and improve the timeliness of data flows.[48]

7.5.2 Processing and analysing routine data

The Medicines and Healthcare produced Regulatory Agency in the UK and the US Food and Drug Administration require trials to show data provenance and integrity. This includes a detailed record of where the data was recorded, how it was processed, and how

consistent such data are through the data flow/lifecycle. This is particular challenging for trials aiming to utilise routine data, where much of the data lifecycle occurs prior to it being sent to the trials unit. Murray et al[48] have designed a preliminary framework to document the provenance and integrity of routine data, and have published reports for both hospital admission data and death registry data in England.

Routinely collected health data are highly complex and specific to the healthcare system they represent. Greater effort needs to be made to share expertise and examples of good practice between trial units. This thesis has highlight how understanding the healthcare system is an important factor when interpreting the data. For example in developing severity factors to determine whether a bleeding event was major, it was important to understand that hospital admission data includes both “hospital inpatient stays” where participants stay overnight, and “day cases” where patients may be in a hospital bed for a small number of hours.

7.6 Implications for future research

To-date, in the UK, few routine data utility SWATs have been conducted; indeed in the systematic review conducted in **CHAPTER 2** only eleven studies were identified that compared the reliability of UK routine data follow-up against trial follow-up data. Within these 11 studies the statistical methods to compare data sources were highly heterogeneous, leading to substantial difficulties comparing results between studies, and most had too few events to reliably assess agreement. Thus, in order to build the evidence base necessary for inform trial design, numerous more SWATS are needed, where the linked trials are of adequate size and results are presented in a standardised format. Additionally, it is important that these studies not only investigate new outcomes in different trial populations, but also carefully verify existing findings.

The relatively high costs of retrospectively linking completed trials to routine data sources are a considerable barrier for many clinical trial units in conducting research, particularly as

trial methodology budgets are small.[45-47] Routine data utility SWATs could instead be included as an add-on to existing routine data applications for other purposes. For example, the primary purpose of linkage for the ASCEND trial was for long-term participant follow-up, the routine data utility SWAT made use of the same data but was included as a secondary purpose. Alternatively, routine data could be assessed in the early phase of a trial, where a decision can be made as to whether adjudication and/or additional sources of follow-up information may be necessary. This approach was taken by the ADAPTABLE trial in the US, where these findings helped inform trial design, leading to the decision that clinical adjudication was not necessary.[107]

Most studies presented in this thesis investigated only a limited number of routine data sources including: routinely collected hospital admission data, cancer and death registry data. As shown in **Figure 1.3**, there are an increasing number of routine data sources becoming available to researchers in the UK. The RECOVERY[26] trial for example, made use of over 25 different routine datasets in which such data was utilised in all areas of the trial's conduct. Future research needs to consider how multiple routine data sources might improve the accuracy and completeness of event ascertainment (and how to resolve apparent conflicts between different data sources). This has already been achieved with two data sources in the Salford Lung Study which developed an algorithm to identify COPD exacerbation using a combination of routinely collected hospital, GP and prescribing data.[72]

For cardiovascular trials, findings from this thesis indicate that UK routine data may be reliable at identifying particular trial outcomes (i.e. major cardiovascular and bleeding events), however, only three previous studies were identified and only two trials were analysed in this thesis. Further research to replicate these findings in a variety of cardiovascular trials would be useful, particular using trials that have used more traditional follow-up methods (including investigator reports and clinic-based follow-up), where routine data can be compared to the current trial "gold-standard". Additionally, routine data utility needs to be assessed in a wider range of countries and broader trial populations, such as

in people with prior cardiovascular events (i.e. secondary prevention trials), where routine data needs to be able to differentiate between “old” and “new” events.

Finally, this thesis has focused primarily on the views of the researchers conducting the trials. Equally important, if not more, are maintaining the trust of the patients, general public and medical profession in the safe and transparent use of people’s routine data in randomised trials. Work is currently ongoing to bring these groups into current discussions on the use of routine data in trials.[50] In the past, when the public and medical professionals have not been involved in decision making processes, this has led to substantial loss of trust and reservations over allowing non-NHS organisations access to their data.[140] This poses a substantial risk to future trials aiming to rely on such data sources for particular areas of trial conduct, greater efforts needs to be made to improve public and patient involvement in running clinical trials and understanding in what purposes they feel comfortable for their data to be used. The RECOVERY[27] trial and #datasaveslives[141] campaign have made some headway in this regard but much more is needed to be done.

7.7 Final summary

Findings from this thesis suggest that UK routinely collected health data may be reliable enough to be the sole methods of follow-up for important outcomes in large-scale cardiovascular trials (i.e. major cardiovascular and bleeding events) and support their wider use in ongoing and future such trials. These relatively novel methods could provide important benefits both for patients more widely and could reduce trial participant burden. The methods should enable researchers to design larger trials for lower cost, ensuring both better powered to detect moderate treatment effects and more efficient use of limited funding resources. Some of the key results from this thesis have already been published and are available open-access to help inform both the streamlined design of future UK cardiovascular trials, to support national aims to develop standardised routine data definitions for cardiovascular trials, and to help broader regulatory acceptance of the methodology. Other work by the wider UK research and health data community remains to

support implementation of thesis findings into practice. These include continued financial support to data providers, greater patient and public involvement, initiatives to reduce the time and complexity of required to establish data access from providers, and to reduce the time lag between hospitalisation and availability of information for researchers.

Appendix 1 Statistics of agreement

A1.1 Background and assumptions

Over the past decade agreement statistics have become the established tool for which trialists have been able to measure the performance of routine data follow-up at accurately and completely ascertaining events in randomised trials.[62,67,89,107,142-145] In this appendix these statistical methods are outlined as an aid to the main thesis.

Assessing routine data follow-up performance using agreement statistics relies on the following assumptions:

- Adjudicated direct follow-up, the "gold-standard" method of event ascertainment in randomised trials, is an approximation of the ground-truth; and therefore the data source used to assess the performance of routine data follow-up.
- Routine data and adjudicated follow-up are independent sources of information (i.e. routine data sources were not used to identify events in adjudicated direct follow-up).
- For each participant, only the first occurrence of an event after the randomisation date is counted.
- Routine data follow-up performance is assessed by measuring:
    - Accuracy (i.e. how often routine data follow-up correctly categorises a participant as having had no event)
    - Completeness (i.e. the proportion of participants with a true event that are correctly identified by routine data follow-up)

When assessing the performance of routine data follow-up, two questions can be asked: (1) whether routine data can correctly categorise a participant as having an event or not; and (2) to what extend the date of event in routine data follow-up deviates from the ground-truth. This methods appendix focuses on question 1 as it is the primary question trialists aim to assess and which standardised methods of assessment exist.[63]

A1.2 Categorising participants

When measuring agreement, participants are categorised as to whether during the trial's follow-up period (commonly starting at randomisation date and ending at date of censoring or death) they have experienced the outcome of interest (i.e. Outcome=Yes, or Outcome=No). For each participant, this categorisation can be made for both events identified in adjudicated direct follow-up and routine data, the resulting categories can be seen in **Table A1.1**. As adjudicated direct follow-up data are assumed to approximate the ground truth, participants are defined as "true positives" where an event has been identified in routine data and adjudicated direct follow-up, while "false positives" are defined as no event being identified in adjudicated follow-up but an event identified in routine data only. Similarly participants who were neither identified as having an event in adjudicated or routine data follow-up are termed "true negatives", and where no event was identified in routine data but was identified in adjudicated follow-up, these are termed "false negatives". In this thesis it was decided to simplify this language by describing these categories as: "outcome in both datasets" for true positives; "outcome in routine data only" as false positives; "outcome in adjudicated follow-up alone" as false negatives; and "no such outcome in either dataset" as true negatives.

Table A1.1: Two-by-two table comparing binary outcomes

		Adjudicated direct follow-up	
		Outcome = Yes	Outcome = No
Routine data	Outcome = Yes	A = True positives	B = False positives
	Outcome = No	C = False negatives	D = True negatives

A1.3 Selection of agreement statistics

Agreement in **Table A2.1** can be summarised using one of the following statistics: sensitivity, specificity, positive predictive value (PPV), negative predictive value (NPV),

crude agreement, and kappa statistic[56]. These statistics are summarised below in **Table A1.2**, in which all six statistics have been used in some form by routine data validation studies previously.[62,67,89,107,142-145] PPV, NPV and crude agreement tend have limited generalisability across studies, due to being dependent on the event rate of the trial population being analysed. Therefore the primary statistics this thesis focuses on are sensitivity, specificity, and kappa as these have greater generalisability and are mostly independent of changes in the event rate. Overall agreement is measured using the kappa statistic, while sensitivity and specificity measure completeness and accuracy.

Table A1.2: Summary of statistics of agreement

Statistic	**Description**	**Formula**
Sensitivity	The proportion of participants with a true event correctly identified by routine data follow-up.	$Sensitivity = \frac{\mathrm{A}}{\mathrm{A}+\mathrm{C}}$
Specificity	The proportion of participants with no true event correctly identified by routine data follow-up.	$Specificity = \frac{\mathrm{D}}{\mathrm{B}+\mathrm{D}}$
Positive predictive value (PPV)	The proportion of participants with an event identified by routine data follow-up that are true positives.	$PPV = \frac{\mathrm{A}}{\mathrm{A}+\mathrm{B}}$
Negative predictive value (NPV)	The proportion of participants with no event identified by routine data follow-up that are true negatives.	$NPV = \frac{\mathrm{D}}{\mathrm{C}+\mathrm{D}}$
Crude agreement	The proportion of all participants that were correctly categorised by routine data follow-up as true positives or true negatives.	$Crude\ agreement\ (\mathrm{p_o}) = \frac{\mathrm{A}+\mathrm{D}}{\mathrm{N}}$
Kappa statistic	The overall level of agreement between routine data and adjudicated direct follow-up, adjusted for agreement by chance.	$\text{Expected agreement } (\mathrm{p_e}) = \left(\frac{\mathrm{A}+\mathrm{B}}{\mathrm{N}}\right)\left(\frac{\mathrm{A}+\mathrm{C}}{\mathrm{N}}\right) + \left(\frac{\mathrm{C}+\mathrm{D}}{\mathrm{N}}\right)\left(\frac{\mathrm{B}+\mathrm{D}}{\mathrm{N}}\right)$ $\text{Kappa statistic } (\kappa) = \frac{\mathrm{p_o}-\mathrm{p_e}}{1-\mathrm{p_e}}$

A1.4 Calculating 95% confidence intervals and comparing subgroups

In order to understand the level of uncertainty around sensitivity and specificity statistics, 95% confidence intervals can be calculated using the Wilson method of binomial proportions.[86] Where p denotes the sensitivity/specificity statistic (see equations in **Table A2.2**) and n the number of participants.

$$\text{Sensitivity/Specificity 95\% confidence interval} = \mathrm{p} \pm 1.96\sqrt{\frac{\mathrm{p}(1-\mathrm{p})}{\mathrm{n}}}$$

Similarly, for the kappa statistic Cohen developed methods to calculate the 95% confidence interval.[56] Where $\mathrm{p_o}$ and $\mathrm{p_e}$ denote the observed and expected levels of agreement (see equations in **Table A2.2**), and n signifies the total number of participants.

$$\text{Kappa 95\% confidence interval} = \kappa \pm 1.96\sqrt{\frac{\mathrm{p_o}(1-\mathrm{p_o})}{\mathrm{n}(1-\mathrm{p_e})^2}}$$

Depending on the aims of the study and the type of data being analysed, the interpretation of the kappa statistic can vary substantially. However, this thesis uses the established approach by Landis and Koch[57] where 0.01-0.20 is considered to represent slight agreement, 0.21-0.40 fair, 0.41-0.60 moderate, 0.61-0.80 strong, and 0.80-0.99 very strong agreement.

Finally, differences in the kappa statistic between subgroups can be assessed using heterogeneity testing,[56,87] in which one can derive the z statistic for a two-way heterogeneity test using the following equation. Where $\mathrm{k_1}$ and $\mathrm{k_2}$ denote the kappa statistics in groups 1 and 2, $\mathrm{SE(k_1)}$ and $\mathrm{SE(k_2)}$ denote the kappa statistic standard errors in groups 1 and 2.

$$z = \frac{\kappa_1 - \kappa_2}{\sqrt{[SE(\kappa_1)]^2 + [SE(\kappa_2)]^2}}$$

A1.5 Summary

To summarise, statistics of agreement are an important tool to measure how reliable routine data follow-up are at ascertaining outcomes in randomised trials. This thesis primarily focuses on sensitivity, specificity and kappa as these statistics are mostly independent of a population's event rate. The kappa measures overall agreement, while sensitivity and specificity measure the completeness and accuracy of routine data follow-up. For these three statistics there are established methods to calculate 95% confidence intervals and test for differences between subgroups.

Appendix 2 Systematic review protocol

A2.1 Review question

How accurate and complete are routinely collected health data in the United Kingdom compared to adjudicated direct follow-up for ascertaining outcomes in randomised trials?

A2.2 Search terms in titles, abstracts, or keywords

[1] Randomised controlled trials:

(clinical trial$ OR randomi#ed trial$ OR randomi#ed controlled trial$)

[2] Routinely collected data (specific terms):

(hospital episode statistics OR patient episode database for wales OR scottish morbidity record$ OR national register OR national registry OR national registries OR public register OR public registry OR public registries OR national patient register OR national patient registry OR national patient registries OR administrative claims OR routinely collected data OR routinely collected electronic patient record$)

[3] Routinely collected data (non-specific terms):

(administrative data OR administrative record$ OR claims data OR claims record$ OR electronic health care record$ OR electronic healthcare record$ OR electronic health data OR electronic health record$ OR electronic medical record$ OR electronic patient record$ OR general practice data OR general practice record$ OR hospital data OR hospital record$ OR hospitali#ation record$ OR inpatient data OR national data$ OR outpatient data OR primary care record$ OR primary care data OR routine data OR secondary care data OR secondary care record$ OR registries OR registry)

[4] Methods of record linkage:

(linked or linkage)

[5] Methods of comparison:

(accuracy OR agreement OR completeness OR concordance OR kappa OR negative predictive value OR positive predictive value OR npv OR ppv OR sensitivity OR specificity OR valid OR validation OR validity OR treatment effect$ OR intervention effect$ OR matched OR miscoding)

A2.3 Search strategy

For an article to be identified it must include a randomised trial term, routine data source term (either a specific dataset, or non-specific routine data term with reference to data linkage), and a term referencing a method of comparison between data sources.

[1] Search for randomised controlled trials

[2] Search for routinely collected data (specific terms)

[3] Search for routinely collected data (non-specific terms)

[4] Search for methods of record linkage

[5] Search for methods of comparison

[6] Combine ([1] AND ([2] OR ([3] AND [4])) AND [5])

A2.4 Databases

The MEDLINE and Embase databases will be used. In which searches will be restricted to English language only and a publication date of 1st January 2010 onwards; for Embase all conference abstracts will also be excluded.

MEDLINE:

1. Access Search Oxford Libraries Online (SOLO) via: http://solo.bodleian.ox.ac.uk/
2. Search for: MEDLINE OVID

3. Open the database: Ovid MEDLINE ® Epub Ahead of Print, In-process & Other Non-Indexed Citations, Ovid EMDLINE ® Daily and Ovid MEDLINE ® 1946 to present
4. Submit syntax: [1] to [6]
5. Restrict to: English language only
6. Publication date: 1st January 2010 to present

Embase:

1. Click on change database
2. Select: Embase (1974 to present)
3. Submit syntax: [1] to [6]
4. Restrict to: English language only
5. Publication date: 1st January 2010 to present
6. Remove conference abstracts

A2.5 Review criteria

An initial search will be conducted, in which to verify that all known articles have been included using the existing search strategy. If some are missing, the search strategy will be updated. Duplicates will be checked manually in Excel. All articles will then be reviewed, first based on the title only, and then by abstract, those that are not excluded will then have full text screening using the following criteria:

1. Only include studies that have conducted a comparison between randomised clinical trial data and routine data follow-up (title/abstract)
2. Exclude studies using non-UK routinely collected data (full text)

A2.6 Data extraction

Information will be extracted using the following categories:

- Trial design

- Participant characteristics
- Routine data linkage
- Algorithms used to define the outcomes
- Criteria for an event to match between the two data sources (e.g. <14 days difference)
- Study results:
    - Participant-level or event-level comparison between routine data and adjudicated follow-up
    - Agreement statistics (e.g. sensitivity and specificity)
    - Randomised comparisons re-run using routine data only
- Author's interpretation of results

A2.7 Summarising the evidence

Study information:

- An outline of each study, will be presented in tabular form

Agreement statistics:

- Information will be summarised in tables, stratified by trial outcome

Randomised comparisons:

- A forest plot for each study will be presented including relative treatment effect estimates derived using trial follow-up data alone and routine data follow-up only, both for the trial's composite outcome and its individual components

Appendix 3 Simulation study methods

A3.1 Formulae to calculate the required event rate for each trial scenario

The event rate required for a simple parallel-group randomised trial (where the proportions of participants with an event at the end of trial are compared) was calculated using the *power.prop.test* package in R. The approach used by the package is to iteratively solve the power equation (below) based on a two-sided test for a difference in proportions.[146]

$$1-\beta = \Phi\left(\frac{(p_1-p_2)\sqrt{n}-z_{1-\frac{\alpha}{2}}\sqrt{(p_1+p_2)\left(1-\frac{p_1+p_2}{2}\right)}}{\sqrt{p_1(1-p_1)+p_2(1-p_2)}}\right) + \Phi\left(\frac{-(p_1-p_2)\sqrt{n}-z_{1-\frac{\alpha}{2}}\sqrt{(p_1+p_2)\left(1-\frac{p_1+p_2}{2}\right)}}{\sqrt{p_1(1-p_1)+p_2(1-p_2)}}\right)$$

Where $\Phi(.)$ is the cumulative distribution function of the standard Normal distribution, n the required sample size in each of the arms, p_1 and p_2 are the proportion of participants who have experienced an event by the end of the trial, α the significance level, β the probability of a type II error, and $z_{1-\frac{\alpha}{2}}$ the (1-α/2)th quantile of the standard normal distribution.

A3.2 Formulae to calculate the probability of an error occurring for each participant

The following formulae were used to convert the true events removed rate (TER) to sensitivity (Se), and false events added rate (FEA) to specificity (Sp). Where $\mathrm{N_e}$ are the total number of participants with a true event and $\mathrm{N_n}$ are the number of participants without a true event.

$$Se = 1 - TER$$

$$Sp = 1 - \left(\frac{N_e}{N_n}\right) FEA$$

The probability of a true event being removed for participants in the placebo group with a true event (p_{pe}) and participants in the active group with a true event (p_{ae}) are:

$$p_{pe} = 1 - Se$$

$$p_{ae} = 1 - Se$$

The false event risk ratio is denoted FRR. Where false events were equally distributed between trial arms (i.e. $\mathrm{FRR}=1$), the probability of a false event being added for participants in the placebo group with no true event (p_{pn}) and participants in the active group with no true event (p_{an}) were:

$$p_{pn} = 1 - Sp$$

$$p_{an} = 1 - Sp$$

Where false events were not distributed equally between trial arms (i.e. $\mathrm{FRR}\neq 1$), the probability of a false event being added for participants in the placebo group with no true event (p_{pn}) and participants in the active group with no true event (p_{an}) were:

$$p_{pn} = \frac{2(1 - Sp)}{FRR + 1}$$

$$p_{an} = \frac{2FRR(1 - Sp)}{FRR + 1}$$

Probabilities were calculated for each participant at the start of each simulated trial using a binomial distribution.

A3.3 Formulae to calculate the minimum number of repetitions required for each scenario

The primary performance measure for the simulation study was the mean difference between the adjudicated data risk ratio and routine data risk ratio. Its variance was calculated, for each scenario, using the following equation where n_{sim} was the number of repetitions, $\hat{\theta}_i$ the estimate of the i_{th} repetition, and $\bar{\theta}$ the mean value of the estimate.

$$Var(\hat{\theta}) = \frac{1}{n_{sim} - 1} \sum_{i=1}^{n_{sim}} (\hat{\theta}_i - \bar{\theta})^2$$

The bias of the performance measure was calculated using the following formula:

$$Bias = \frac{1}{n_{sim}} \sum_{i=1}^{n_{sim}} \hat{\theta}_i - \bar{\theta}$$

The Monte Carlo standard error for bias was then calculated using the formulae below:

$$Monte\ Carlo\ standard\ error\ = \sqrt{\frac{1}{n_{sim}(n_{sim} - 1)} \sum_{i=1}^{n_{sim}} (\hat{\theta}_i - \bar{\theta})^2}$$

By pre-setting Monte Carlo standard error threshold, the above formula was rearranged to calculate the minimum number of repetitions (i.e. simulated trials) required to meet this threshold.

$$n_{sim} = \frac{Var(\hat{\theta})}{Monte\ Carlo\ SE^2}$$

References

1. Collins R, MacMahon S. Reliable assessment of the effects of treatment on mortality and major morbidity, I: clinical trials. *Lancet* 2001; **357**(9253): 373-80.

2. Collins R, Bowman L, Landray M, Peto R. The Magic of Randomization versus the Myth of Real-World Evidence. *N Engl J Med* 2020; **382**(7): 674-8.

3. Collins R, Reith C, Emberson J, et al. Interpretation of the evidence for the efficacy and safety of statin therapy. *Lancet* 2016; **388**(10059): 2532-61.

4. Herrington W, Lacey B, Sherliker P, Armitage J, Lewington S. Epidemiology of Atherosclerosis and the Potential to Reduce the Global Burden of Atherothrombotic Disease. *Circ Res* 2016; **118**(4): 535-46.

5. Baigent C, Blackwell L, Emberson J, et al. Efficacy and safety of more intensive lowering of LDL cholesterol: a meta-analysis of data from 170,000 participants in 26 randomised trials. *Lancet* 2010; **376**(9753): 1670-81.

6. Hicks KA, Mahaffey KW, Mehran R, et al. 2017 Cardiovascular and Stroke Endpoint Definitions for Clinical Trials. *J Am Coll Cardiol* 2018; **71**(9): 1021-34.

7. Bowman L, Mafham M, Stevens W, et al. ASCEND: A Study of Cardiovascular Events iN Diabetes: Characteristics of a randomized trial of aspirin and of omega-3 fatty acid supplementation in 15,480 people with diabetes. *Am Heart J* 2018; **198**: 135-44.

8. ASCEND Study Collaborative Group, Bowman L, Mafham M, et al. Effects of Aspirin for Primary Prevention in Persons with Diabetes Mellitus. *N Engl J Med* 2018; **379**(16): 1529-39.

9. ASCEND Study Collaborative Group, Bowman L, Mafham M, et al. Effects of n-3 Fatty Acid Supplements in Diabetes Mellitus. *N Engl J Med* 2018; **379**(16): 1540-50.

10. Sharp Collaborative Group. Study of Heart and Renal Protection (SHARP): randomized trial to assess the effects of lowering low-density lipoprotein cholesterol among 9,438 patients with chronic kidney disease. *Am Heart J* 2010; **160**(5): 785-94 e10.

11. 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* 2011; **377**(9784): 2181-92.

12. 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**(10): 725-41.

13. SEARCH Study Collaborative Group, Bowman L, Armitage J, Bulbulia R, Parish S, Collins R. Study of the effectiveness of additional reductions in cholesterol and homocysteine (SEARCH): characteristics of a randomized trial among 12064 myocardial infarction survivors. *Am Heart J* 2007; **154**(5): 815-23, 23 e1-6.

14. HPS2-THRIVE Collabartive Group. HPS2-THRIVE randomized placebo-controlled trial in 25 673 high-risk patients of ER niacin/laropiprant: trial design, pre-specified muscle and liver outcomes, and reasons for stopping study treatment. *Eur Heart J* 2013; **34**(17): 1279-91.

15. REVEAL Collaboative Group, Bowman L, Chen F, 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, and baseline characteristics. *Am Heart J* 2017; **187**(Supplement C): 182-90.

16. Carson P, Teerlink JR, Komajda M, et al. Comparison of Investigator-Reported and Centrally Adjudicated Heart Failure Outcomes in the EMPEROR-Reduced Trial. *JACC: Heart Failure* 2023; **11**(4): 407-17.

17. Nabel EG, Braunwald E. A tale of coronary artery disease and myocardial infarction. *N Engl J Med* 2012; **366**(1): 54-63.

18. Lopez AD, Adair T. Is the long-term decline in cardiovascular-disease mortality in high-income countries over? Evidence from national vital statistics. *International Journal of Epidemiology* 2019; **48**(6): 1815-23.

19. Roberts DA, Kantarjian HM, Steensma DP. Contract research organizations in oncology clinical research: Challenges and opportunities. *Cancer* 2016; **122**(10): 1476-82.

20. Eisenstein EL, Collins R, Cracknell BS, et al. Sensible approaches for reducing clinical trial costs. *Clin Trials* 2008; **5**(1): 75-84.

21. Sertkaya A, Wong H-H, Jessup A, Beleche T. Key cost drivers of pharmaceutical clinical trials in the United States. *Clinical Trials* 2016; **13**(2): 117-26.

22. Collins R. Back to the future: the urgent need to re-introduce streamlined trials. *European Heart Journal Supplements* 2018; **20**(suppl_C): C14-C7.

23. Pathak J, Kho AN, Denny JC. Electronic health records-driven phenotyping: challenges, recent advances, and perspectives. *Journal of the American Medical Informatics Association* 2013; **20**(e2): e206-e11.

24. Mc Cord KA, Al-Shahi Salman R, Treweek S, et al. Routinely collected data for randomized trials: promises, barriers, and implications. *Trials* 2018; **19**(1): 29.

25. Pletcher MJ, Forrest CB, Carton TW. PCORnet's Collaborative Research Groups. *Patient Relat Outcome Meas* 2018; **9**: 91-5.

26. Group RC, Horby P, Lim WS, et al. Dexamethasone in Hospitalized Patients with Covid-19. *N Engl J Med* 2021; **384**(8): 693-704.

27. Pessoa-Amorim G, Campbell M, Fletcher L, et al. Making trials part of good clinical care: lessons from the RECOVERY trial. *Future Healthc J* 2021; **8**(2): e243-e50.

28. NHS Digital Transforming health and care through technology: Delivering the strategy set out in Personalised Health and Care 2020. https://digital.nhs.uk/about-nhs-digital/our-work/transforming-health-and-care-through-technology. Accessed February 17, 2021.

29. Patient Episode Dataset for Wales. Health Data Research Innovation Gateway. https://www.healthdatagateway.org/dataset/4c33a5d2-164c-41d7-9797-dc2b008cc852. Accessed March 1, 2021.

30. General Acute Inpatient and Day Case - Scottish Morbidity Record (SMR01). Health Data Research Innovation Gateway. https://www.healthdatagateway.org/dataset/98cda353-0011-45b2-80ca-4ed24cd084bf. Accessed March 1, 2021.

31. Hospital Episode Statistics Admitted Patient Care. Health Data Research Innovation Gateway. https://web.www.healthdatagateway.org/dataset/2b6409db-a669-4bef-9fd5-39c2b6f8d5e9. Accessed March 1, 2021.

32. Civil Registration - Deaths. Health Data Research Innovation Gateway. https://www.healthdatagateway.org/dataset/050163dc-1728-4ac5-a7d9-4dd3ca0ca12a. Accessed March 1, 2021.

33. National Records of Scotland (NRS) - Deaths Data. Health Data Research Innovation Gateway. https://www.healthdatagateway.org/dataset/e600dae2-a83c-4b7a-8d23-af4ac31ca374. Accessed March 1, 2021.

34. Clinical Classifications. NHS Digital. https://digital.nhs.uk/services/terminology-and-classifications/clinical-classifications. Accessed March 1, 2021.

35. Terminology Services and Clinical Coding: Scottish Clinical Coding Standards. Public Health Scotland. https://www.isdscotland.org/Products-and-Services/Terminology-Services/Clinical-Coding-Guidelines/. Accessed March 1, 2021. .

36. Welsh Clinical Coding Standards and Guidance. NHS Wales: Digital Health and Care Wales. https://dhcw.nhs.wales/information-services/information-standards/clinical-classifications-and-terminology-standards/welsh-clinical-coding-standards-and-guidance/. Accessed January 24, 2023.

37. (WHO) WHO. ICD-10 International Statistical Classification of Diseases and Related Health Problems: v. 2: Instruction Manual by World Health Organization (WHO): World Health Organization (WHO); 1993.

38. Comission A. Improving coding, costing and comissioning: Annual report on the Payment by Results data assurance programme 2010/11, 2011.

39. Burns EM, Rigby E, Mamidanna R, et al. Systematic review of discharge coding accuracy. *Journal of public health (Oxford, England)* 2012; **34**(1): 138-48.

40. Storey BC, Staplin N, Harper CH, et al. Declining comorbidity-adjusted mortality rates in English patients receiving maintenance renal replacement therapy. *Kidney Int* 2018; **93**(5): 1165-74.

41. Classification Browser ICD-10-5TH-Edition. NHS Digital. https://classbrowser.nhs.uk/#/book/ICD-10-5TH-Edtion. Accessed January 24, 2023.

42. Classification Browser OPCS-4.9. NHS Digital. https://classbrowser.nhs.uk/#/book/OPCS-4.9. Accessed January 24, 2023.

43. Sudlow C, Gallacher J, Allen N, et al. UK biobank: an open access resource for identifying the causes of a wide range of complex diseases of middle and old age. *PLoS Med* 2015; **12**(3): e1001779.

44. Green J, Reeves GK, Floud S, et al. Cohort Profile: the Million Women Study. *Int J Epidemiol* 2019; **48**(1): 28-9e.

45. Lensen S, Macnair A, Love SB, et al. Access to routinely collected health data for clinical trials - review of successful data requests to UK registries. *Trials* 2020; **21**(1): 398.

46. McKay AJ, Jones AP, Gamble CL, Farmer AJ, Williamson PR. Use of routinely collected data in a UK cohort of publicly funded randomised clinical trials. *F1000Res* 2020; **9**: 323.

47. Macnair A, Love SB, Murray ML, et al. Accessing routinely collected health data to improve clinical trials: recent experience of access. *Trials* 2021; **22**(1): 340.

48. Murray ML, Love SB, Carpenter JR, et al. Data provenance and integrity of health-care systems data for clinical trials. *Lancet Digit Health* 2022; **4**(8): e567-e8.

49. Roe MT. Leveraging electronic health record data for pragmatic randomized trials. *Clinical Trials* 2020: 1740774520931036.

50. Williams ADN, Davies G, Farrin AJ, et al. A DELPHI study priority setting the remaining challenges for the use of routinely collected data in trials: COMORANT-UK. *Trials* 2023; **24**(1): 243.

51. Clark TG, Bradburn MJ, Love SB, Altman DG. Survival Analysis Part I: Basic concepts and first analyses. *British Journal of Cancer* 2003; **89**(2): 232-8.

52. Braunwald E. Gliflozins in the Management of Cardiovascular Disease. *New England Journal of Medicine* 2022; **386**(21): 2024-34.

53. Peto R, Pike MC, Armitage P, et al. Design and analysis of randomized clinical trials requiring prolonged observation of each patient. II. analysis and examples. *Br J Cancer* 1977; **35**(1): 1-39.

54. Yusuf S, Peto R, Lewis J, Collins R, Sleight P. Beta blockade during and after myocardial infarction: an overview of the randomized trials. *Prog Cardiovasc Dis* 1985; **27**(5): 335-71.

55. Peto R, Pike MC, Armitage P, et al. Design and analysis of randomized clinical trials requiring prolonged observation of each patient. I. Introduction and design. *Br J Cancer* 1976; **34**(6): 585-612.

56. Cohen J. A Coefficient of Agreement for Nominal Scales. *Educational and Psychological Measurement* 1960; **20**(1): 37-46.

57. Landis JR, Koch GG. The measurement of observer agreement for categorical data. *Biometrics* 1977; **33**(1): 159-74.

58. Hlatky MA, Ray RM, Burwen DR, et al. Use of Medicare data to identify coronary heart disease outcomes in the Women's Health Initiative. *Circ Cardiovasc Qual Outcomes* 2014; **7**(1): 157-62.

59. Herzog CA, Littrell K, Arko C, Frederick PD, Blaney M. Clinical characteristics of dialysis patients with acute myocardial infarction in the United States: a collaborative project of the United States Renal Data System and the National Registry of Myocardial Infarction. *Circulation* 2007; **116**(13): 1465-72.

60. Herzog CA, Asinger RW, Berger AK, et al. Cardiovascular disease in chronic kidney disease. A clinical update from Kidney Disease: Improving Global Outcomes (KDIGO). *Kidney Int* 2011; **80**(6): 572-86.

61. Rodrigues C, Odutayo A, Patel S, et al. Accuracy of Cardiovascular Trial Outcome Ascertainment and Treatment Effect Estimates from Routine Health Data: A Systematic Review and Meta-Analysis. *Circulation: Cardiovascular Quality and Outcomes* 2021; **14**(5): e007903.

62. Edwards D. EXPLORING THE USE OF ROUTINE DATA FOR RECRUITMENT AND FOLLOW-UP IN LARGE RANDOMISED TRIALS: University of Oxford; 2019.

63. Harper C, Mafham M, Herrington W, et al. Comparison of the Accuracy and Completeness of Records of Serious Vascular Events in Routinely Collected Data vs Clinical Trial–Adjudicated Direct Follow-up Data in the UK: Secondary Analysis of the ASCEND Randomized Clinical Trial. *JAMA netw* 2021; **4**(12): e2139748-e.

64. Shepherd J, Cobbe SM, Ford I, et al. Prevention of coronary heart disease with pravastatin in men with hypercholesterolemia. West of Scotland Coronary Prevention Study Group. *N Engl J Med* 1995; **333**(20): 1301-7.

65. Barry SJ, 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. *PLoS One* 2013; **8**(9): e75379.

66. 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**(9326): 7-22.

67. Blake SR, Roome C, Shahzad A, et al. A comparison of hospital episode statistics and traditional methods to identify outcomes in a randomized trial; a sub-study of HEAT-PPCI. *Journal of public health (Oxford, England)* 2020; **42**(1): 175-82.

68. Shahzad A, Kemp I, Mars C, et al. Unfractionated heparin versus bivalirudin in primary percutaneous coronary intervention (HEAT-PPCI): an open-label, single centre, randomised controlled trial. *The Lancet* 2014; **384**(9957): 1849-58.

69. Merriel SWD, Turner EL, Walsh E, et al. Cross-sectional study evaluating data quality of the National Cancer Registration and Analysis Service (NCRAS) prostate cancer registry data using the Cluster randomised trial of PSA testing for Prostate cancer (CAP). *BMJ Open* 2017; **7**(11): e015994.

70. Thorn JC, Turner E, Hounsome L, et al. Validation of the Hospital Episode Statistics Outpatient Dataset in England. *Pharmacoeconomics* 2016; **34**(2): 161-8.

71. Battersby C, Statnikov Y, Santhakumaran S, Gray D, Modi N, Costeloe K. The United Kingdom National Neonatal Research Database: A validation study. *PLoS One* 2018; **13**(8): e0201815.

72. Sperrin M, Webb DJ, Patel P, et al. Chronic obstructive pulmonary disease exacerbation episodes derived from electronic health record data validated using clinical trial data. *Pharmacoepidemiol Drug Saf* 2019.

73. Powell GA, Bonnett LJ, Smith CT, Hughes DA, Williamson PR, Marson AG. Using routinely recorded data in a UK RCT: a comparison to standard prospective data collection methods. *Trials* 2021; **22**(1): 429.

74. Wright-Hughes A, Graham E, Cottrell D, Farrin A. Routine hospital data - is it good enough for trials? An example using England's Hospital Episode Statistics in the SHIFT trial of Family Therapy vs. Treatment as Usual in adolescents following self-harm. *Clin Trials* 2018: 1740774517751381.

75. Kilburn LS, Aresu M, Banerji J, Barrett-Lee P, Ellis P, Bliss JM. Can routine data be used to support cancer clinical trials? A historical baseline on which to build: retrospective linkage of data from the TACT (CRUK 01/001) breast cancer trial and the National Cancer Data Repository. *Trials* 2017; **18**(1): 561.

76. Kalsi JK, Ryan A, Gentry-Maharaj A, et al. Completeness and accuracy of national cancer and death registration for outcome ascertainment in trials-an ovarian cancer exemplar. *Trials* 2021; **22**(1): 88.

77. Mehran R, Rao SV, Bhatt DL, et al. Standardized bleeding definitions for cardiovascular clinical trials: a consensus report from the Bleeding Academic Research Consortium. *Circulation* 2011; **123**(23): 2736-47.

78. Higgins JPT, Altman DG, Gøtzsche PC, et al. The Cochrane Collaboration's tool for assessing risk of bias in randomised trials. *BMJ* 2011; **343**: d5928.

79. 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.

80. Li L, Binney LE, Luengo-Fernandez R, Silver LE, Rothwell PM, Oxford Vascular S. Temporal trends in the accuracy of hospital diagnostic coding for identifying acute stroke: A population-based study. *Eur Stroke J* 2020; **5**(1): 26-35.

81. Seminog OO, Scarborough P, Wright FL, Rayner M, Goldacre MJ. Determinants of the decline in mortality from acute stroke in England: linked national database study of 795 869 adults. *BMJ* 2019; **365**: l1778.

82. European Society of Cardiology: Clinical Practice Guidelines. https://www.escardio.org/Guidelines/Clinical-Practice-Guidelines. Accessed February 17, 2021.

83. 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): 1073-80.

84. SCOT-HEART Investigators, Newby DE, Adamson PD, et al. Coronary CT Angiography and 5-Year Risk of Myocardial Infarction. *N Engl J Med* 2018; **379**(10): 924-33.

85. Herrett E, Gallagher AM, Bhaskaran K, et al. Data Resource Profile: Clinical Practice Research Datalink (CPRD). *Int J Epidemiol* 2015; **44**(3): 827-36.

86. Brown LD, Cai TT, DasGupta A. Interval Estimation for a Binomial Proportion. *Statistical Science* 2001; **16**(2): 101-17.

87. Fleiss J.L. LB, Paik M. C. Statistical methods for rates and proportions. 3rd ed. Hoboken: John Wiley & Sons; 2003.

88. Brazzelli M, Shuler K, Quayyum Z, et al. Clinical and imaging services for TIA and minor stroke: results of two surveys of practice across the UK. *BMJ Open* 2013; **3**(8): e003359.

89. Kjoller E, Hilden J, Winkel P, et al. Agreement between public register and adjudication committee outcome in a cardiovascular randomized clinical trial. *Am Heart J* 2014; **168**(2): 197-204 e1-4.

90. Ndounga Diakou LA, Trinquart L, Hrobjartsson A, et al. Comparison of central adjudication of outcomes and onsite outcome assessment on treatment effect estimates. *Cochrane Database Syst Rev* 2016; **3**(3): MR000043.

91. ICH Harmonised tripartite guideline. Clinical safety data management: definitions and standards for expedited reporting E2A. Oct 1994.

92. Health Data Research UK. https://www.hdruk.ac.uk/. Accessed February 17, 2021.

93. Frobert O, Lagerqvist B, Olivecrona GK, et al. Thrombus aspiration during ST-segment elevation myocardial infarction. *N Engl J Med* 2013; **369**(17): 1587-97.

94. Jernberg T, Attebring MF, Hambraeus K, et al. The Swedish Web-system for enhancement and development of evidence-based care in heart disease evaluated according to recommended therapies (SWEDEHEART). *Heart* 2010; **96**(20): 1617-21.

95. Jones WS, Mulder H, Wruck LM, et al. Comparative Effectiveness of Aspirin Dosing in Cardiovascular Disease. *N Engl J Med* 2021; **384**(21): 1981-90.

96. Rannikmäe K, Ngoh K, Bush K, et al. Accuracy of identifying incident stroke cases from linked health care data in UK Biobank. *Neurology* 2020; **95**(6): e697-e707.

97. Landray MJ, Bax JJ, Alliot L, et al. Improving public health by improving clinical trial guidelines and their application. *European Heart Journal* 2017; **38**(21): 1632-7.

98. Kikkert WJ, Tijssen JGP, Piek JJ, Henriques JPS. Challenges in the adjudication of major bleeding events in acute coronary syndrome: a plea for a standardized approach and guidance to adjudication. *Eur Heart J* 2016; **37**(14): 1104-12.

99. Margolis KL, Mahady SE, Nelson MR, et al. Development of a standardized definition for clinically significant bleeding in the ASPirin in Reducing Events in the Elderly (ASPREE) trial. *Contemp Clin Trials Commun* 2018; **11**: 30-6.

100. Schulman S, Kearon C. Definition of major bleeding in clinical investigations of antihemostatic medicinal products in non-surgical patients. *J Thromb Haemost* 2005; **3**(4): 692-4.

101. Wiviott SD, Antman EM, Gibson CM, et al. Evaluation of prasugrel compared with clopidogrel in patients with acute coronary syndromes: design and rationale for the TRial to assess

Improvement in Therapeutic Outcomes by optimizing platelet InhibitioN with prasugrel Thrombolysis In Myocardial Infarction 38 (TRITON-TIMI 38). *Am Heart J* 2006; **152**(4): 627-35.

102. Pasea L, Chung SC, Pujades-Rodriguez M, et al. Bleeding in cardiac patients prescribed antithrombotic drugs: electronic health record phenotyping algorithms, incidence, trends and prognosis. *BMC Med* 2019; **17**(1): 206.

103. Pasea L, Chung SC, Pujades-Rodriguez M, et al. Personalising the decision for prolonged dual antiplatelet therapy: development, validation and potential impact of prognostic models for cardiovascular events and bleeding in myocardial infarction survivors. *Eur Heart J* 2017; **38**(14): 1048-55.

104. Timmis A, Rapsomaniki E, Chung SC, et al. Prolonged dual antiplatelet therapy in stable coronary disease: comparative observational study of benefits and harms in unselected versus trial populations. *Bmj* 2016; **353**: i3163.

105. Hippisley-Cox J, Coupland C. Predicting risk of upper gastrointestinal bleed and intracranial bleed with anticoagulants: cohort study to derive and validate the QBleed scores. *BMJ* 2014; **349**: g4606.

106. Button LA, Roberts SE, Evans PA, et al. Hospitalized incidence and case fatality for upper gastrointestinal bleeding from 1999 to 2007: a record linkage study. *Alimentary pharmacology & therapeutics* 2011; **33**(1): 64-76.

107. Marquis-Gravel G, Hammill BG, Mulder H, et al. Validation of Cardiovascular End Points Ascertainment Leveraging Multisource Electronic Health Records Harmonized Into a Common Data Model in the ADAPTABLE Randomized Clinical Trial. *Circulation: Cardiovascular Quality and Outcomes* 2021; **14**(12): e008190.

108. Mullen L, Meah MN, Elamin A, et al. Risk of Major Bleeding With Potent Antiplatelet Agents After an Acute Coronary Event: A Comparison of Ticagrelor and Clopidogrel in 5116 Consecutive Patients in Clinical Practice. *J Am Heart Assoc* 2021; **10**(8): e019467.

109. NHS DigiTrials. https://digital.nhs.uk/services/nhs-digitrials. Accessed September 22, 2022.

110. Heckbert SR, Kooperberg C, Safford MM, et al. Comparison of Self-Report, Hospital Discharge Codes, and Adjudication of Cardiovascular Events in the Women's Health Initiative. *American Journal of Epidemiology* 2004; **160**(12): 1152-8.

111. Ninomiya T, Donnan G, Anderson N, et al. Effects of the end point adjudication process on the results of the perindopril protection against recurrent stroke study (PROGRESS). *Stroke* 2009; **40**(6): 2111-5.

112. Tyl B, Sendon JL, Borer JS, et al. Comparison of Outcome Adjudication by Investigators and by a Central End Point Committee in Heart Failure Trials. *Circulation: Heart Failure* 2020; **13**(7): e006720.

113. Gaba P, Bhatt DL, Giugliano RP, et al. Comparative Reductions in Investigator-Reported and Adjudicated Ischemic Events in REDUCE-IT. *Journal of the American College of Cardiology* 2021; **78**(15): 1525-37.

114. Stevens PE, O'Donoghue DJ, de Lusignan S, et al. Chronic kidney disease management in the United Kingdom: NEOERICA project results. *Kidney Int* 2007; **72**(1): 92-9.

115. Baigent C, Herrington WG, Coresh J, et al. Challenges in conducting clinical trials in nephrology: conclusions from a Kidney Disease-Improving Global Outcomes (KDIGO) Controversies Conference. *Kidney Int* 2017; **92**(2): 297-305.

116. Haynes R, Lewis D, Emberson J, et al. Effects of lowering LDL cholesterol on progression of kidney disease. *J Am Soc Nephrol* 2014; **25**(8): 1825-33.

117. Peplow M. The 100,000 Genomes Project. *BMJ* 2016; **353**: i1757.

118. Denaxas SC, George J, Herrett E, et al. Data resource profile: cardiovascular disease research using linked bespoke studies and electronic health records (CALIBER). *Int J Epidemiol* 2012; **41**(6): 1625-38.

119. Harper C, Mafham M, Herrington W, et al. Comparison of the Accuracy and Completeness of Records of Serious Vascular Events in Routinely Collected Data vs Clinical Trial-Adjudicated Direct Follow-up Data in the UK: Secondary Analysis of the ASCEND Randomized Clinical Trial. *JAMA Netw Open* 2021; **4**(12): e2139748.

120. Ndounga Diakou LA, Trinquart L, Hrobjartsson A, et al. Comparison of central adjudication of outcomes and onsite outcome assessment on treatment effect estimates. *Cochrane Database Syst Rev* 2016; **3**: MR000043.

121. Pogue J, Walter SD, Yusuf S. Evaluating the benefit of event adjudication of cardiovascular outcomes in large simple RCTs. *Clin Trials* 2009; **6**(3): 239-51.

122. Harper C, Mafham M, Herrington W, et al. Reliability of major bleeding events in UK routine data versus clinical trial adjudicated follow-up data. *Heart* 2023: heartjnl-2023-322616.

123. Herrington WG, Harper C, Staplin N, et al. Impact of Outcome Adjudication in Kidney Disease Trials: Observations From the Study of Heart and Renal Protection. *Kidney International Reports* 2023.

124. Buyse M, Squifflet P, Coart E, Quinaux E, Punt CJ, Saad ED. The impact of data errors on the outcome of randomized clinical trials. *Clin Trials* 2017; **14**(5): 499-506.

125. Godolphin PJ, Bath PM, Partlett C, et al. Outcome assessment by central adjudicators in randomised stroke trials: Simulation of differential and non-differential misclassification. *European Stroke Journal* 2020; **5**(2): 174-83.

126. Van den Bor RM. Innovations in monitoring and data quality control in clinical trials. GVO drukkers & vormgevers B.V.: Utrecht University; 2017.

127. Herrington WG, Staplin N, Haynes R. Kidney disease trials for the 21st century: innovations in design and conduct. *Nature Reviews Nephrology* 2020; **16**(3): 173-85.

128. Korn EL, Dodd LE, Freidlin B. Measurement error in the timing of events: effect on survival analyses in randomized clinical trials. *Clin Trials* 2010; **7**(6): 626-33.

129. Morgan AD, Sinnott SJ, Smeeth L, Minassian C, Quint J. Concordance in the recording of stroke across UK primary and secondary care datasets: a population-based cohort study. *BJGP Open* 2020.

130. A Study of Cardiovascular Events iN Diabetes Plus (ASCEND PLUS). https://clinicaltrials.gov/ct2/show/NCT05441267. Accessed November 25, 2022.

131. Aspirin to Target Arterial Events in Chronic Kidney Disease (ATTACK). https://clinicaltrials.gov/ct2/show/NCT03796156. Accessed November 25, 2022.

132. Aspirin after hospitalisation with Pneumonia to prevent cardiovascular Events randomised Controlled Trial (ASPECT). https://fundingawards.nihr.ac.uk/award/NIHR132968. Accessed November 25, 2022.

133. BHF Data Science Centre. https://www.hdruk.ac.uk/helping-with-health-data/bhf-data-science-centre/. Accessed April 21, 2023. .

134. Sydes MR, Barbachano Y, Bowman L, et al. Realising the full potential of data-enabled trials in the UK: a call for action. *BMJ Open* 2021; **11**(6): e043906.

135. MRC-NIHR TMRP Health Informatics. https://www.methodologyhubs.mrc.ac.uk/research/working-groups/health-informatics/. Accessed April 21, 2023.

136. Levin A, Agarwal R, Herrington WG, et al. International consensus definitions of clinical trial outcomes for kidney failure: 2020. *Kidney Int* 2020; **98**(4): 849-59.

137. Group TE-KC. Design, recruitment, and baseline characteristics of the EMPA-KIDNEY trial. *Nephrol Dial Transplant* 2022.

138. Powell GA, Bonnett LJ, Tudur-Smith C, Hughes DA, Williamson PR, Marson AG. Using routinely recorded data in the UK to assess outcomes in a randomised controlled trial: The Trials of Access. *Trials* 2017; **18**(1): 389.

139. Facile R, Muhlbradt EE, Gong M, et al. Use of Clinical Data Interchange Standards Consortium (CDISC) Standards for Real-world Data: Expert Perspectives From a Qualitative Delphi Survey. *JMIR Med Inform* 2022; **10**(1): e30363.

140. Care.data: How did it go so wrong? BBC News. https://www.bbc.co.uk/news/health-26259101. Accessed November 25, 2022.

141. Data saves lives. https://www.herc.ac.uk/get-involved/data-saves-lives/. Accessed April 25, 2023.

142. Brennan JM, Wruck L, Pencina MJ, et al. Claims-based cardiovascular outcome identification for clinical research: Results from 7 large randomized cardiovascular clinical trials. *Am Heart J* 2019; **218**: 110-22.

143. Butala NM, Strom JB, Faridi KF, et al. Validation of Administrative Claims to Ascertain Outcomes in Pivotal Trials of Transcatheter Aortic Valve Replacement. *JACC Cardiovasc Interv* 2020; **13**(15): 1777-85.

144. Guimarães PO, Krishnamoorthy A, Kaltenbach LA, et al. Accuracy of Medical Claims for Identifying Cardiovascular and Bleeding Events After Myocardial Infarction : A Secondary Analysis of the TRANSLATE-ACS Study. *JAMA Cardiol* 2017; **2**(7): 750-7.

145. Anderson GL, Burns CJ, Larsen J, Shaw PA. Use of administrative data to increase the practicality of clinical trials: Insights from the Women's Health Initiative. *Clin Trials* 2016; **13**(5): 519-26.

146. Cohen J. Statistical Power Analysis for the Behavioral Sciences. 2nd Edition ed. New York: Routledge; 1988.